RDM PROGRAMME @ EDINBURGH
- Service Interoperation

Stuart Macdonald
RDM Service Coordinator
University of Edinburgh

stuart.macdonald@ed.ac.uk
Background

• EDINA and University Data Library (EDL) together are a division within Information Services (IS) of the University of Edinburgh.

• EDINA is a Jisc-designated centre for digital expertise & online service delivery - [http://edina.ac.uk/](http://edina.ac.uk/)

• The Data Library assists Edinburgh University users in the discovery, access, use and management of research datasets: [http://www.ed.ac.uk/is/data-library](http://www.ed.ac.uk/is/data-library)

• Research and Learning Services offer specific services to the University with a focus on enabling research (OA publications, research data, bibliometrics) and resource discovery for learners (resource search systems).
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](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.
An RDM Policy Implementation Committee was set up by the VP of Knowledge Management charged with delivering services that will meet RDM policy objectives:

- Membership from across Information Services
- Iterate with researchers to ensure services meet the needs of researchers

The VP also established a Steering Committee led by Prof. Peter Clarke with members of the Research Committee from the 3 colleges, IS, and the Research Office (ERI).

Their role is to:

- Provide oversight to the activity of the Implementation Committee
- Ensure services meet researcher requirements without harming research competitiveness

- Cross-divisional collaboration

- Services already in place:
  - Data management planning
  - Active working file space = DataStore
  - Data publication repository = DataShare

- Services in development:
  - Long term data archive = DataVault
  - Data Asset Register (DAR)

- RDM support: Awareness raising, training & consultancy

http://edin.ac/1u3sKqy
Research Data Management Planning

Performed at the conceptual stage before research data are created (what, where, who, how)

Customised instance of DCC’s **DMPonline toolkit** for University of Edinburgh use:

- Funders and local (non-funder) DMP templates
- Institutional guidance (storage, services, support)
- Piloting customised school-level guidance - end of Jan. 2015

Tailored DMP assistance for researchers submitting research proposals (F-2-F)
DataStore

- NAS facility to store data that are actively used in current research activities
- 1.6PB storage initially. 0.5 TB (500GB) per researchers, PGR upwards
- Up to 0.25TB of each allocation can be used for “shared” group storage
- Cost of extra storage: £200 per TB per year = 1TB primary storage, 10 days online file history, 60 days backup, DR copy
- Infrastructure in place. Allocation of space devolved to IT departments of respective Schools overseen by Heads of IT from each College.
- De-allocation policy detailing responsibilities and storage costs for ‘orphaned data’ - pending approval by Steering Committee

DataShare

- Edinburgh DataShare is the University’s open access multi-disciplinary data repository: [http://datashare.is.ed.ac.uk](http://datashare.is.ed.ac.uk)
- Assists researchers disseminate their research, get credit for data publication, and preserve their data for the long-term (DOI, licence, citation)
- Help researchers comply with funder requirements to preserve and share your data and complies with Edinburgh’s RDM Policy
Data Vault

- Safe, private, store of data that is only accessible by the data creator or their representative
- Secure storage: File security; Storage security; Additional security: encryption
- Long term assurance
- Automatic versioning
- Front-end application requirements (authorisation, retention & deletion, file structure, file transfer, integration

Data Asset Register (DAR)

- A catalogue of data assets produced by University of Edinburgh researchers
- A key component of the University of Edinburgh RDM systems
- Will give researchers a single place to record the existence of the data assets they produce for discovery, access, and re-use as appropriate.
- Paper proposing the adoption of PURE as the University’s DAR provisionally accepted the RDM Steering Committee (Oct. 2014)
Interoperation

Systems do not live in isolation, and become more powerful and more likely to be used if they are integrated with each other.

However, the last thing that we want is to introduce further systems that need to be fed with duplicate information.

This means interoperation for some or all of the components.
RDM Support

Making the most of local support!

- RDM team will work with the Research Administrators in each School.
- Academic Support Librarians (who represent each of the 22 Schools).
- IT staff in each School.
- ERI staff. They will be receiving RDM training.
- Each School’s Ethics Committee
- Queries can be sent to the IS Helpline who will direct them as appropriate.
Awareness raising

- Introductory sessions on RDM services and support for research active and research admin staff in Schools / Institutes
- RDM website: [http://www.ed.ac.uk/is/data-management](http://www.ed.ac.uk/is/data-management)
- RDM blog: [http://datablog.is.ed.ac.uk](http://datablog.is.ed.ac.uk)
- RDM wiki: [https://www.wiki.ed.ac.uk/display/RDM/Research+Data+Management+Wiki](https://www.wiki.ed.ac.uk/display/RDM/Research+Data+Management+Wiki)
MANTRA

- MANTRA is an internationally recognized self-paced online training course developed here for PGR’s and early career researchers in data management issues.

