Data Management Plans

Friday, 27 January 2017

11:08 AM

* As we have seen, data management covers a range of activities other than its direct use:
  + Managing backups
  + Archiving for long-term preservation
  + Data sharing
  + Ensuring security of confidential data
* Data management requires planning; if done at last minute, it typically is not thought through, becoming inefficient, time-consuming and error-prone.
* Taking the time to prepare a data management plan at the start of a research project has major benefits in terms of improving efficiency, security, data quality and exposure.

* Data management plan is required to comply with the ***Australian Code for the Responsible Conduct of Research***.
* Also is a requirement to receive funding from funding bodies such as the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC).

* As developed at most research institutions, data management plans act as a check-list so that research teams can consider all the relevant issues that are likely to arise during the course of, or subsequent to, a project and to put the necessary arrangements in place for managing their data.

Some institutions have developed automated tools to assist with all or part of the process, but as often as not, even these serve to act as a checklist.

* Typically, data management plans describe:
  + What data will be created.
  + What policies will apply to the data.
  + Who will own and have access to the data.
  + What data management practices will be used.
  + What facilities and equipment will be required.
  + Who will be responsible for each of these activities.

* In Australia, and internationally, governments and major research funding bodies have established organisations to assist researchers and research institutions in becoming more aware of the need to better manage research data and provide them not just with the tools to do so, but to establish international level infrastructure to support best practice. Organisations include:
  + Australian National Data Service (ANDS)
  + Digital Curation Centre (DCC) - UK
  + DataOne - USA

* Although each institution will need to produce it's own data management plan checklist to reflect its specific operational environment, ANDS suggessts that any **data management plan should include at minimum**:
  + Survey of existing data
    - What existing data will need to be manage?
  + Data to be created
    - What data will your project create?
  + Data owners and stakeholders
    - Who will own the data created and who would be interested in it?
  + File formats
    - What file formats will you use for your data?
  + Metadata
    - What metadata will you keep?
    - What format or standard will you follow?
  + Access and security
    - Who will have access to your data?
    - If sensitive, how will you protect from unauthorised access?
  + Data organisation
    - How will you name your data files?
    - How will you organise your data into folders?
    - How will you mange 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?
  + Storage
    - Where will your data be stored?
    - Who will pay for the hardware?
    - Who will manage it?
  + Backups
    - Probably the most important item on the list, hard drives on PCs fail regularly.
    - Must have a credible backup strategy of regular backups, and of course then it must be followed.
    - Consider including an off-site backup so that your data will not be lost if your building burns down.
    - Rather than relying on memory, consider an automated backup process.
  + Bibliography management
    - What bibliography management tools will you use?
    - How will you share references with the other members of your group?
  + Data sharing, publishing and archiving
    - What data will you share with others?
    - How will you do this?
  + Destruction
    - What data will you destroy, when and how?
  + Responsibilities
    - Who will be responsible for each of the items in this plan?
  + Budget
    - What will the plan cost?
    - Possible costs include hardware for backups, research assistant time for data curation, metadata creation and archiving etc.

**ANDS Guide :: Data Management Plans**

* A data management plan is a document that describes:
  + What data will be created
  + What policies will apply to the data
  + Who will own and have access to the data
  + What data management practices will be used
  + What facilities and equipment will be required
  + Who will be responsible for each of these activities.

**Why do I need a data management plan?**

* Improvements to efficiency, protection, quality and exposure.

* Data management in some form is an unavoidable consequence of working with data.
* Typically data management is done at the last minute and using the first method that comes to mind.
* This approach is usually time-consuming and error-prone.
* Taking time at the start of a research project to put in place robust, easy-to-use data management procedures will usually pay off several times over in the later stages of the project.
* Inadequate data management can also lead to catastrophes like the loss of data or violation of people's privacy.

* Basic data management is required by the Australian Code for the Responsible Conduct of Research.

**Mercury :: Research Data Management In Practice**

* Reuse and management of research data is becoming important due to lodging publications and data in public repositories.

