Towards a Unified University Infrastructure: The Data Management Roll-out at the University of Oxford (Damaro)

Wednesday 16th January, 2013

James A J Wilson
James.wilson@oucs.ox.ac.uk
RDM at the University of Oxford

• Work on scoping university RDM requirements began in 2008
• Several JISC-funded projects
• Loose programme of infrastructure development

Opportunistic → Coordinated

• Principles:
  – researchers need to be at the core of development
  – intra-institutional collaboration amongst service providers
• University RDM Policy now in place
• Working group formed to create and present business case for integrated suite of services
Policy

- Indicates value of research data
- Sets minimum of 3 years retention period for data after publication or public release of the work of the research (but must meet funder requirements if longer)
- Researchers responsible for documenting procedures for collection, storage, use, re-use, access, retention and destruction of research data and records
- University responsible for:
  - Providing access to services and facilities for the storage, backup, deposit and retention of research data and records that allow researchers to meet their requirements under this policy and those of the funders of their research
  - Providing researchers with access to training, support and advice in research data and records management
  - Providing the necessary resources to those operational units charged with the provision of these services, facilities and training
Research data lifecycle

Research idea → data gathering → documentation → curation → access and re-use

Planning → literature/data review → file organisation & local storage → repository storage

Funding bid → analysis & research outputs → deposit → discovery

Tools & Services

DataStage → Databank → DataFinder

External data repositories

DMP Online
Oxford
Welcome to DataBank

DataBank is a repository that will keep data safe in the long term.

You can use DataBank to store, retrieve and publish data at any time, from anywhere on the web. You can accomplish these tools using

The simple web interface of a hosted instance of DataBank, so it this website you are viewing.

From another service, using the DataBank API.

Run your own instance of DataBank.

Using the web interface

To get started, see this document on the DataBank wiki.

The admin credentials used in this demo instance of DataBank are

Username: admin
Password: test

Using the DataBank API

The API provides a powerful web interface to all of DataBank’s features. Documentation on the API

DataBank API documentation

Running your own instance of DataBank

DataBank is written in Python. The DataBank code is licensed under MIT and can be checked out from https://github.com/datacite/Databank

For instructions on installing DataBank, follow this document

Get in touch with the developers of DataBank

contact

Copyright provides the naked authoritative nature. If the work be libens, unless otherwise noted.

Databank API
Research data lifecycle:
- Research idea
- Planning
- Literature/data review
- Data gathering
- File organisation & local storage
- Documentation
- Curation
- Repository storage

Tools & Services:
- DMP Online
- Oxford
- DataStage
- ORDS (Online Research Database Service)
- Departmental data stores
- External data stores
- Colwiz (& other external research collaboration tools)
- External data repositories

Institutional RDM Policy
Training & Support
Sustainability
Understanding Researchers

- Previous projects have looked at specific research disciplines, but now need to generalize across institution
- Large-scale benchmarking survey undertaken in November / December 2012 – 314 responses from researchers generating data

Data Management Roll-out at Oxford (DaMaRO)
Research Data at Oxford

Types of data

- Textual
- Numerical
- Statistical
- Geospatial
- Images
- Audio
- Multimedia
- Bibliographic
- Other (please specify)

How data is stored

- In word processor files
- In tables or spreadsheets
- In relational databases
- In document databases (or other 'unstructured' forms of database)
- With XML mark-up
- As RDF triples
- Other (please specify)
Researcher Attitudes toward RDM

Importance of RDM

- Essential -- My research would suffer significantly if my data were not properly managed
- Important -- My research benefits from the time spent managing data
- Helpful up to a point -- Time spent managing research data can make life easier further down the line, but it's not a very significant aspect of research
- Not important -- Devoting time to managing research data would be a distraction from the real work of research

But fewer than a quarter had received any information about RDM from the University.

Ever been inspired to undertake new research after looking at past shared data?

- Yes
- No
- Seeing existing research data may have played a part in shaping new research ideas, but has never been a particularly significant factor
- Don't know / can't recall

Data Management Roll-out at Oxford (DaMaRO)
Researcher Attitudes Toward Data Sharing

Have you (or has your research team) ever deposited your research data in a dedicated repository or data store?

- **Yes**
- **No**
- **Can't recall**

If ‘no’, why not?

- Not aware of an appropriate data store
- Do not want to share it publicly
- First project generating data
- Nothing worth preserving previously
- Not had time to get around to it
- Other

‘Other’s included confidentiality agreements, security concerns, copyright, and ‘will do so shortly’

Data Management Roll-out at Oxford (DaMaRO)

[University of Oxford logo]
Sustainability

• “It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.”

  Research Councils UK Common Principles on Data Policy

• Different elements of infrastructure require different models
  – Cost recovery (paid by funder or researchers themselves) – possibly with subsidy
  – Centrally funded (included in overheads)

• Value of whole is greater than the sum of parts

• Non-institutional RDM solutions may appear cheaper, or appear to give more flexibility
Sustainability Solutions

- Still very much work in progress at Oxford!
- Lightweight, open source software (DataStage)
- Full service with support (ORDS)
- Materials embedded in existing services (training)
  - Important where disciplinary differences loom large
Final Thoughts

• Funders increasingly driving agenda, but researchers need to shape the solutions, else they won’t be used

• Central services require assiduous promotion to researchers
  – And this should be factored into cost models

• Try to assess potential uptake of services
  – Typically high fixed costs, low variable costs

• Different business models for different RDM services
  – Adopt a modular structure so components can be added (or removed) in future
http://damaro.oucs.ox.ac.uk/

damaro@oucs.ox.ac.uk

Please email us to receive progress updates