



Facilitate Open Science Training for European
Research

**Open access and research data management:
Horizon 2020 and beyond
University College Cork, April 14th & 15th 2015**

Looking After Your Data: RDM @ EDINBURGH

Stuart Macdonald
RDM Services Coordinator
University of Edinburgh

stuart.macdonald@ed.ac.uk



- **Background:** Definitions, Data Audit Framework implementation project, Univ. of Edinburgh RDM policy
- **Governance**
- **Roadmap:** Policy implementation
- **RDM services**
- **Support:** Awareness raising, training
- **Resources & staffing**
- **Service interoperation & hosting**
- **Future activity**



Research data

“Recorded, factual material commonly retained by and accepted in the [research] community **as necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.”

Engineering and Physical Sciences Research Council (EPSRC)

- There is no single accepted definition of research data. Several definitions exist. It may be useful to consider not just what material would be required to **validate research findings** but also think about what information is needed to **enable re-use** of the data.
- Not all definitions of data will be appropriate for all disciplines.
- Data can be generated for one purpose and used for a completely different one.



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/>



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.



Governance

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).

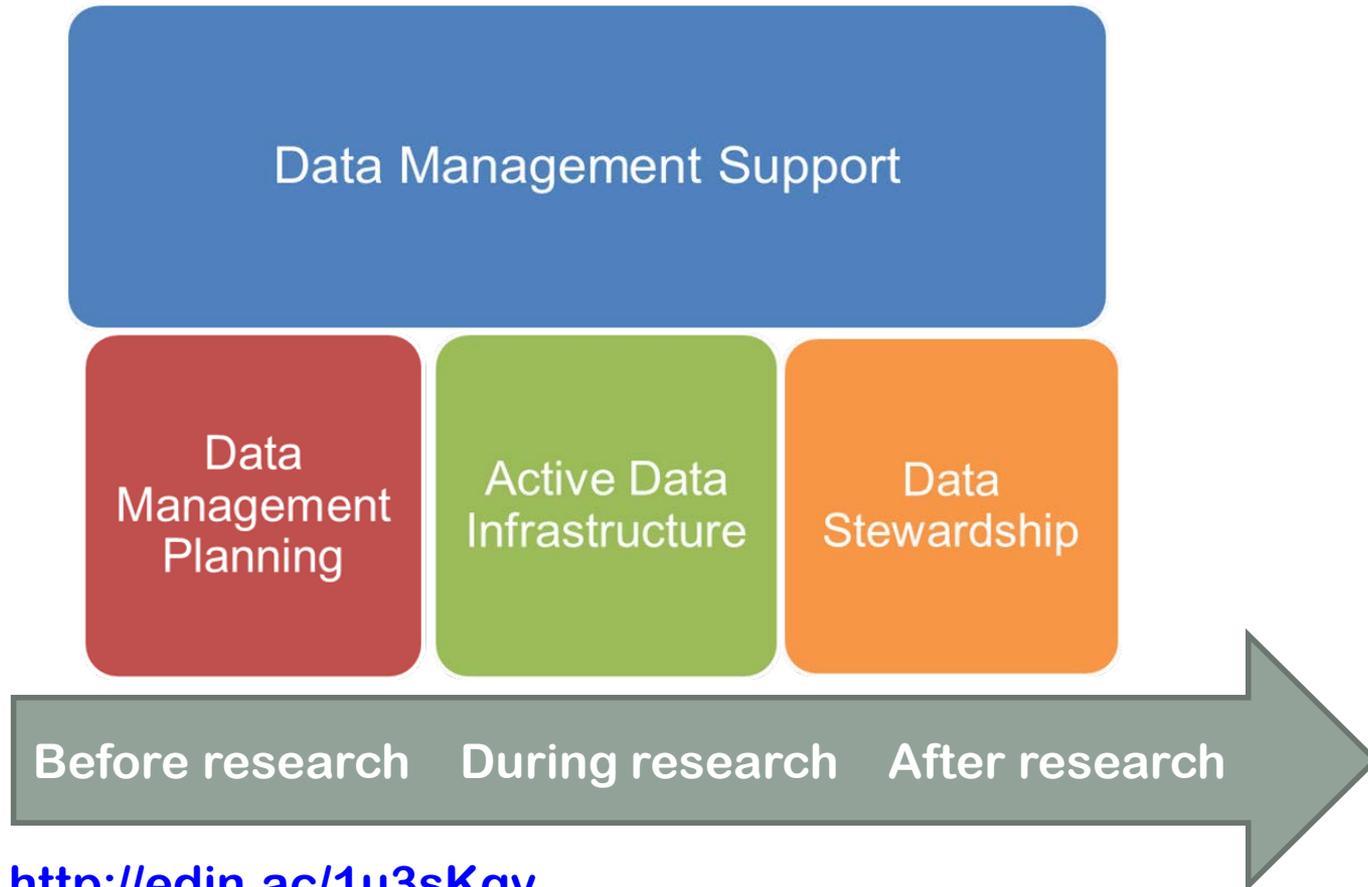
Their role is to:

