Growing an Institution’s Research Data Management Capability through Strategic Investments in Infrastructure

Anthony Beitz
General Manager, Monash e-Research Centre
15 January 2013
Look, half the work is done!
All you need to do is fill in the top part so we can legally say the bottom part.

CHOCOLATE CO-OP INC.
RESEARCH AND DEVELOPMENT

DATA:

CONCLUSION:
EATING CHOCOLATE WILL MAKE YOU LOOK YOUNGER AND THINNER
Why are Researchers caring more about their data?

- Achieving new research outcomes and higher research impact
- Safe guarding their data
- Dealing with the data deluge
- Facilitating collaboration and enabling reuse
- Validating their research
- Conformance
  - funding body policies, institutional policies, legal obligations, codes of conduct, and culture
Why are institutions caring more about their Researchers’ data?

- Enabling leading-edge research
- Achieving the University’s research aspirations
- Increasing research outcomes and impact
- Reducing legal risk
- Attracting the best researchers
- Attracting additional research income
Monash University

- Large
  - 3,800+ academic staff, 4,200+ higher degree by research students
- Diverse
  - 10 faculties, many disciplines, 6 Victorian and 3 overseas campuses
- Research intensive
  - $AU264 million in externally funded research income in 2010
- Considers research data as the data, records, files or other evidence, irrespective of their content or form that comprise a research project’s observations, findings or outcomes.
- Strong history of leadership in e-research & scholarly communications, including research data management
Statement of intent

Monash University recognises research data that is better managed, more discoverable and available for re-use will contribute to increased research impact, enhanced research practice (including collaboration) and improved education outcomes. The University aims to maintain its national leadership role around research data management and to fulfil compliance requirements and community expectations.

All members of the Monash University community share responsibility to improve research data management in a coordinated and integrated way.

This strategy supports the research, education and professional services strategies developed as part of the Monash Futures program.
Australian environment

- **Australian Code for Responsible Conduct of Research (2007)** outlines data responsibilities of researchers and their institutions
- Strong publication repositories community
- Limited history of sector/discipline archives
- Data management plans not compulsory
- $AU72M investment into date re-use infrastructure and capability building - Australian National Data Service (ANDS)
- $AU50M investment into national research data storage – Research Data Storage Infrastructure (RDSI)
- $AU47M investment into collaborative research infrastructure - National eResearch Collaboration Tools and Resources (NeCTAR)
- Outputs from National Health and Medical Research Council (NHMRC) funded projects must be deposited into an institutional repository within 12 months of publication
- Publications from Australian Research Council (ARC) funded projects must be deposited into an open access institutional repository within 12 months of publication. The ARC strongly encourages the deposition of project data into an appropriate publicly accessible subject and/or institutional repository.
The Australian national e-infrastructure platform

Slide by Dr. Rhys Francis  
(Executive Director, Australian eResearch Infrastructure Council (AeRIC))
Research Data Management Lifecycle

- **Conceive**
- **Design**
- **Experiment**
- **Analyse**
- **Collaborate**
- **Publish**
- **Expose**

**Data Management Planning**

**Research Data Management Platform**

- National Repository or Institutional Repository or Electronic Journal or Community Repository

---

Growing an Institution’s RDM Capability through Strategic Investments in Infrastructure
Growing an Institution’s RDM Capability through Strategic Investments in Infrastructure
Research Data Management Landscape
About Researchers

- Focused on research outcomes
- Work in an interpretive mode
  - Characteristics include: an iterative process; open-ended; and thrive on ambiguity
- Their requirements and goals may change over time
- Generally more loyal to research community than their institution
- May require an ICT capability for only a short period of time
- Very resourceful and driven
Selecting a RDM Platform Solution

• Must fit in with researchers’ tools, instruments, environment, and workflows

*No single institutional RDM platform will fit all researchers’ needs*

• Must be a good cultural fit and should facilitate external collaboration

• Shouldn’t re-invent the wheel

• Developing a new product may be expensive, may be costly to maintain, and may split researchers from their community

*Adopt/Adapt/Develop - if a research community already has a RDM platform (or there is an emerging one) and it meets the needs of your researchers in that community, then adopt it.*
Selecting a RDM Platform Solution (2)

- Each University has many research disciplines, so, unrealistic to support a RDM platform for each discipline.
- Not all disciplines need or are ready for a RDM platform.