* Publication and re-usability of research data bring great benefits such as research and researcher effectiveness, enhancing the reputation of researchers and institutions, meeting obligations to funders, and compliance with Open Access agendas.
* To achieve this, it is critical that research data is properly management from the pre-research planning stages, through to post research completion.

* ANDS has commissioned "Research and Data Management Practice Guide" as a practical starting point that focuses on the 'Why' and 'How' of a good data and risk management, with plenty of references for further reading for readers who need more detail.
* It seems likely that Australia's research funders will follow their international counterparts in requiring a research data management plan as a part of the research process.

* No single person or even business unit is responsible for all aspects of research data management and that a collaborative approach is required.
* In all cases this will involve the researcher/data creator.

**Why Manage Research Data?**

* Research is becoming more data intensive.
* Not all data can be made publicly available, however, most can be shared if appropriate action is taken to ensure that ethical and other concerns are addressed.
* Research outputs and data is now being lodged in public repositories, this is best achieved using standardised approaches to data capture, storage, attribution and metadata.
* Recent studies show that benefits of public sector information, which includes publicly funded research data, outweigh the costs.

* Effective data management plays a vital role in managing research risk.
* All research is subject to a range of data related risks such as data loss or corruption, and privacy or copyright breaches with potential catastrophic impacts.
* Effective research data management can go a long way towards preventing and managing such risks.

* Effective research data management benefits both researchers and institutions by minimising the risk of loss or deterioration of research findings (including the data).

**Steps in Research Data Management**

Machine generated alternative text:
Pre Research 
Ma n ae.m.nt 
FOure I Key steps in Research Data Managerænt 
Research 
post Research 
publish Data 
Resist. 
(ROAI 
Ongoing 
Mon it or". 
primarily 

*Research Data Curation*

* With traditional publication, most curation activities occur at the end of the research cycle.
* In contrast, digital curation of data is characterised by activities planned for, from the outset and occurring throughout the data lifecycle.

Machine generated alternative text:
Private Research 
Domain 
Laboratory 
Information 
S e Sea 
M anagement 
S stem 
May link to data obiects stored in 
Research Data 
Store 
This domain involves the core research 
team IS they undertake the research. 
usually within a single institution. 
Access is often tightly controlled as 
hypotheses and analyses are developed. 
Migration Pr-ocess 
Auth Orised by 
research team 
leader 
Performed by 
research team IT 
support 
Involves Object 
selection. 
m igration. 
assignment of 
access controls, 
augmentation of 
Collabo ration 
Curation Boundary 
Shared Research 
Domain 
Collaboration 
Support System 
(Plane. TWIki, 
Sharepoint) 
May link to data objects stored in 
ta Store 
This domain involves researchers 
Outside the Core team as they 
collaborate with colleagues, often 
across institutions. Access is more 
open, but not everything is shared. 
Migration Process 
Authorised by 
research team 
leader 
research team 
supporülR staff 
Invoh•es Object 
selection. migration. 
a s Of 
persistent identifier, 
creation of fixity 
metadata. 
augmentation of 
other metadaca 
Publication 
Curation Boundary 
Public Domain 
Institutional 
document 
May link cc data objects stored in 
public Data 
Store p os 
This domain involves the public sphere 
(publication in the sense of making 
public). Access will usually be open to 
Version I A. 
Figure 2 Research Data Curation Continuum. Source: htto://ands.orq.au/quides/curation.continuum.html 

* Digital data curation requires the capturing of rich metadata, and depositing in appropriate formats into well-managed stores as defined in the Research Data Management Plan.