- 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 - 3 Phases (Aug 2012 – May 2015)



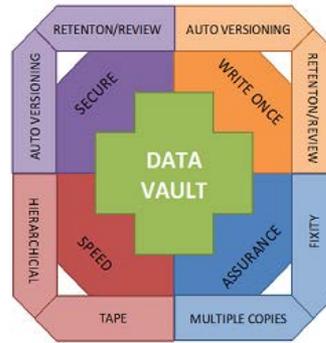
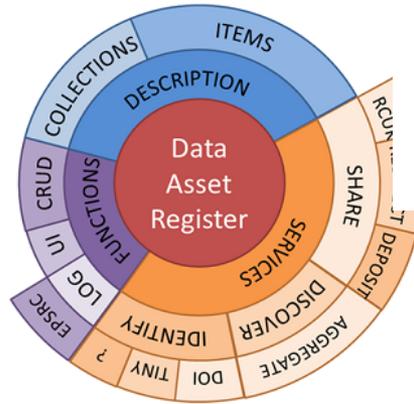
<http://edin.ac/1u3sKqy>





<https://dmponline.dcc.ac.uk/>

DAR



Data Vault

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



DataStore

RESEARCH DATA MANAGEMENT AT THE UNIVERSITY OF EDINBURGH

www.ed.ac.uk/is/data-management

<http://datalib.edina.ac.uk/mantra/>

<http://datablog.is.ed.ac.uk/>



Beta testing now!

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)

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 School IT departments of 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>
- 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 Asset Register (DAR)

- A catalogue of data assets produced by researchers at the University as part of the research process.
- 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 for launch in April 2015
- Researchers / research administrators familiar with PURE
- Take advantage of Dataset Content type
- No dataset upload nor DOI minting
- Complies with EPSRC RDM expectation V. and the University of Edinburgh RDM Policy point 6.



Data Vault Project*

- Funded by the Jisc Data Spring with local contributions (partnering with Manchester University)
- To define and develop an Open Source Data Vault to allow data creators to describe and store data safely in one of a number of archival storage options.
- Including: cloud solutions, shared storage systems, or local tape back-up infrastructure.
- The Data Vault will sit alongside DataStore and the data archiving storage
- Current focus on web front-end application requirements incl. authorisation, retention & deletion, file structure, file transfer, integration, encryption



* <http://libraryblogs.is.ed.ac.uk/jiscdatavault/>

The project is split into phases to align with the phases of the Jisc Data Spring Programme

Phase 1: April – June 2015

- Identify use cases
- Develop workflows for how the Data Vault could be used in the research process
- Examine capabilities of archival storage systems to ensure they support Data Vault features and functionality
- Define requirements specifications for Data Vault design
- Develop proof of concept

