

# **DSpace for Data Revisited**

**Stuart Macdonald**  
**Robin Rice**

**EDINA & Data Library**  
**University of Edinburgh**

# Context

- **EDINA and Data Library (EDL) together are a division within Information Services (IS) of the University of Edinburgh.**
- **EDINA is a jisc-funded National Data Centre providing national online resources for education and research.**
- **The Data Library service (established in 1983) assists Edinburgh University users in the discovery, access, use and management of research data assets.**

# Background



- **DISC-UK DataShare Project – funded by JISC (Mar. 07 – Mar. 09) - a collaborative project exploring new pathways to assist researchers wishing to share data via institutional repositories**
- **Edinburgh Datashare is an online institutional repository of multi-disciplinary research datasets produced at the University of Edinburgh, 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 is allowing enhancements**

**DataShare Supports University of Edinburgh RDM Policy (clause 5)**

**Promoted as part of RDM programme, one of a range of RDM Services being developed for University of Edinburgh researchers**

# **Metadata and Discoverability**

**DataShare is a customised DSpace instance with a selection of 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 (Handle) on dataset landing page**

**Conforms to DataCite minimum fields (DOIs soon)**

**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: KISS workflow**
  - **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)**
- **Open Data Commons Attribution licence by default; open metadata**
- **Preservation policy; depositor agreement; service level definition**



## Information Services

Edinburgh DataShare



### Search Edinburgh DataShare

Go

Advanced Search

### Browse

- Edinburgh DataShare
- Research Communities
- Date Issued
- Data Creators
- Titles
- Subjects

### My Account

- Login
- Register

### RSS Feeds

- RSS 1.0
- RSS 2.0
- Atom

### What is Edinburgh Data Share?

Edinburgh DataShare is an online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh, hosted by the [Data Library](#) in Information Services.

Edinburgh University researchers who have produced research data associated with an existing or forthcoming publication, or which has potential use for other researchers, are invited to upload their dataset for sharing and safekeeping. A persistent Identifier and suggested citation will be provided.

### Deposit Your Data

Deposit

- How to deposit

### Research Communities in Edinburgh Data Share

Select a community to browse its collections.

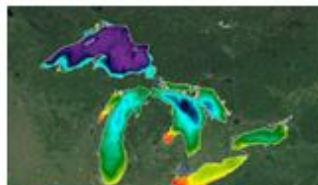
- Business School
- Edinburgh College of Art
- Edinburgh Law School
- Information Services (IS)
- Moray House School of Education
- Royal (Dick) School of Veterinary Studies
- School of Biological Sciences
- School of Biomedical Sciences
- School of Chemistry
- School of Clinical Sciences
- School of Divinity
- School of Economics
- School of Engineering
- School of Geo Sciences
- School of Health in Social Science
- School of History, Classics and Archaeology
- School of Informatics
- School of Literatures, Languages and Cultures
- School of Mathematics
- School of Molecular, Genetic and Population Health Sciences
- School of Philosophy, Psychology and Language Sciences
- School of Physics and Astronomy
- School of Social and Political Science

### Search

Enter some text in the box below to search Edinburgh DataShare.

Go

### Spotlight



### Item submission

Describe | Upload | Review | License | Deposit Agreement | Complete

Great! New Rept Temp of Ge (EEC)

Info

Lat

### Describe Item

**Title:**   
Enter the main title of the dataset.

**Type:**   
Dataset  
Image  
Interactive Resource  
Moving Image  
Service  
Software  
Select the type(s) of content of the item. To select more than one value in the list, you may have to hold down the 'CTRL' or 'Shift' key. Please select 'Dataset' if in doubt.

**Depositor:**  Macdonald  Stuart  
Surname First name(s)  
Enter the name of the person entering this record.

**Data Creator:**   Add  
Surname First name(s)  
Enter the names of the data creators / principal investigators of this item below. If the name of the data creator is unknown please be sure to fill out the Publisher field.

**Data Publisher:**   
Enter the name of the organisation or service responsible for making the data available. Example: University of Edinburgh, School of GeoSciences, Institute of Geography. Note: Publisher is a mandatory field if no Creator is specified.