*Pre Research*

*Research Data Management Framework*

Machine generated alternative text:
Data 
management 
framework 
Policies and 
procedures 
infrastructure 
Support 
services 
Metadata 
management 

* The research institution is responsible for providing an adequate Research Data Management Framework, providing the basic elements required within an institutional context to support effective data management. These elements comprise of four categories:
  + **Institutional policy and procedures**
    - These should be up to date, addressing data-related issues, and be publicised to all those who have a data creation and/or management role.
  + **IT Infrastructure**
    - The hardware, software and other facilities which underpin data-related activities, as we ll as identity management and access control
  + **Support Services**
    - People and other means of providing advice and support, such as web-pages.
  + **Metadata management**
    - So that data records can be used for both internal and external purposes.
* The researcher can benefit from an understanding of the policies, facilities and services of the institution.
* For more information :: <http://ands.org.au/guides/dmframework/data-management-framework.html>

*Risk Management Plan*

Machine generated alternative text:
Risk management 
plan 
Identify 
Assess 
Mitigate 
Monitor 

* Risk management is the process of identifying, assessing and responding to risks and communicating the outcomes of the processes in a timely manner.
* Developing a risk management plan is important to avoid preventable losses associated with research data management.
* Risk can arise from
  + Data loss
  + Corruption
  + Under-utilisation of research outcomes
  + Breaches of privacy, confidentiality or copyright.
  + Among others.
* A risk management plan identifies and **assesses risk based on likelihood of occurrence and respective impact**.
* It identifies effective mitigation measures based on risk criticality, which are allocated to specific people, and monitored throughout the research project.

* An effective risk management plan adheres to the international risk management standard AS/NZS ISO 31000:2009.
* An example of a guide that buids on this standard is provided by the Queensland government:: <https://www.treasury.qld.gov.au/publications-resources/risk-management-guide/guide-to-risk-management.pdf>

* An example of an institutional research data risk management plan, including illustrations of risk mitigation strategies, by the Alfred campus in Melbourne can be found at :: <https://www.alfredhealth.org.au/research>.

*Research Data Management Plan*

Machine generated alternative text:
Data 
management 
plan 
Organisation 
Metadata 
Backups 
Archiving 
Version 
control 
Sharing 
Security 
Governance 

* Development of a Research Data Management Plan is a critical aspect of the pre-research stage.
* It lays out :
  + What data will be created.
  + What policies will apply to the data.
  + Who will own and have access to the data.
  + What research data management practices will be used.
  + What facilities and equipment will be required.
  + Who will be responsible for each of these activities.

* It will include activities such as :
  + Data organisation and storage.
  + Metadata standards and guidelines.
  + Backups.
  + Archiving for long-term preservation.
  + Version control and derived data products.
  + Data sharing or publishing intentions, including licensing through AusGOAL (<http://www.ausgoal.gov.au/>).
  + Ensuring security of confidential data.
  + Data synchronisation.
  + Governance, roles and responsibilities.

* The plan usually defines all research data management related activities during and subsequent to the research activity.

* The ANDS Data Management Planning and Awareness Guide provides a detailed checklist of topics to be covered, as well as links to some sample plans :: <http://www.ands.org.au/resource/data-management-planning.html>.

*Ethical Clearance*

Machine generated alternative text:
Ethical clearance 
Informed 
consent 
Anonymise 
Access control 
Licensing 

* Sharing of research data that relates to people can often be achieved using a combination of obtaining consent, anonymising data and regulating data access.

* Research data even sensitive and confidential data can be shared ethically and legally if researchers pay attention, from the beginning of research to four important aspects:
  + Including provision for data sharing when gaining informed consent.
  + Protecting people's identities by anonymising data where needed.
  + Considering controlling access to data.
  + Applying appropriate licence.
* More information :: <http://ands.org.au/guides/ethics-working-level.html>

*Training and Induction*

Machine generated alternative text:
Training and 
induction 
All staff 
Induction 
Refresher 

* Effective implementation of an institutional research data management framework requires that all institutional staff receive adequate training.
* All staff involved in the research project would benefit from up to date training.

* Furthermore, for project specific issues such as risk management and the implementation of the research data management plan, all researchers and support staff will need to be inducted.
* Inductions take place not only at project commencement, but also for any personnel coming on board during the project.
* For longer projects, refresher training and inductions should be considered.

* ANDS provides guides, support and training services online, as well as webinars and custom courses.
* More information can be found at :: <http://ands.org.au>.

