RDM PROGRAMME @ EDINBURGH
- an institutional approach

Stuart Macdonald
RDM Services Coordinator
University of Edinburgh

stuart.macdonald@ed.ac.uk
DAF Implementation Project

Edinburgh Data Audit Framework (DAF) Implementation Project (May – Dec 2008)

A JISC-funded pilot project produced 6 case studies from research units across the University in identifying research data assets and assessing their management, using DAF methodology developed by the Digital Curation Centre.

2 main outcomes:

- Develop university research data management policy
- Develop services & support for RDM (in partnership IS)

DAF Implementation Project: [http://ie-repository.jisc.ac.uk/283/](http://ie-repository.jisc.ac.uk/283/)
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](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.
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
**Policy implementation: RDM Roadmap**

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 DMP templates
- Local (non-funder) DMP template
- Institutional guidance (storage, services, support)
- Piloting customised guidance (for funders and schools) 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
- Provision: 1.6PB storage initially (currently 2.3PB?)
- 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 incl. back-up and DR copies
- Infrastructure in place. Allocation of space devolved to IT departments of respective Schools overseen by Heads of IT from each College.

DataShare

- Edinburgh DataShare is the University’s OA multi-disciplinary data repository hosted by the Data Library: [http://datashare.is.ed.ac.uk](http://datashare.is.ed.ac.uk)
- Assists researchers who want to share their data, get credit for data publication, and preserve their data for the long-term (DOI, licence, citation)
- It can 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
- Current focus on 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
- Will give researchers a single place to record the existence of the data assets they produce for discovery, access, and re-use as appropriate.
- Adoption of PURE as the University’s DAR accepted the RDM Steering Committee for launch in mar. 2015
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.
• Bespoke RDM email address or queries can be sent to the 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 / Research Centres

• Contact Cuna Ekmekcioglu at cuna.ekmekcioglu@ed.ac.uk for a session for your School/Research Centre

• RDM website: http://www.ed.ac.uk/is/data-management

• RDM blog: http://datablog.is.ed.ac.uk

• RDM wiki: https://www.wiki.ed.ac.uk/display/RDM/Research+Data+Management+Wiki
Training: MANTRA

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

- Anyone doing a research project will benefit from at least some part of the training – discrete units

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

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

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

- Creating a data management plan for your grant application

- Handling data using SPSS

- Managing your research data: why it is important and what should you do? **NEW**

- Publishing and sharing sensitive data (pilot) **NEW**

  [http://edin.ac/1kRMPv3](http://edin.ac/1kRMPv3)
RDM Programme resourcing & staffing

RDM Programme:
Funded internally (c. £1.2 Million)
75% - infrastructure / storage
25% - staffing (recurrent for 3 years)

MANTRA and DataShare – originally Jisc project funding

2014 DCC RDM Survey* - 90% of institutions used internal funding for new appointments in RDM, for training for infrastructure

From RDM Programme (fixed term):
Data Library: 1.5 FTE equivalent (+ 2.5 FTE equivalent core funding)
IT Infrastructure: 2 FTE equivalent

Research & Library Services: 2 FTE equivalent

Following RDM training the job description of all Academic Support Librarians have been restructured to incorporate DMP Support as part of their role.

2014 DCC RDM Survey:
Overall provision for RDM is currently
4.4 FTE on average (across library, IT, research office)

4.7 FTE being the average in Russell Group institutions and
2.6FTE in other target group institutions.

**RDM staffing is expected to double to 9.5 FTE in Russell Group institutions in next year, split roughly equally across 3 groups**

For others institutions in target group RDM staffing is expected to rise only slightly to 3 FTE
Current and future activity

Discipline-specific training – based on school-level & funder DMP guidance

DataSync - to allow cross-platform sharing of data on DataStore
- Secure drop-box type functionality
- uses open source ‘ownCloud’ technology
- desktop and mobile machines synchronize files with the ownCloud server
- Files can also be accessed using a web browser without any additional software. File updates are pushed between all devices connected to a user's account.

