



Constructing a Research Project Data Management Plan

Belinda Weaver

Manager, UQ eSpace

[Whose job is it?]

“The immediate and **critical** issue for the stewardship of research data in Australia is the **lack of administrative responsibility** for the task ...”

Sustainability Issues for Australian Research Data: The Report of the Australian eResearch Sustainability Survey Project

Report findings

- Researchers both provide and consume data
- Disincentives to manage data long term
- Little recognition for good data management
- No national data management system
- Infrastructure is decentralised and uncoordinated
- Significant technological challenges
- Need for a coherent stewardship system backed by Government | policy-creators | funding bodies | research institutions

[If we don't manage it ...?]

- Data is lost
- Data becomes unusable
- Data is inaccessible
- Data is wasted or forgotten
- Data might have to be re-collected
- Data may be released improperly

[A Data Management Plan]

- An instrument to help researchers **manage** their data
- Created at **beginning** of research project, updated as necessary
- Captures some technical, access and descriptive **metadata** at the start of a project

[Can you answer ... ?]

- What data have we got?
- Who can use it?
- How can I grant access to others?
- Where is it stored?
- Is it safe and secure?
- What are my legal responsibilities?
- Will it still be usable in five-ten years?
- Do I have to do this to comply with a grant?
- Do I have to do this to comply with legislation?

You'll need to know

http://www.apsr.edu.au/long_lived_collections/kaarin_anstey.pdf

[Components of a plan]

- Originators and owners of the data
- Description of project
- Metadata used (schema, standards)
- Types of data to be collected
- Volume of data (estimate of disc and/or tape storage required)
- Retention requirements
- Format/s of and software used in creation and use of the data
- Access policies and provisions
- Confidentiality requirements
- Storage, preservation and archiving of data

When is the plan needed?

- An **effective Data Management Plan**, based on an understanding of existing database models and data management schemes for the storage and management of data, **should be in place from the conception and commencement of a research project**

Data management plans should cover ...

- how data will be generated or collected
- how data will be integrated and stored in the database
- data ownership and legal controls
- how data will be managed and disseminated
- any regulatory mechanisms that apply to the data
- quality assurance and control measures
- data security measures
- and ...

Data management plans should cover ...

- how access will be provided
- how data will be disseminated
- the timeline(s) for distribution of data
- roles and responsibilities of researchers / database managers
- database infrastructure
- sustainability of data
- expenditures on equipment, storage, staff ...

Stakeholders

- the creators of the data
- the collectors of the data
- the compilers of the data
- the consumers or end users of the data who will seek access to the data
- the parties funding the data collection
- the managers of the database or the repository housing the data

[More questions]

At the outset, identify

- the **scope** of the data required
- what is to be **done** with the data?
- how individual researchers will **contribute** data to the project
- what incentives may be needed for **compliance**

[**And these too ...**]

- information technology requirements
- the kinds of data that will be collected / generated
- what data is required from other sources
- any special practices in a research discipline
- who will be contributing to data collection
- any contractual arrangements or agreements

[The database itself]

- where is the data to be stored?
- who is responsible for the data in the database?
- arrangements for the integration of data that is collected or generated by the research project (original data)
- arrangements for the integration of data that is acquired from other parties (non-original data)

Who owns the data?

- the person who collects, creates or generates the data (the researcher) ?
- the employer of the researcher, under the terms of the researcher's employment contract ?
- the funder of the research, under the terms of the funding agreement ?
- the database owner or provider, if ownership in the data is assigned upon inclusion in the database ?

[Data management]

Data must be managed in accordance with

- legislative
- administrative
- contractual
- legal requirements

[Balancing acts]

Legislative requirements must balance two public interests:

- protecting sensitive, personal and health information
- providing wide access to government and to the public that has funded the research

Researchers must comply with

- requirements stipulated by research project funders
- licensing agreements concerning the release of data and datasets

[Other considerations]

- cultural protocols
- established practices and procedures within a specific discipline

Contracts that may apply

- confidentiality agreements
- copyright assignments and licences, especially with third party-obtained data
- deposit agreements, where data is to be included in a database or digital repository
- access agreements to provide users with access, subject to certain conditions

[**Big stick ?**]

The DMP should address

- statutory licences and open source licences such as Creative Commons licences
- ensuring researchers are not breaching licence terms

[**Quality assurance**]

Applies to

- The data itself
- The way the data is managed

Quality assurance involves

- the **data standards** and quality assurance arrangements for the research project
- how quality is to be **assessed**
- whether there is a **minimum standard** of quality
- whether data must be **authenticated** or **verified** by other researchers or industry professionals before being publicly released
- provisions for the regular **auditing** of the various parties' compliance with data management and access requirements

[Data security involves]

- the **type** of data
- **why** it needs to be protected
 - confidentiality or privacy ?
 - needing de-identification ?
 - copyright and licensing ?
 - cultural protocols ?
 - patent application pending ?
 - competitiveness ?
- the **level** of protection required

[Data security responsibilities]

- researchers
- curators and database managers
- other relevant parties ?