During Research

*Policy Compliance Monitoring*

Machine generated alternative text:
Policy compliance 
monitoring 
Compliance 
Audits 
Corrective 
action 

* Acknowledging the policies and guidelines as defined in the institution's research data management framework at the start of the research project is of little us in itself.
* Demonstrating compliance through review or audit frameworks allows non-compliance to be identified early and corrective action to be taken.
* This allows the organisation to respond to compliance breaches in a systematic rather than ad hoc fashion.

* Policy compliance is normally dealt with as a part of the review of the research data management plan.

*Risk Monitoring and Communication*

Machine generated alternative text:
Risk monitoring 
and 
communication 
Monitor 
Review 
Re-assess 
Communicate 

* Continuous monitoring and review are vital components of an effective risk management process.
* In terms of the research project, review of risk would normally be incorporated into the research data management plan.
* The primary purpose of monitoring and review is to determine whether risks still exist, whether new risks have emerged, and to reassess the risk priorities.

*Research Data Collection and Analysis*

Machine generated alternative text:
Data collection 
and analysis 
Sharing with 
collaborators 
Derivation 
Curation 

* Through the research cycle, data will be collected; data analysis will generate derived data and in many cases data will be shared between researchers and institutions for collaboration purposes.

* Data at this point may be made available to collaborators, as specified in the research data management plan.

*Metadata Generation*

Machine generated alternative text:
Metadata 
generation 
Rich metadata 
from the 
outset 
At collection 
and object 
level 

* Data discovery and access is dependent on the availability of rich metadata.
* Metadata is collected at both collection and object level and can be stored separately, or embedded in the data collection.
* Collection level metadata is generated by the researcher as a part of the research process and supplemented by object level metadata for data publication, in most cases, by librarians or other data professionals.

* Good metadata creation can be supported by tools designed to simplify metadata input and to enhance interoperability.

* Metadata is also useful in tracking the history of derived data products.
* Further reading :: <http://www.ands.org.au/metadata/index.html>

*Storage and Access*

Machine generated alternative text:
• 
Storage and 
access 
Considered 
choice 
Metadata rich 
Curated 
Wide 
community 
scope 

* The choices made regarding the approach to data storage have implications for cost, security and further access.
* It is an institutional responsibility to ensure that adequate and appropriate storage facilities are available.

* The goals of "reusing and sharing data more often" are met by storage solutions which make data discoverable and accessible over the long term, which means the tendency should be towards more metadata rich, curated stores with a wide community scope.
* More information in the ANDS data storage guide :: <http://ands.org.au/guides/storage.html>

Post Research

*Publish Research Data*

Machine generated alternative text:
Publish data 
Well- 
described 
Citable 
Discoverable 
Re-usable 
Ethics & 
copyright 

* There is an increasing expectation that the outputs of publicly funded research, including data, will be made available for others to use.
* The means published data should be well-described (metadata), citable, discoverable and re-usable whenever possible.

* Potential refusers of research data need to have clear guidance about what they can and cannot do with the data: this is normally achieved via a licence.
* ANDS supports the Australian Governments Open Access and Licensing Framework (AusGOAL) which provides support and guidance to facilitate open access to publicly funded information.
* More information on copyright and licensing can be found at :: <http://ands.org.au/publishing/licensing.html>

* Research data can be published in the form of collection descriptions, citable and online accessible data elements or citable other objects such as web services, APIs, or concept definitions.
* Sometimes just the existence of data collections is published; this occurs when the data cannot be access or accessed under strict conditions.
* Online data publication for download or web-service access is desirable for those kinds of data which are not restricted.

* ANDS provides an overview of the many aspects of data publication here :: <http://www.ands.org.au/publishing/>

*Register Research Data*

Machine generated alternative text:
Register data 
Research Data 
Australia 
Cite my data 
DO's 
Identifiers 

* The process of registration and citation depends on the type of data published and how it is published.
* Data can be considered as 'published' when it is generally discoverable.
* ANDS recommends that registration with Research Data Australia is one form of data publication.