DataStore De-allocation policy detailing responsibilities and storage costs for ‘orphaned data’ - pending approval by Steering Committee
Service interoperations:

DataShare (also DataStore & Data Vault) with RSpace Electronic Lab Notebooks

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

Refresh of ECDF Computing Cluster (‘Eddie’) complete with ‘Data Centric Computing’ business model – integrate Eddie storage & HPC, parallel and cloud computing layers with DataStore i.e. data transferred from DataStore for analysis on Eddie and then data ported back to DataStore (DataVault)

Linking of SDA toolkit with numeric ASCII data held in DataShare for the purposes of analysis (re-use)

Facility to embargo variables within numeric files (in statistical analysis package formats) for subsequent open deposit into DataShare of de-sensitised version
Metrics

Research Data Management Training Breakdown for May – December 2014

<table>
<thead>
<tr>
<th>No. of staff / students attended RDM courses</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research staff</td>
<td>55</td>
<td>16</td>
<td>8</td>
<td>12</td>
<td>14</td>
<td>8</td>
<td>27</td>
<td>12</td>
<td>152</td>
</tr>
<tr>
<td>Postgraduate research students</td>
<td>21</td>
<td>11</td>
<td>9</td>
<td>5</td>
<td>15</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td>Research support staff</td>
<td>34</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>Distribution by colleges (total for 8 months)</td>
<td>CHSS 117</td>
<td>CSE 50</td>
<td>MVM 135</td>
<td>Other* 19</td>
<td>321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* ERI 16, OLL 1, EDINA 1, L&UC 1
DMPonline usage statistics Jan – December 2014

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan 14</th>
<th>Feb 14</th>
<th>Mar 14</th>
<th>Apr 14</th>
<th>May 14</th>
<th>Jun 14</th>
<th>Jul 14</th>
<th>Aug 14</th>
<th>Sep 14</th>
<th>Oct 14</th>
<th>Nov 14</th>
<th>Dec 14</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>52</td>
</tr>
</tbody>
</table>

Total of users: 52

New users 2014

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan 14</th>
<th>Feb 14</th>
<th>Mar 14</th>
<th>Apr 14</th>
<th>May 14</th>
<th>Jun 14</th>
<th>Jul 14</th>
<th>Aug 14</th>
<th>Sep 14</th>
<th>Oct 14</th>
<th>Nov 14</th>
<th>Dec 14</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>52</td>
</tr>
</tbody>
</table>

Total of plans based on UBC templates: 30
Total of plans created by all staff: 185
Total of views of DMPonline (all): 205
DataStore Statistics for May – December 2014

Total unique users in a month

Date

# Users


SCE SAS NSU MVM ISG HSS CSG
Unique users per month
by College

# Users

Date

Unique users per month
by College

# Users

Date


CSG
HSS
ISG
MVM
NSU
SAS
SCE
Number of users
The number of users connecting to DataStore has continued to increase rapidly, with the majority of growth due to the migration of users off the College File Servers onto DataStore as their default managed desktop storage provider.

<table>
<thead>
<tr>
<th>Date</th>
<th>CSG</th>
<th>HSS</th>
<th>ISG</th>
<th>MVM</th>
<th>NSU</th>
<th>SAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-12-31</td>
<td>994</td>
<td>2447</td>
<td>514</td>
<td>2268</td>
<td>100</td>
<td>468</td>
</tr>
</tbody>
</table>

Table 1: Number of users by College

Volume of data
Data growth volumes continue to be sustainable, with rapid growth as more data is migrated to DataStore from other sources (including the College File Servers and ECDFNAS).

Each file system currently has 437TB of capacity.

<table>
<thead>
<tr>
<th>Date</th>
<th>SG Usage</th>
<th>CHSS Usage</th>
<th>CSCE Usage</th>
<th>CMVM Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-01-01</td>
<td>28</td>
<td>38</td>
<td>240</td>
<td>158</td>
</tr>
</tbody>
</table>

Table 2: Usage (TB) by File System
DataShare Repository statistics (June – September 2014)

**DataShare**

Recent deposits into DataShare are shown below, along with the rise in use of the system as a whole since 2011.

<table>
<thead>
<tr>
<th>Items deposited in DataShare</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>34</td>
</tr>
</tbody>
</table>

DataShare system usage (sessions per week, depositors and viewers)
THANK YOU!

Acknowledgements:

Dr. Cuna Ekmekcioglu (Research & Learning Services)
Sarah Jones (Digital Curation Centre)
Stuart Lewis (Research & Learning Services)
Kerry Miller (Research & Learning Services)
Robin Rice (EDINA & Data Library)
Dr. Orlando Richards (IT Infrastructure)
Dr. John Scally (Library and Collections)
Tony Weir (IT Infrastructure)