**Funder:**   
AHRC - Arts and Humanities Research Council  
BBSRC - Biotechnology and Biological Sciences Research Council  
Carnegie Trust for the Universities of Scotland  
EPSRC - Engineering and Physical Sciences Research Council  
ERC - European Research Council  
ESA - European Space Agency  
Select each agency that funded the creation of the dataset.

Optimise the 'findability' of your submission via search engines and link to published papers by adding supplementary information.

- Provide an abstract and attach keywords.
- Link to publications and other versions of data item.
- Acknowledge sources, place an embargo, describe file structure.
- Add geography dates, and language information.

Save & Exit | Next >

Entries marked with a \* are mandatory

# Edinburgh Datashare: Enhancements

**Case-studies based on 3 piloted research groups (Roslin Institute, Clinical Psychology, School of Philosophy, Psychology and Language Sciences) were used to capture user requirements**

- **Sept. 2013 - Streamlined usability and deposit workflow e.g. collapsible non-required metadata fields, clear and simple licence information, streamlining initial self-deposit questions**
- **Nov. 2013 - Load balancing between 2 remote sites (with automatic failover)**
- **Dspace upgrade to v. 3.2**
- **Developmental server established behind University authentication – for depositors to test repository functionality**



- **Feb. 2014 – SWORD (Push) – utilising SWORD API for batch deposit of large and/or many files from remote computers**
- **June 2014 – Internal batch ingest of many/large files (Pull)**
  - currently 2.1GB limit via the web interface
  - Use of checksums to determine that delivered object mirrors deposited object
- **July 2014 – Upgrade to DSpace 4.1**
- **Request Copy Button – for user to request hidden item direct from depositor.**
  - Data Vault planned for this. Eventual plan that this becomes functionality within the DAR.

# September 2014 - End-user interface improvements

- **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 - Datashare want's to learn from the Dspace community about how to handle this material**
- **Faceted browsing – by community and collection**
- **Usability and user testing.**

# Future

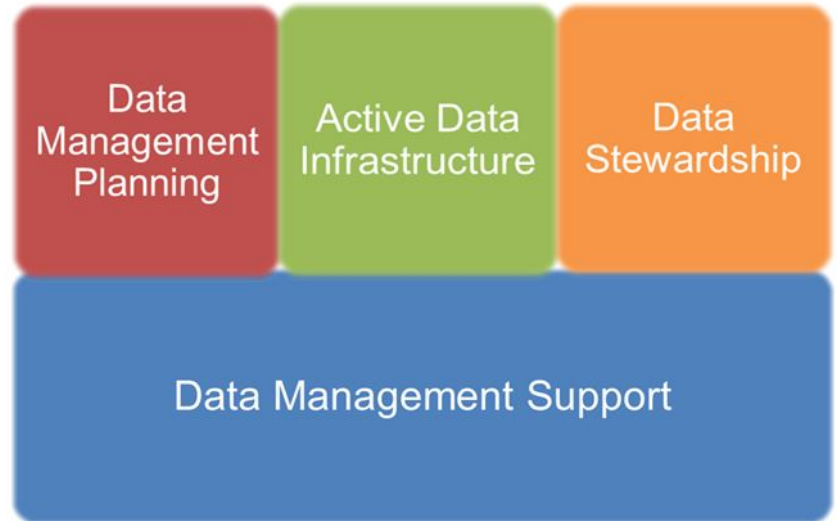
- Pursuing Data Seal of Approval as part of RDM Roadmap
- Joining DataCite via British Library: will offer DOIs shortly
- Research data deposit from RSpace electronic notebook interface into DataShare
- Working with F1000Research to define a workflow to alert depositors to opportunities to get credit for data as research output
  - Published new list of data journals for our depositors

# University of Edinburgh RDM Programme

- **University of Edinburgh is one of the first Universities in UK who adopted an RDM 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**
- **DataShare Supports University of Edinburgh RDM Policy (clause 5)**
- **Promoted as part of RDM programme, one of a range of RDM Services being developed for University of Edinburgh researchers**