- Data handling exercises with open datasets in 4 analytical packages: R, SPSS, NVivo, ArcGIS

- CC License & embed units in VLE’s e.g. Moodle

http://datalib.edina.ac.uk/mantra
Training: Tailored Courses

- A range of training programmes on research data management (RDM) in the form of workshops, power sessions, seminars and drop in sessions to help researchers with research data management issues

- Creating a data management plan for your grant application
- Research Data Management Programme at the University of Edinburgh
- Good practice in Research Data Management
- Handling data using SPSS
- Handling data with ArcGIS

http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management/rdm-training

http://edin.ac/1kRMPv3
Service Integration

• **DataShare** is a customised DSpace instance with a selection of OAI-PMH compliant DCMI metadata fields for data discovery through Google and other search engines
  • Records are harvested by Data Citation Index

• SWORD API utilised for batch deposit of large and/or many files from remote computers (‘Push using http’)

• Internal batch ingest of many/large files to circumvent 2.1GB limit via the web interface (‘Pull via command line interface’)
  • Use of checksums to determine that delivered object mirrors deposited object

• Working with F1000Research to define a workflow for depositors to get credit for data as research output by publishing data articles - [http://f1000research.com/](http://f1000research.com/)
  • Published new list of data journals for our depositors
DSpace GITHUB plugin* - allows software to be archived from GitHub (or similar) source code repository into DataShare, which can then be assigned a DOI to facilitate citation - using the SWORD deposit protocol

DataSync - to allow sharing of data on DataStore:

• drop-box type functionality
• uses open source ‘ownCloud’ technology
• desktop and mobile machines synchronize files with the ownCloud server
• file updates are pushed between all devices connected to a user's account.

Research data deposit from RSpace Electronic Lab Notebook (ELN) interface into DataShare (and Datastore & Data Vault) using SWORD

Integrating an electronic lab notebook with a university research infrastructure: Case Study with RSpace at the University of Edinburgh

Rory Macneil
Research Space
rmacneil@researchspace.com
Overview

- ELNs – where the demand coming from
- RSpace – origins and overview
- RSpace at Edinburgh
  - Linking to files in Edinburgh DataStore
  - Depositing content in Edinburgh DataShare
  - Archiving in Edinburgh DataVault
- Platform for integration with other RDM infrastructures
Who and what is driving demand for ELNs?

- **Researchers**
  - Utility and convenience of paper lab book + online capabilities
  - On multiple devices
  - File management/integration

- **Groups/PIs**
  - Controlled sharing
  - Collaboration
  - Group management
  - File management/integration

- **Institutions: data librarians, research admins, IT, commercialisation offices**
  - Enterprise features: Scalable deployment, Single Sign On
  - IP protection: audit trail, signing
  - Publishing
  - Archiving
  - Repository integration
  - File management/integration
RSpace

- Conceived in response to Wisconsin RFP and trial 2011 - 2012

- Developed with Wisconsin by Research Space 2012 - 2013
# Researcher experience

<table>
<thead>
<tr>
<th>Feature</th>
<th>Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sketching</td>
<td>✓</td>
</tr>
<tr>
<td>Image annotation</td>
<td>✓</td>
</tr>
<tr>
<td>Chemical structures</td>
<td>✓</td>
</tr>
<tr>
<td>Notebook</td>
<td>✓</td>
</tr>
<tr>
<td>Forms</td>
<td>✓</td>
</tr>
<tr>
<td>Templating</td>
<td>✓</td>
</tr>
<tr>
<td>Snippets</td>
<td>✓</td>
</tr>
<tr>
<td>PDF export</td>
<td>✓</td>
</tr>
<tr>
<td>Export to html</td>
<td>✓</td>
</tr>
<tr>
<td>File gallery</td>
<td>✓</td>
</tr>
<tr>
<td>Journal view</td>
<td>✓</td>
</tr>
<tr>
<td>Tablet friendly</td>
<td>✓</td>
</tr>
<tr>
<td>Clean design</td>
<td>✓</td>
</tr>
<tr>
<td>Performance</td>
<td>✓</td>
</tr>
<tr>
<td>Round trip editing</td>
<td>✓</td>
</tr>
<tr>
<td>Offline access</td>
<td>✓</td>
</tr>
</tbody>
</table>
PI/Lab support

<table>
<thead>
<tr>
<th>Feature</th>
<th>Enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharing</td>
<td>✓</td>
</tr>
<tr>
<td>Messaging</td>
<td>✓</td>
</tr>
<tr>
<td>Lab set up enabled</td>
<td>✓</td>
</tr>
<tr>
<td>Group management</td>
<td>✓</td>
</tr>
<tr>
<td>Inter-group collaboration</td>
<td>✓</td>
</tr>
</tbody>
</table>
Institutional requirements (IT, data librarians, research admins, commercialisation offices)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sign on</td>
<td>✓</td>
</tr>
<tr>
<td>Tiered admin</td>
<td>✓</td>
</tr>
<tr>
<td>Group set up</td>
<td>✓</td>
</tr>
<tr>
<td>IP support</td>
<td>✓</td>
</tr>
<tr>
<td>Export to XML</td>
<td>✓</td>
</tr>
<tr>
<td>Archiving</td>
<td>✓</td>
</tr>
<tr>
<td>Repository integration</td>
<td>✓</td>
</tr>
</tbody>
</table>
RSpace design advantages