[Retention and disposal]

- Components of this section
 - Description
 - Specific categories
 - Retention period
 - Storage/custody
- For example
 - Research data of archival significance / not lodged in national or international repository / should be retained permanently / in central store
 - Research data / involving psychological testing or intervention with adults / should be retained for 7 years after publication of results / in secure, accessible form in research location

Data not all created equal

- Not all data from a researcher or research project will be managed in the same way – some will need greater (or lesser) levels of management

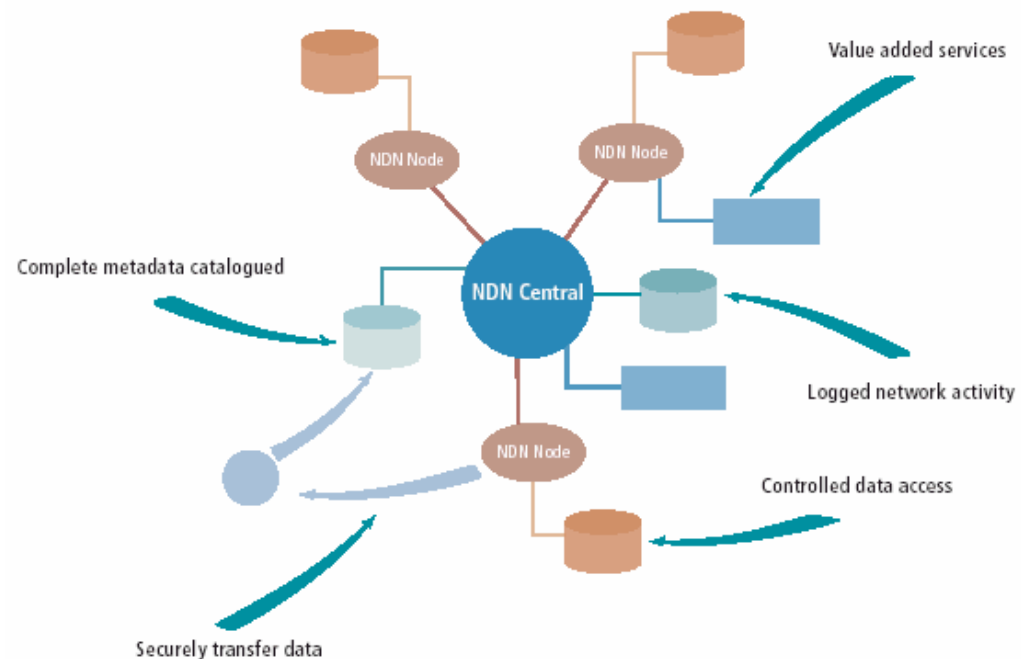
[Why manage and protect?]

- Data from publicly funded research is a valuable **national asset** and should be made freely available
- Data re-use can lead to new areas of research

National Data Network

A distributed library of data holdings relevant to policy analysis and research.

A range of web based services, protocols, procedures, and tools to manage and share data in a way that ensures security and privacy.



[Good management]

- Supports individual or group research
- Optimises the investment
- Allows reuse and recombination
- Helps manage compliance, legal and financial risks, e.g. IP, privacy concerns
- Allows the dissemination of results to the wider community
- Maintains the scholarly record

[Data management principles]

- Centralised
- Secure
- Share-able
- Preserved and curated
- Sustainable
- Use of Open Standards formats
- Management of the data's life cycle

[The data life cycle]

- Created
- Stored
- Used
- Archived
- Migrated to new format?
- Merged?
- Today, tomorrow and beyond ...

eResearch life cycle

- Data and information **creation**
- **Preservation** and **curation** of data for long-term use
- **Search** and navigation
- Virtual research communities to enable **collaboration** across institutional, disciplinary and national boundaries
- Networks, computers and data storage to support **distributed access** to resources
- Authentication, authorisation and accounting middleware and digital rights management to ensure that access is **safe, secure** and **legally compliant** ...

[Expenditures]

- Costs of
 - Creating
 - Acquiring
 - Managing compliance
 - Long term maintenance and preservation

NCRIS Summary – Institutions should:

- Develop and implement a policy on data ownership
- Provide guidelines to researchers on ownership, what to keep and researcher responsibilities
- Maintain durable records on what research data has been held and ensure that research data is under the control of the institution where the work was performed
- Provide secure systems for holding data and for granting access to that data

NCRIS Summary – Researchers should:

- Determine what data to keep, considering research community practice and any project or legal requirements
- Ensure research data is retained (for at least 5 years from publication of results) using institutionally provided mechanisms
- Ensure at the end of employment (for whatever reason) data retention passes to the institution
- Maintain confidentiality where it exists

[Help @ UQ]

- **UQ eSpace repository**
 - Publications
 - Research data
 - Metadata about research data held elsewhere
 - Advice on metadata / data storage
 - Object history logs, AONS, TDR
- Any staff member can deposit in either
 - Open collections
 - Closed collections

[Have your say]

- Complete the UQ data management survey
- Nominate for an eResearch focus group

[Components of a plan]

- Originators and owners of the data
- Description of project
- Metadata used (schema, standards)
- Types of data to be collected
- Volume of data (estimate of disc and/or tape storage required)
- Retention requirements
- Format/s of and software used in creation and use of the data
- Access policies and provisions
- Confidentiality requirements
- Storage, preservation and archiving of data

[UQ eSpace]

W: <http://espace.uq.edu.au/>

Contact:

Belinda Weaver, Manager

T: +617 3365 8281

E: espace@library.uq.edu.au

Upcoming event

Clever Collections: a national showcase of technical innovations for digital collections

The University of Melbourne, Melbourne
November 28 - 29, 2007

<http://www.apsr.edu.au/clevercollections/>



Useful links

- [Guidance on data management](#) and other documents - Economic and Social Data Service, UK
- [Guide to Social Science Data Preparation and Archiving](#) - Inter-university Consortium for Political and Social Research (ICPSR)
- [Guide to preparation of Data Management Plans](#) – Geoscience Australia
- [Rural Economy Land Use Guidance on Data Management](#) – excellent publications on data management
- [Ecology Centre](#), UQ – data management and access policies [maybe a case study?]