



Research Data and the *Code*: The Researcher Perspective

Outline

- About ANDS
- The *Code* & your obligations
- Data sharing & re-use
- Open Data Best Practice: Managing copyright – Professor Anne Fitzgerald, QUT
- Break
- Data management & data management planning
- Discussion
- Report back
- Summing up

About us

- Margaret Henty
Capabilities Program Leader
ANDS, ANU
- Dr Ian Barnes
Research Analyst
ANDS, ANU



About ANDS

Why ANDS?

- Data is an increasingly important component of the research enterprise as new means of data capture and storage become available
- Research collaboration is vital to crack some of the major challenges facing humanity in the twenty-first century

Therefore, the ANDS vision is:

‘More researchers sharing and re-using data more often’

How are we working to achieve this?

- By lowering the costs of capture and finding data and increasing the benefits of sharing data
- Through creating a populated data commons — the Australian Research Data Commons — with tools and processes to enable its effective use

What is ANDS?

- ANDS is an unincorporated Joint Venture to implement the National Collaborative Research Investment Strategy (NCRIS) ANDS program and the Education Investment Fund (EIF) ARDC program
- A collaboration between Monash University, the Australian National University and CSIRO
- Operational locations at Monash & ANU

Enablers

- Shared desire to change
- Professional services – research data analysts, research data carers, professional programmers
- Change partners such as eResearch organisations
- Changed status of research data

The ANDS constituency

■ Organisations

- Universities, research institutions, large instrument facilities, government departments, galleries, libraries, archives and museums

■ People

- Researchers and their support staff
- Repository and data centre managers and technical staff
- Research administrators, government policy-makers and funders



The Code

The Australian Code for the Responsible Conduct of Research

- The *Code* describes the responsibilities of institutions and researchers in the management of research data and primary materials.
- Institutions are to retain research data, provide secure data storage, identify ownership, and ensure security and confidentiality of research data.
- Researchers are to retain research data and primary materials, manage storage of research data and primary materials, maintain confidentiality of research data and primary materials.

The *Code* is not currently being enforced, however all universities are signatories.

ANDS and the *Code*

- ANDS has no formal responsibility for the *Code*.
- Many of the aims of the *Code* coincide with the aims of ANDS.
- The *Code* makes significant references to research data and responsibilities related to it.
- ANDS is, therefore, interested in promoting awareness and supporting the uptake of the *Code*.
 - ANDS has conducted fora on institutional responses to the data aspects of the *Code*.

What this workshop is **not** about

- Provisions in the *Code* that don't relate to data
- Data that isn't digital
- The institutional perspective

What does the *Code* say about data? (1)

- 'The responsible conduct of research includes the **proper management and retention** of the research data.'

What does the *Code* say about data? (2)

- ‘The researcher must decide which data and materials should be retained, although in some cases this is determined by law, funding agency, publisher or by convention in the discipline. 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, particularly where the research would be difficult or impossible to repeat.’

Responsibilities of institutions

- Retain research data and primary materials
- Provide secure research data storage and record-keeping facilities
- Identify ownership of research data and primary materials
- Ensure security and confidentiality of research data and primary materials

Responsibilities of researchers (1)

- Retain research data and primary materials
 - As long as interest and discussion persist
 - Make it available to others
 - At least as long as the institution requires
 - If research is challenged, everything must be kept until the challenge is resolved
 - If there is an allegation of misconduct, everything must be kept
 - Secure and safe disposal

Responsibilities of researchers (2)

- Manage storage of research data and primary materials
 - Keep clear and accurate records
 - Safe and secure storage
 - Keep primary records too (e.g. lab notebooks)
 - Keep data in durable, indexed and retrievable form
 - Keep a catalogue
 - Follow ethical protocols and legislation

Responsibilities of researchers (3)

- Maintain confidentiality of research data and primary materials
 - Only use confidential data in ways agreed with providers
 - Take particular care when making confidential data available for discussion



Data sharing and re-use

Data sharing and re-use

What the *Code* has to say

- ‘A strong research culture will demonstrate ... good stewardship of public resources used to conduct research.’
- ‘Policies are required that address the ownership of research materials and data, their storage, their retention beyond the end of the project, and appropriate access to them by the research community.’
- ‘The potential value of the material for further research should [...] be considered, particularly where the research would be difficult or impossible to repeat.’

What the *Code* has to say (2)

- ‘Research data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters.’

The ANDS position: data should be made available unless there is a good reason not to.

Data sharing and re-use – the practicalities

- How will the data be made available?
- Where will the data be stored?
- How will the dataset and its accompanying documentation be described?
- Who will provide that description?
- Will the data be openly available or available only on request?
- Are there embargo periods?