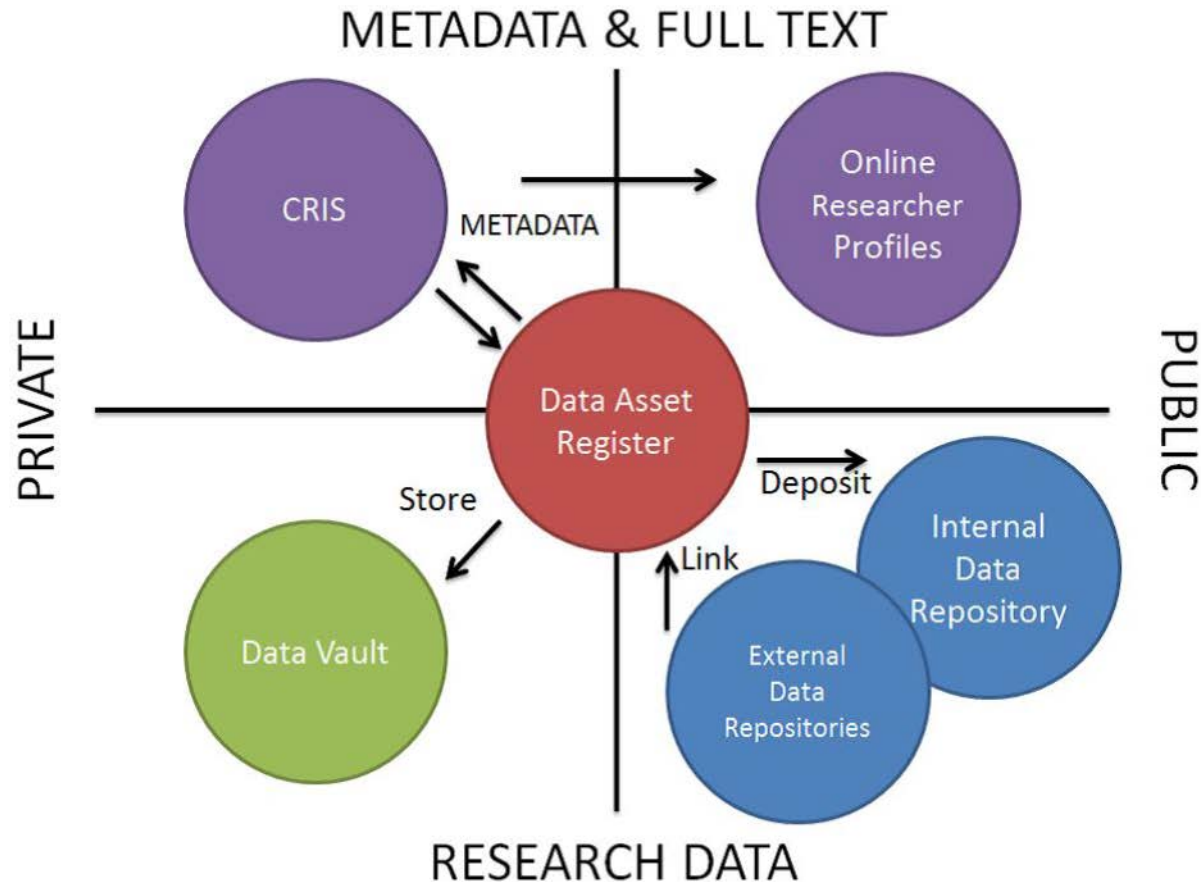
- **Services already in place:**
  - Data management planning
  - Active working file space = DataStore
  - Data publication repository = DataShare
- **Services in development:**
  - Long term data archive = DataVault
  - Data Asset Register (DAR)
- **RDM support:**  
Awareness raising,  
training & consultancy

## RDM Roadmap



<http://edin.ac/1u3sKqy>

# DataShare & related Univ. of Edinburgh RDM services



# Thanks!

## Links:

- Data Library services: <http://www.ed.ac.uk/is/data-library>
- EDINA: <http://edina.ac.uk/>
- Edinburgh DataShare: <http://datashare.is.ed.ac.uk/>
- Edinburgh University data policy: <http://www.ed.ac.uk/is/research-data-policy>
- Research Data MANTRA: <http://datalib.edina.ac.uk/mantra>

**stuart.macdonald@ed.ac.uk**