Must be strategic in selecting which discipline-specific RDM platforms to support and focus on ready/willing/able.

Offer a versatile solution for those in the long tail.

- Varying approaches to research make it difficult to find generic terms to describe the management of data.

Provide one or more versatile RDM platforms.
Selecting a RDM Platform Solution (3)

- Many researchers are only looking for a safe holding for their data which they can easily share with colleagues.
- Researchers don’t want to waste time in thinking about which data to keep and what to throw out.

Institutions should at least provide their researchers with access to a very large and reliable research data store.
Institutional Data Collection

- Generally speaking, it is a collection of an institution’s publicly listed digital data collections. Each registered collection has only a high-level description. Associated data is stored in another repository.

- Covers a variety of research disciplines.

- Main purpose is to aid in the assessment of an institution’s research output, which may affect research funding. It can also be used to show case an institution’s researchers.

- Not really for data reuse, as researchers more likely to browse comprehensive discipline-specific data collections and will generally search on attributes other than those provided in high-level descriptive metadata.
Deployment Considerations

- Many infrastructure hosting options, including: department, institution, state, national, and cloud
- Cost/Benefit of an institutional deployment versus a state/national/international
- Differing ethical, legal, and security obligations; RDM policies & strategies; capabilities; and sustainability models
- For very large data sets, consider co-locating instruments, data storage and management, compute, analysis tools, and visualisation.

_Institutions must be flexible in their approach to provisioning RDM infrastructure_
Supporting Technical RDM Services

- eResearch services are generally bespoke systems addressing unique needs
- There aren’t commercial vendors which the support call can be escalated to, nor are there large online communities who self support
- Broad and diverse ecosystem of users, providers, and intermediary organisations which need to be engaged
- Supporting eResearch is a different paradigm for IT Directors and CIOs and this capability usually doesn’t exist within their division

Form a separate specialised support group for RDM infrastructure

Adapted from a CAUDIT slide by Richard Northam (CEO CAUDIT)
Monash’s Technical RDM Infrastructure

- Website
- Large research data store
- Monash University Research Repository
- Range of RDM platforms
- Versatile RDM platform
- Researcher information systems
- Shibboleth IDP
- Resourced through a partnership between Monash’s Library, central IT division, and e-Research Centre
Monash RDM Website

Monash University is leading efforts to improve the management of research data. Well-managed research data is more discoverable and available for re-use, and contributes to increased research impact, enhanced research practice (including collaboration) and improved education outcomes.

Managing research data - a roadmap

The Monash University Research Data Management Strategy and Strategic Plan 2012-2015 was publicly released in April 2012. It outlines future initiatives at Monash.
Monash’s Research Data Store – LaRDS (Large Research Data Store)

- 4 Petabytes
- Market and vault capability
- Secure, reliable, and backed up
  - 4 copies of any file over 2 data centres
- No cost, for most research uses
- LaRDS is a common infrastructure supporting a multiplicity of applications and access modes
- To be augmented in 2013 by national research data storage
Monash University Research Repository

- 71,450 items
  - Publications required for Excellence in Research Australia (ERA)
  - Journal collections
  - Research data
  - Newspapers – Lot’s Wife
  - Monash University patents
  - PhD theses
  - Faculty of Business and Economics working papers
RDM Platforms @ Monash
MyTARIDS

- Solving the problem of archiving, accessing, and citing raw bio science data
- Easy sharing, access and publication of terabyte datasets
- More than 2 million files, 20 terabytes collected automatically for access - and growing
- 18 code contributors (python, [github.com/mytardis](https://github.com/mytardis))

Slide by Steve Androulakis
(Data Consultant, MeRC - steve.androulakis@gmail.com)
Protein Crystallography Research Data and Metadata Workflow

Growing an Institution’s RDM Capability through Strategic Investments in Infrastructure
MyTardis - Connecting scientific instrument data to people

- 11 deployments in Australia for several types of data
- New project ‘Bioscience Data Platform’ bringing MyTardis into the cloud with links to supercomputers and rich data publication platform

