Research Data Services @ Edinburgh: MANTRA & Edinburgh DataShare

Stuart Macdonald
EDINA & Data Library
University of Edinburgh
**Context**

- **EDINA and Data Library** are a division within Information Services (IS) of the University of Edinburgh.

- **EDINA** is a Jisc-funded centre for digital expertise providing national online resources for education and research.

- **Data Library & Consultancy** assists Edinburgh University users in the discovery, access, use and management of research datasets.

- The **Data Library** is part of the new *Research Data Service* – the culmination of a 48 month RDM Roadmap (Phases 0 - 4) to implement the University’s RDM Policy and develop a suite of RDM Services that map onto the research lifecycle to support our researchers.

- **Data Library Services**: [http://www.ed.ac.uk/is/data-library](http://www.ed.ac.uk/is/data-library)

- **EDINA**: [http://edina.ac.uk/](http://edina.ac.uk/)
University of Edinburgh RDM Policy

• University of Edinburgh is one of the first Universities in UK to adopt a policy for managing research data: http://www.ed.ac.uk/is/research-data-policy

• The policy was approved by the University Court on 16 May 2011.

• It’s acknowledged that this is an aspirational policy and that implementation will take some years.

Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University’s commitment to research excellence.

2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).

3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.

4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.

5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.

6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.

7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.

8. The legitimate interests of the subjects of research data must be protected.

9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.

10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.
Policy implementation: RDM Roadmap

Research Data Management Roadmap (v.2)
http://www.ed.ac.uk/information-services/about(strategy-planning/rdm-roadmap)
RDM AT THE UNIVERSITY OF EDINBURGH

Data catalogue in PURE
http://edin.ac/1OF8Aug

Data Vault
Ready by mid-2016

DataSync
http://www.ed.ac.uk/is/datasync

DataStore

MANTRA
http://datalib.edina.ac.uk/mantra/

A Guide to the Research Data Service

The University of Edinburgh

Research Data Management Policy

RDMP online
https://dmponline.dcc.ac.uk/

Data blog
http://datablog.is.ed.ac.uk/

Research Data Management Programme

Ready by mid-2016

http://www.ed.ac.uk/is/data-sync

http://www.ed.ac.uk/files/atoms/files/rdm_service_a5_booklet_0.pdf

http://www.ed.ac.uk/is/research-data-policy

http://datashare.is.ed.ac.uk/

http://www.ed.ac.uk/is/data-management
Research Data MANTRA

MANTRA is a free online course for those who manage digital data as part of their research project.

http://datalib.edina.ac.uk/mantra/
MANTRA overview

Project funded by Jisc Managing Research Data Programme (2010-2011)

Partnership between:

• Data Library
• Institute for Academic Development

Grounded in three disciplinary contexts: social science, clinical psychology and geoscience.

Aims to develop online interactive open learning resources for PhD students and early career researchers that will:

• Raise awareness of the key issues related to research data management.
• Provide guidelines for good research practice.
Online learning module

Eight units with activities, scenarios and videos:

- Research data explained
- Data management plans
- Organising data
- File formats and transformation
- Documentation and metadata
- Storage and security
- Data protection, rights and access
- Preservation, sharing and licensing

Four data handling practicals: SPSS, NVivo, R, ArcGIS

Xerte Online Toolkits – University of Nottingham
Online learning module

- Delivered online – self-paced, available ‘anytime, anyplace’.
- One hour per unit.
- Read and work through scenarios & online activities (incl. videos etc).
- CC licence to allow manipulation of content for re-use with attribution.
- Portable content in open standard formats (e.g. SCORM).
- Learning materials deposited with an open licence in JorumOpen and Xpert OER repositories.
User Profiles

Research student:
May want to use MANTRA for:

• Introduction to concepts and terminology of RDM
• Provide an overview of
  • how to collect, manage data for dissertations, reports, fieldwork
  • how to plan and develop research projects (data gathering, analysis and storage)
• Learn how to use R, SPSS, NVivo or ArcGIS.

Career researcher:
May want to use MANTRA for:

• Reflect on your current data management practice.
• Help develop DMPs.
Senior academic:

May want to use MANTRA for:

- Discovering content that might help students and be useful in teaching and learning activities.
- Checking content and recommended resources to revise DMPs.
- Gaining awareness of good RDM practices and benefits of sharing and licensing of their own data.