- Easy data entry
- Easy and flexible data structuring
- Multiple ways of getting data out (and back in)
  - Export PDF
  - Export to html
  - Export Zip (XML)
  - Re-import, preserving structure
  - Archive (with metadata)
Business Model

- Free public cloud for labs and individuals
- Institutional deployments @$100/user/year
- Seamless movement of groups and data between different RSpaces
RSpace at Edinburgh

- Linking to files in Edinburgh DataStore
- Depositing content in Edinburgh DataShare
- Archiving in Edinburgh DataVault
Linking to DataStore

“My plan for workflow would be generally to deposit my data in DataStore either from the wet lab instruments (gel photos, elisa data, etc, and also possibly directly from an iPad) or from in silico data analysis I’ve been doing, and then link to it from within RSpace.”
Linking to DataStore

Experiment

Procedure

Results

Results.xls

ELN

UoE DataStore
Exposing DataStore File Roots
Linking to a DataStore File

Here we look at recordings from a particular voice. The data file is here:
Linking to a DataStore File

Here we look at recordings from a particular voice. The data file is here:

File Store: Neurovoice

- .Temporary/Items
- .Trash-628401
- quota.txt
- recordings
  - database
  - guests
- MNDA_27_06_11.zip
Linking to a DataStore File

Here we look at recordings from a particular voice. The data file is here:

MNDA_27_06_11.zip
DataStore integration: Designing for the User

- Single Sign On via EASE
- Seamless file access
  - Common Internet File Standard with user credentials
- Multiple file roots per user
  - Idiosyncratic organisation
  - Sharing between users/groups
  - Accessing external files (DataStore, Box, Dropbox)
Exporting to DataShare

RSpace

UoE DataShare
Adding metadata

<table>
<thead>
<tr>
<th>Options</th>
<th>Type</th>
<th>Owner</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>admin</td>
<td></td>
<td>Energy Lab</td>
</tr>
<tr>
<td>✔️</td>
<td>admin</td>
<td></td>
<td>Non-Domestic</td>
</tr>
<tr>
<td></td>
<td>admin</td>
<td></td>
<td>TEDDI project</td>
</tr>
</tbody>
</table>

**Required Information for Repository Deposit**

- Enter your name. Put any forename(s) in the first box and your last name in the second.

- Enter the publishing institution, followed by any school or department, e.g. University of Edinburgh, School of GeoSciences, Institute of Geography.

- Enter a title for the deposit.

- Enter the names of any contributors who created the contents of the deposit, with first and then surname. Use 'Add' to add contributors and 'X' to remove.
RSpace – DataShare integration: Backend platform

– Edinburgh DataShare has three interfaces/APIs
  ● Web-UI
  ● Python
  ● SWORD (simple Java based web-service which supports repository deposits)
– RSpace uses the SWORD Interface
– The SWORD server accepts a file for deposition if a METS description file is provided
Four part METS implementation in RSpace – DataShare integration

• RSpace uses the standard METS header
• DMD -- field definitions are based on Dublin Core
  – Four required fields in Edinburgh DataShare -- contributor, publisher, title, and data creator -- must be completed as part of the deposit through RSpace
  – Additional optional fields can be filled in later by DataShare administrator:
    • FUNDER, SPATIAL_COVERAGE, TIME_PERIOD, DATA_CREATOR, AVAILABLE_DATE, DESCRIPTION_ABSTRACT, DESCRIPTION_TOC, LANGUAGE, RELATION_VERSION_OF, RELATIONREFERENCED_BY, SUPERCEDES, RIGHT, SOURCE, SUBJECT_KEYWORDS, SUBJECT_CLASSIFICATION, ALTERNATIVE_TITLE
• All zipped files and their mime-types (e.g. application/pdf, text/html) are included
• A structure map describes the full structure and relationships between the above three elements
RSpace – DataShare integration: Workflow

• Front end trigger
  – An RSpace user selects files/folders/notebooks to be deposited from RSpace, and starts the deposit process
• Backend to support the user workflow
  – RSpace extracts the associated data and resources from its database and file-store
  – These are turned into xml files
  – METS is used to describe the zip file and each selected file
  – The xml, resource, and METS files are zipped into a zip file for archiving
  – The DSpace SWORD client deposits the zip file to DataShare after an authentication and validation
Archiving in Edinburgh DataVault

- DataVault functionality/API not yet specified
- Anticipate use of XML zip archive
- Many requirements to be determined
  - e.g., searching, restoration
RSpace and Edinburgh RDM

DataStore

DataShare

RSpace server

DataVault

User / Browser
RSpace and RDM: Other institutions

- File store
- Interface
- RSpace server
- Interface
- DSpace/other repositories
- XML
- Archive
- User/Browser