Slide by Steve Androulakis
(Data Consultant, MeRC - steve.androulakis@gmail.com)

bioscience-data-platform.posterous.com
Growing an Institution’s RDM Capability through Strategic Investments in Infrastructure

Slide by Dr. Wojtek Gosciniski
(External Collaborations Manager, MeRC)

15 January 2013
Slide by Dr. Wojtek Gosinski
(External Collaborations Manager, MeRC)
Characterisation Virtual Lab Desktop

Slide by Dr. Wojtek Goscinski (External Collaborations Manager, MeRC)
Roles of the RDM Capability Partners

- **Library**
  - Leadership – steering committee, policy, and strategy
  - Communication and awareness raising
  - Skills development – Researchers and Library staff
  - Provide advice to Researchers (direct and via the website)
  - Lead data planning
  - Manual curation of strategically important data collections

- **MeRC (Monash e-Research Centre)**
  - Enable innovation in RDM technical infrastructure
  - Provide advice to Researchers (technical in nature)

- **eSolutions (central IT Division)**
  - Provides Platforms-as-a-Service, Software-as-a Services, software development capability, and technical support
Monash’s RDM Strategy (2012 – 2015)

Excellence and impact
- More research data discoverable and available for re-use
- Re-use of Monash data contributes to formal and informal measures of research quality and impact
- Systems and policies help make research data available more quickly and easily

World class infrastructure
- Systems and facilities that support data management are expanding and improving
- Researchers make more use of these systems and facilities
- A range of institutional and discipline-specific needs are catered for
- Local infrastructure leverages national and international services and facilities
- Infrastructure supports the management of data, regardless of format

Skills and knowledge
- Researchers have the knowledge and skills they need to manage data well, and understand the benefits of making data discoverable and available for re-use
- Data management skills are seen as essential for research and transferable to other workplaces
- Research data contributes to the educational outcomes of students from an early stage in their academic career
- Data management professionals have career paths and development opportunities
- Professional development opportunities meet the needs of researchers from different disciplines and at different career stages

Integrity and professionalism
- Managing data well is seen as a key part of research integrity and professional practice
- Compliance with Section 2 of the Code for Responsible Conduct of Research is improved
- All Monash researchers understand their obligations and take practical steps - as individuals and teams - to improve how research data is managed
- Research data management advisory and technical services are increasingly coordinated and integrated

Leadership and collaboration
- Monash University leads and actively participates in global, national and regional research data initiatives
- The University is regarded as a partner of choice for collaborative work in this area
- Data management technologies developed at Monash University are successfully adopted by other organisations
- Monash University is seen by other institutions as an authoritative source of information and advice
Conclusion

- At the very least, an institution should provide its researchers with access to a large and reliable research digital data store.

- Research institutions should expect that they will need to engage with a range of RDM platforms to accommodate their researchers.

- Adopt/Adapt/Develop

- Must be strategic in selecting which discipline-specific RDM platforms to support and focus on ready/willing/able.

- Offer a versatile solution/s for those in the long tail.

- Institutions must be flexible in their approach to provisioning and supporting RDM infrastructure.
Acknowledgements

- Dr. Rhys Francis – Executive Director, Australian eResearch Infrastructure Council (AeRIC)
- Dr. Richard Northam – CEO, CAUDIT
- Prof. Paul Bonnington – Director, Monash e-Research Centre (MeRC)
- Wilna Macmillan – Director, Client Services, Monash Library
- Dr. Wojtek Goscinski – External Collaborations Manager, MeRC
- Steve Androulakis – Data Consultant, MeRC
- Sam Searle – eResearch Specialist, Research Data Services Unit, Griffith University
Contact:

Anthony Beitz
General Manager, Monash e-Research Centre
Phone: +613 9905-8681
Email: Anthony.Beitz@monash.edu
Twitter: twitter.com/a_beitz
LinkedIn: au.linkedin.com/pub/anthony-beitz/1/7a4/78a

Website:

http://www.monash.edu.au/eresearch/
Aspects of this work have been supported by the Australian National Data Service (ANDS)

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative

Australian Government
Department of Innovation
Industry, Science and Research