Prepare for the second Jisc Data Spring Sandpit and possible phase 2 funding

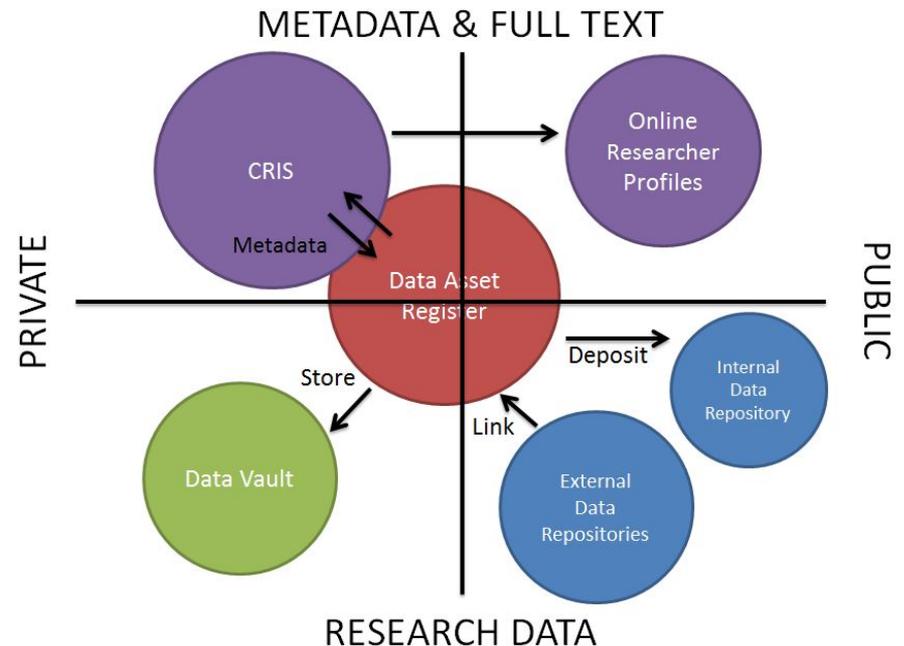


Interoperation

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 briefed on RDM Programme.
- RDM queries can be sent to the Helpline who will re-direct them to appropriate personnel via RDM knowledge base.



Awareness Raising

- Introductory sessions on RDM services and support for research active and research admin staff in Schools / Institutes / Research Centres
- 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>



THE UNIVERSITY of EDINBURGH

Schools & departments

Search

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



The screenshot shows the MANTRA website interface. At the top, the logo 'MANTRA' is displayed in a stylized font, with 'Research Data Management Training' written below it. The background features a colorful, abstract pattern. Below the logo, a text block states: 'MANTRA is an online course designed for researchers or others planning to manage digital data as part of the research process.' To the right of this text are four icons representing different user roles: 'Research Student', 'Career Researcher', 'Senior Academic', and 'Information Professional'. Below this is a navigation bar with links for 'HOME', 'ABOUT', 'ACKNOWLEDGEMENTS', 'DIY TRAINING KIT FOR LIBRARIANS', and 'FEEDBACK'. The main content area is titled 'LEARNING UNITS' and contains a grid of nine units, each with an icon and a dropdown arrow: 'Research data explained', 'Data management plans', 'Organising data', 'File formats & transformation', 'Documentation & metadata', 'Storage & security', 'Data protection, rights & access', 'Sharing, preservation & licensing', and 'Software practicals'. The 'Software practicals' unit is highlighted in orange.

<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>
- 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>



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

* Digital Curation Centre's 2014 RDM Strategy to Action Survey:

<https://zenodo.org/collection/user-dcc-rdm-2014>



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

(c.f. 4.7 FTE being the average in Russell Group institutions)

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:

RDM staffing is expected to **double** to 9.5 FTE in Russell Group institutions in next year.



Service interoperations

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

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 ported back to DataStore (→ DataVault)

Hosting

UoE host DSpace publication repositories for Scottish Digital Library Consortium (SDLC) members – can be configured for datasets for free - costs for additional disk space (£ hundreds not thousands)

IT Infrastructure can provide costed access to default DataStore as a service (authorisation and authentication permitting) - £200 per TB per year



Future activity

DataSync - to allow cross-platform synchronisation & sharing of data on DataStore – for launch April 2015

- Secure drop-box type functionality for non-sensitive data
 - Uses open source 'ownCloud' technology
 - Sync using local clients.
 - Share using local clients or web URL with colleagues anywhere
-
- **DatStore De-allocation policy detailing responsibilities and storage costs for 'orphaned data'** - pending approval by Steering Committee
-
- Gain Trusted Digital Repository status via **Data Seal of Approval** for DataShare



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)



Open access and research data management: Horizon 2020 and beyond

This event was funded by FOSTER through the European Union's Seventh Framework Programme
<http://www.fosteropenscience.eu>

and organised by

- University College Cork <http://www.ucc.ie>
- Teagasc <http://www.teagasc.ie>
- Repository Network Ireland <http://rni.wikispaces.com>

