DATA management, Monash policy and planning

Cathrine Harboe-Ree, 4 May 2007
Why manage data?

- Support individual or group research
- Optimise the investment
- Allow reuse
- Manage compliance, legal and financial risks
  - Funding, IP, privacy, etc
- Disseminate results to wider community
A “P” view of the world

• People
• Partnerships
• Purpose / Principles
• Policies
• Practice
• Places (to put stuff)
• Publication
In the beginning, there were …

Researchers

• Wanting to publish research data
• Wanting to collaborate more effectively
• Needing enduring access to data
• Needing to be able to re-use data
• Not understanding the data management environment, and …
• Sometimes wanting to be left alone
There was an ITS Division …

• Needing to provide storage solutions
• Needing to find archiving solutions
• Worried about how to pay for storage
• Needing resolution of access and authorisation issues
• Being expected to help the University meet its regulatory obligations
There were librarians …

• Aware of data management issues but short on solutions
• Interested in whole of data life cycle (research to publication)
• Experimenting with / establishing repositories
• Building expertise in metadata, persistent identifiers, standards, copyright, IP, version control, etc
There were senior university people …

• Wanting appropriate storage and data management regimes
• Extremely interested in e-research capability
• Watching the synchrotron with its attendant data storage and management requirements rapidly become a reality
And the Government was …

- Funding research infrastructure through
  - Systemic Information Infrastructure
    > APSR, ARROW, DART, ARCHER, MAMS, RUBRIC, etc
- Acknowledging the importance of data management
- Pursuing policy through
  - e-Research Coordinating Committee
  - PMSEIC Working Group on Data for Science
  - NCRIS KCA 16 Platforms for Collaboration
  - RQF
  - Accessibility Framework
Monash’s response

• Established a forum, through the E-Research Steering Committee, for discussing the issues
• Developed a proposal for storing large volumes of data
  – LaRDS (Large Research Data Store)
• Developed a data management policy …
• … building on ARROW, DART and ARCHER
The forum

• Comprises
  – ITS
  – Library
  – E-Research Centre
  – Records and Archives
  – Researchers (through consultative processes)

• Consults with
  – University Solicitor
  – Research Office
LaRDS

• Addresses institutional and researcher needs
• Formulates a set of principles to guide cost modelling and sustainable funding options
• Assumes commitment to storage in perpetuity (“as long as required”)
• Adopts a central storage model …
  – Centrally funded basic allowance, plus
  – Directly charged excess allowance
• … in parallel with decentralised storage
Monash information management principles

• Corporate importance – information is valued
• Information sources reliable and findable
• User centred
• Available as appropriate
• Staff and student skills developed
• Supports productivity and efficiency
• Complies with statutory requirements
• Information and systems are trustworthy
• Right information retained and disposed of
• Supported by information and technology
Specific Data Management Policy principles

• Centralised data storage
• Security of data
• Sharing research data
• Preservation when required
• Sustainability
• Information lifecycle management
• Use of Open Standard formats
Data Management Plan

• Instrument to help researchers manage their data
• Completed at beginning of research project, updated as necessary
  – May become mandatory
• Captures some technical, access and descriptive metadata at the beginning of a research project
Data Management Plan components

• 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 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
Data Management Plan requirements

• The plan process must not be onerous
• It must have visible benefits
• It must assist in providing complete research data solutions for researchers
Retention and disposal guidelines

• Components
  – Description
  – Specific categories
  – Retention period
  – Storage/custody

• For example
  1. Research data of archival significance / not lodged in national or international repository / should be retained permanently / in central store
  2. 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
Retention and disposal guidelines – cont.

• Curation continuum
  – Researcher → LaRDS → ARROW
  – From light data management regime to extensive data management regime

• Not all data from a researcher or research project will be managed in the same way – some will need greater levels of management
Repository continuum dimensions

- less metadata ← more metadata
- less harvesting ← more harvesting
- less curation ← more curation
- less preservation ← more preservation
- more items ← less items
- less organisation ← more organisation
- researcher manages ← university manages
- items continually updated ← items static/snapshot
Current state of play

• LaRDS established
• Data Management Policy accepted in draft form
• Retention and disposal guidelines completed
• Testing underway of Data Management Plan with selected researchers
  – 2 recognising need and reasonably well informed
  – 1 sceptical of need but working on research considered significant by Monash
• ITS establishing e-Research support group
• Library has staff committed to the project
• ARCHER and ARROW investigating tools, requirements and management of data and datasets related to published material
See also

- Prairie Cluster Prototype Long-Term Ecological Monitoring Program
  - [http://nature.nps.gov/im/units/prcl/pdf/PrairieClusterDMP.pdf](http://nature.nps.gov/im/units/prcl/pdf/PrairieClusterDMP.pdf)
- Astro-D Guest Observer Facility
- Rural Economy Land Use Guidance on Data Management
  - [http://relu.ac.ak/about/data.htm](http://relu.ac.ak/about/data.htm)
Some terms

- **Data management** ≠ data storage
- **Data curation** ≠ archiving and preservation
- **Data stewardship** might be a good term
Back to the “P” view of the world

• **People** - awareness, engagement, skills
• **Partnerships** - involving the right people
• **Purpose** - objectives
• **Policies** - as in data management, storage, principles
• **Practice** - support services and researchers
• **Places (to put stuff)** - storage, institutional and national
• **Publication** – access, links, harvesting
Thankyou