

Two vertical panels are positioned on the left side of the slide. The left panel shows a close-up of a textured surface with a grid-like pattern of wavy lines, possibly representing a biological or geological structure. The right panel shows a traditional Aboriginal rock painting, featuring a central figure with a sunburst-like head and a body composed of vertical lines, set against a reddish-brown background.

# Lunch Break

Please be back by 1:30pm



# Data Management and Data Management Planning

# Data management & data management planning

- ANDS website contains
  - Guides
  - Links to other resources
    - Australian universities
    - UK Data Archive *Managing and Sharing Data*
    - MIT Libraries *Data Management & Publishing*
    - DCC *Data Management Plan Content Checklist*



# What do we mean by data?

- ‘facts and statistics collected together for reference or analysis.’

*(Oxford Dictionary of English, 2nd ed.*

Revised.)

# What do we mean by *research* data?

- Any data that is of interest to researchers
- NOT administrative data
- NOT teaching data

# Data management – what is it?

- Data management is a broad catch-all term used by different people in different contexts. It can be used to describe a variety of activities such as:
  - Data storage,
  - Data curation,
  - Data preservation,
  - Database design,
  - Data modeling and more.
- Sometimes it can be used to refer to data management policy and sometimes to the practice of data management.

# Data management for the researcher

- All those activities which a researcher can undertake
  - to organise and manage their data
  - to facilitate their own research, and
  - to provide a foundation for the longer-term sustainability of the data

# What's in it for the researcher?

- meeting obligations
- efficiency
- protection
- quality
- exposure
- avoiding catastrophe

# Why is data management important to the *Code*?

- ‘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, particularly where the research would be difficult or impossible to repeat..’
- ‘good stewardship of public resources...’



# Data Management Planning

# Creating a data management plan

## The introduction

- About the project
- Owners and stakeholders
- The aims of the research

# Re-using and creating data

- Survey of existing data you might use
  - Can you get access?
  - Under what conditions?
- Data to be created
  - Describe in detail the different forms

# Copyright, IP and other legal issues

- Copyright and IP
  - usually defined by your employer
  - will the data be created in Australia or overseas?
  - will the research project involve partners from other institutions, Australian or otherwise?
  - who will own the data?
  - how will the data be licensed?
- Ethical and policy issues

# Access and security

- Will others have access to the data during the life of the project?
  - Inside your institution or externally?
  - As part of the research team?
  - On open access?
- How will their access by yourself and others be controlled?
  - Internally and externally

# File formats (1)

- *A file format* is a way of organising meaningful information into a sequence of bits and bytes for storage in a computer system
- Can often be identified from the file name extension, e.g.
  - .doc
  - .html or .htm
  - .jpg or .jpeg

## File formats (2)



*“My data system isn’t speaking to your data system.”*

- What’s the problem?
  - Durability
  - Proprietary vs. open access
  - Degradation
  - Compression
  - Standards, within your discipline and more broadly
  - Compatibility with collaborators
  - Choosing the right one
- Planning implications
  - Get it sorted before you start

Any questions at this stage?

# Documentation & record keeping

- Data asset inventory
- File naming protocols
- Metadata standards
- Contextual information
- Protocols used
- Criteria for quality assurance
- Software used
- Lab notebooks

# Metadata (1)

- The term *metadata* refers to information used to describe items and groups of items.
- It is data about data.
- It can be used to describe physical items as well as digital items.
- It provides meaning and context.

# Metadata (2)

- Types
  - Descriptive
  - Technical
  - Access or rights
  - Preservation
- Standards
  - Disciplinary and other
  - Vocabularies
  - Ontologies

# 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?

# Bibliography management

- What bibliographic management tools will you use?
- How will you share references with the other members of your group?

# Short term storage and backup

- Where to store?
  - Know your institution
  - Know your faculty/school
- How to back up?
  - How much is enough?
- Don't forget your non-digital data

# Retention and the *Code* (1)

The *Code* says

- In general, the minimum recommended period for retention of research data is 5 years from the date of publication.
- However, in any particular case, the period for which data should be retained should be determined by the specific type of research.

## Retention and the *Code* (2)

- For example:
  - for short-term research projects that are for assessment purposes only, such as research projects completed by students, retaining research data for 12 months after the completion of the project may be sufficient
  - for most clinical trials, retaining research data for 15 years or more may be necessary
  - for areas such as gene therapy, research data must be retained permanently (e.g. patient records)
  - if the work has community or heritage value, research data should be kept permanently at this stage, preferably within a national collection.

# Retention and disposal

- Longer-term retention
  - ‘The institutional policy on the secure and safe disposal of primary materials research data must be followed’ (*Code*)
  - Is recommended by ANDS in order to encourage data sharing and re-use
  - Is recommended where the data has intrinsic properties which make it unique
  - May be affected if data is not properly managed from the outset

# Responsibilities

- Who is responsible for each part of the plan?
- Who is responsible for
  - Approval of the plan?
  - Modification and review?
  - Storage of the plan?

# Budget

Once the plan is complete you are in a position to estimate any costs for which you may be responsible

- Documentation and metadata
- Equipment and software
- Storage
- ...?

# Other issues

- Meeting the requirements of funding bodies
- Understanding disciplinary differences

Any questions?



# Small group discussion

# Discussion topics

- How to engage with your institution around data? Who to talk to? What to ask for? How to negotiate the balance of responsibility between the individual researcher (or research group or department) and the institution as a whole?
- What different approaches could institutions take to fulfil their responsibilities? How might each of these approaches impact researchers in those institutions?
- (For medical and human subjects research disciplines only) How do you manage the tension between ethical/confidentiality requirements for keeping data private and the *Code's* retention and sharing requirements?

# How to participate in the online forum

- <http://community.ands.org.au/>
- Log in
  - Username = table1 etc; Password = table1 etc
- To participate in the online discussion
  - Click 'ANDS Roadshow 2009-10', then 'Canberra', then 'Research data and the Code'.
  - Click on 'Post Reply'.
  - Then start typing...
  - When finished, click 'Submit' (*not* 'Save')

# Thank you for your attention

## Questions?

Contact ANDS via:

<http://ands.org.au/contact.html>

ANDS Resources:

<http://ands.org.au/resource/>