Benefits of data sharing for you

- Greater recognition of your work through increased citations and follow-on research
- Multiple copies protect the data from accidental erasure
- May lead to new collaborations with data users
- Demonstrates the value of your work by allowing continued re-use of the data, which in turn may lead to further funding
- Meets the contractual obligations of research funders who have data sharing policies
- Leads to satisfaction as a result of benefiting the broader community

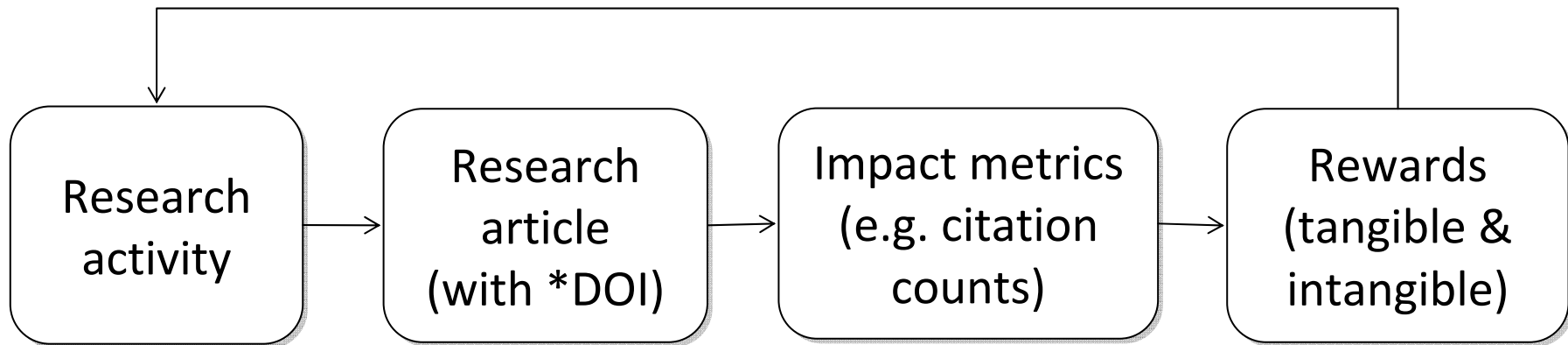
Broader benefits of data sharing

- Productivity increases through reduction of costs.
- A community of researchers can develop around a common resource.
- Independent lines of research based on the same data might result in a better research outcome.
- The risk of duplicating data collection efforts is reduced.

Obstacles to data sharing

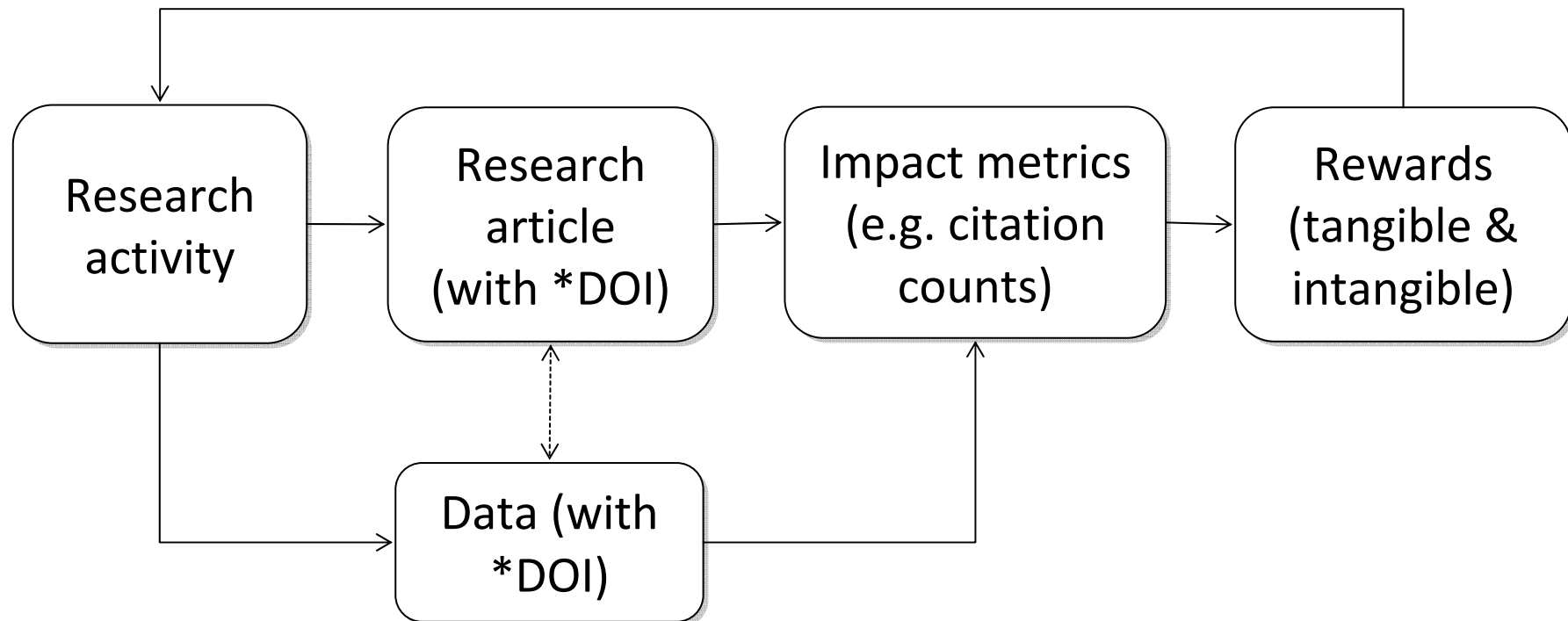
- Not yet finished publishing
- Lack of time
- Lack of resources
- Confidentiality
- Lack of reward