Information professional:

May want to use MANTRA for:

- Training support staff to increase awareness of institutional data management requirements.
- Assisting academics and research students preparing DMPs.
- Gaining awareness of the benefits of data sharing and licensing, and digital preservation practices.
• **DIY Training Kit for Librarians:** an RDM course for librarians covering 5 topics involving reading assignments from MANTRA, reflective writing, and 2-hour F-2-F training sessions, including group exercises.

• Fourth release (Sept. 2014) of MANTRA - revised and updated with new content, videos, reading lists, and interactive quizzes. Three of the data handling tutorials were rewritten and tested for newer software versions.

• Oct. 2015 - Research Data MANTRA Forum: [http://www.jiscmail.ac.uk/mantra-forum](http://www.jiscmail.ac.uk/mantra-forum)
Learners can also proceed at their own pace. No formal credit is assigned for the MOOC, Statements of Accomplishment will be available to any learner who completes a course for a small fee.
Background

• DISC-UK DataShare Project – funded by the Jisc Repositories and Preservation Programme (Mar. 07 – Mar. 09)
• A collaborative project exploring new pathways to assist researchers wishing to share data via institutional repositories
• Edinburgh DataShare is an open institutional repository of multi-disciplinary datasets produced at the University of Edinburgh.
• Tangible deliverable from the project and hosted by the Data Library.
• Researchers producing research data associated with a publication, or which has potential use for other researchers, can upload their dataset for sharing and safekeeping.
Scope

• Available for University of Edinburgh researchers & their collaborators primarily for research projects without a domain repository.
• No limits in terms of subject matter or data types.
• An IS service since 2010 - RDM Programme funding for development allows enhancements.
• DataShare Supports University of Edinburgh RDM Policy (clause 5).
• Promoted as part of Research Data Service, one of a range of RDM Services developed for University of Edinburgh researchers
• DataShare not for potentially disclosive, commercially sensitive data
• Link in PURE Data Catalogue from publication to data record in DataShare
Benefits for stakeholders (funder, researcher, institution)

• Edinburgh DataShare acts as a trusted digital repository for research data, where none is designated by a funder.

• Data will be discoverable and accessible for others to use beyond the life of a research project.

• A permanent identifier can be recorded with your funder to ensure persistent access.

• In addition, some publishers require that the data on which a publication is based is made available by the author.

• By depositing once you can meet all future requests by researchers wanting a copy of your data.
Metadata and Discoverability

• DataShare is a customised DSpace instance.

• Selection of DataCite-compliant DCMI metadata fields for discovery of datasets through Google and other search engines via OAI-PMH.

• Records are harvested by Data Citation Index.

• Citation field automatically generated based on specified metadata values.

• Persistent identifier minting (DataCite DOI).

• Discovery metadata only; documentation files required to allow re-use (part of manual QA check).
Policies

• No mandate for deposit.
• Open data or embargo.
• Self-deposit model:
  – Guidance, such as checklist for deposit, user guide with screenshots.
  – Meetings to discuss data welcome; assisted deposit where warranted.
• Basic quality assurance checks by staff (documentation exists, file formats, file integrity).
• Creative Commons 4.0 licence by default; open metadata
• Preservation policy; depositor agreement; service level definition; recommended file formats, submission policy.
Edinburgh Datashare: Enhancements

- Load balancing between 2 remote sites (with automatic failover)
- Developmental server established behind University authentication – for depositors to test repository functionality
- SWORD (Push) – utilising SWORD API for batch deposit of large and/or many files from remote computers
- NEW - Implemented HTML5 resumable upload in the DataShare web interface to allow depositors to easily and quickly deposit individual files up to 15 GB – multiple files can be uploaded by drag ‘n’ drop.
- Faceted browsing by data creator, subject classification, keywords, funder for community and collection
- Awarded Data Seal of Approval Certification (Oct. 2015)
- Research data deposit from RSpace electronic notebook interface into DataShare (prototype)
Future

• Streaming multi-media files (files too big to play in browsers) – dependent upon browser choice, plug-ins loaded, network speed

• Display multimedia gallery for images

• Integrating an SFTP server to allow users to retrieve filesets larger than our current 20 GB limit.
  • All files downloadable as a zip file.
  • We anticipate making numerous filesets around 100 GB available in this way in the medium term.
  • Storage rather than network/browser timeout will become the limiting factor on fileset size.
  • Move DSpace asset store to a location where more storage space is available
Thanks!

stuart.macdonald@ed.ac.uk