Rewards: what we have...



*DOI: Digital Object Identifier

Rewards: what we need...

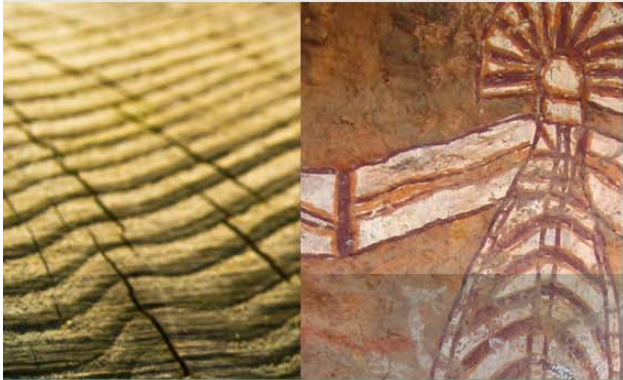


*DOI: Digital Object Identifier

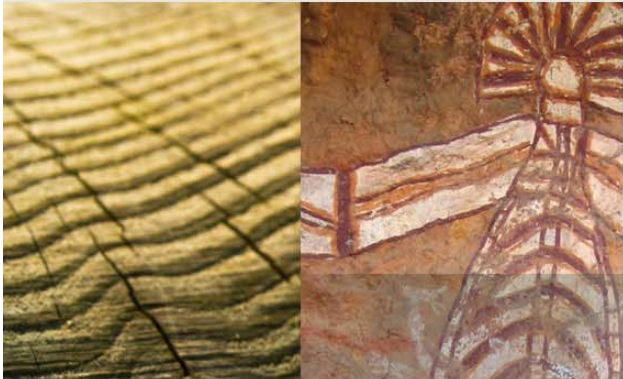


Copyright and data Prof. Anne Fitzgerald (QUT)

Slides available separately



Morning Tea Break



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*

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 you?

- 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)

- 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

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.

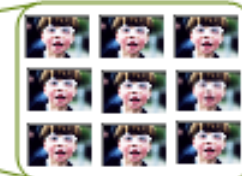
Metadata (2)



Digital Photograph



Digital Photo Album



Collection of Digital Photos



Digital File



Compact Disc



Collection of Digital Files

Metadata (3)

- 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. For example:’

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?

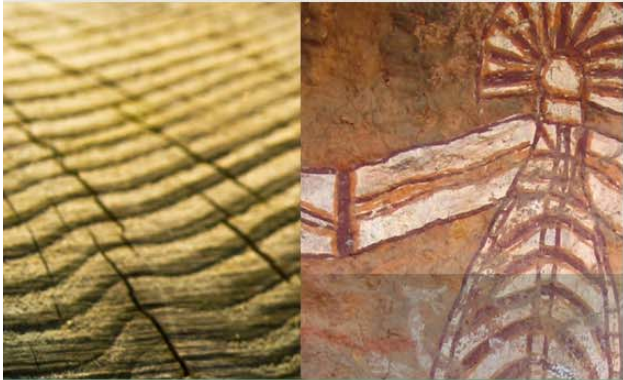
Budget

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



Small group discussion

Discussion topics

- What do *you* need to do in order to comply with the *Code*?
What parts do you (your department, your research group) already have well covered? Where do you need to do some work? Where do you need help?
- (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? De-identification issues.
- Issues around international partnerships, outside funding arrangements (e.g. medical research funded by commercial interests).

Discussion topics (continued)

- 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?
- What can you do to make your data available for re-use? What do you need from your institution to support this?

Thank you for your attention

Questions?

Contact ANDS via:

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

ANDS Resources:

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