Opportunities & Challenges in Research Data Management:

Developing a Roadmap & Services at the University of Edinburgh

Professor Jeff Haywood, Vice Principal, CIO & Librarian

University of Edinburgh, UK

jeff.haywood@ed.ac.uk

http://homepages.ed.ac.uk/jhaywood
My thanks to the many people at the University of Edinburgh who have contributed to this work.

Especial thanks go to Kevin Ashley, Cuna Ekmekcioglu, Robin Rice, John Scally, Tony Weir, Peter Clark, Stuart Lewis
You never actually own a Patek Philippe. You merely look after it for the next generation.

The patented Annual Calendar displays day month and date, which only needs resetting once a year. A sapphire crystal back reveals the self-winding movement which, like all Patek Philippe mechanical watches, is distinguished by the Geneva Seal. Ref. 5066.

Patek Philippe Geneve
Tel: (41) 22 384 20 30  www.patek.com
So, what do we mean by ‘research data’?

Anything & everything produced in the course of research
**Scale of research data creation – a proxy measure**

<table>
<thead>
<tr>
<th>In academic year 2012-13, the University of Edinburgh had:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research applications submitted</td>
</tr>
<tr>
<td>Research grants awarded</td>
</tr>
<tr>
<td>Value of awards</td>
</tr>
</tbody>
</table>

Most of these were collaborative internally (ie teams of researchers)

Many were collaborative externally, often internationally
Why keep / make available research data?

Organisation for Economic Co-operation and Development describes data as a public good that should be made available.

Research Councils UK in its 2009 code of good research conduct says data should be preserved and accessible for 10 years +

Research Funder data policies increasingly demanding of institutional commitment and provisions...
# The complex compliance context

<table>
<thead>
<tr>
<th>Research Funders</th>
<th>Policy Coverage</th>
<th>Policy Stipulations</th>
<th>Support Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Published outputs</td>
<td>Data</td>
<td>Time limits</td>
</tr>
<tr>
<td>AHRC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>BBSRC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>CRUK</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>EPSRC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>ESRC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>MRC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>NERC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>STFC</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

RDM, OpenAIRE COAR Athens – May 2014
An early win – the university RDM Policy 2011

- University of Edinburgh is one of the first few Universities in UK who adopted 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.
Where do I safely keep my data from my fieldwork, as I travel home?

How do I best keep years worth of research data secure and accessible for when I and others need to re-use it?

How do we ensure we have access to our research data after some of the team have left?

How can our research collaborations share data, and make them available once complete?

How do we ensure compliance to funders’ requirement for several years of open access to data?

PhD student

individual researcher

research team

university

supra-university

LEVEL

Seeking the real win + win + win + win + win + win......
Setting out on the path to RDM: an agreed roadmap

Research Data Management (RDM) Roadmap
August 2012 – May 2015

Information Services RDM Policy Implementation Committee

University of Edinburgh

January 2014: Version 1.2

Document Status
This is a living document of the IS Research Data Management (RDM) Policy Implementation Committee which has been approved by the RDM Steering Committee.
The stages of the data lifecycle we have to address

- Data Management Planning
- Creating data
- Documenting data
- Accessing / using data
- Storage and backup
- Sharing data
- Preserving data
Making the roadmap manageable

• RDM Programme in 3 phases:
  – **Phase 0**: August 2012 – August 2013: Planning phase, with some pilot activity and early deliverables.
  – **Phase 1**: September 2013 – May 2014: Initial rollout of primary services.
  – **Phase 2**: June 2014 – May 2015: Continued rollout; maturation of services.

• Full details of the programme is available at:
  [http://edin.ac/1eE3sav](http://edin.ac/1eE3sav)
Putting the pieces together

- **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

RDM Roadmap
Data Management Planning

Free and open web-based tool to help researchers write plans: https://dmponline.dcc.ac.uk/

It features:

- Templates based on different requirements
- Tailored guidance (disciplinary, funder etc.)
- Customised exports to a variety of formats
- Ability to share DMPs with others
Datastore

- The facility to store data that are actively used in current research activities
- Provision: 1.6PB storage initially
- 0.5 TB (500GB) per researchers, incl PhD
- Up to 0.25TB of each allocation can be used to create “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
- Early adopters in three Colleges: Geosciences; Philosophy, Psychology, & Language Sciences; Population Health Sciences; Biological Sciences; Clinical Neurosciences; and Social & Political Sciences
- CHSS: all personal spaces rolled out, waiting for information required for construction of group spaces. CHSS IT will roll it out to individual staff and ensure they have the right information to use the space.
- All other Schools /research centres in CSE & CMVM will follow.
Datashare

- Edinburgh DataShare is the University data repository for publishing your research data openly: [http://datashare.is.ed.ac.uk](http://datashare.is.ed.ac.uk)
- It will help you disseminate your research, get credit for your data collection efforts, and preserve your data for the long-term.
- It backs up the University Research Data Management policy.
- It can help you comply with funder requirements to preserve and share your data.
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

http://datablog.is.ed.ac.uk/2013/12/20/thinking-about-a-data-vault
Data Asset Register (DAR):

- is a catalogue of data assets produced by researchers working for the University of Edinburgh,
- will be a key component of the University of Edinburgh Research Data Management (RDM) system,
- will give researchers a single place to record the existence of data assets they have produced so that they can be discovered, accessed, and reused as appropriate.

http://datablog.is.ed.ac.uk/2013/12/12/thinking-about-research-data-asset-registers
Awareness raising for RDM

- 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 for RDM

- 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 (and you can pick and choose)

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

http://datalib.edina.ac.uk/mantra
Where are we at May 2014?

✓ Data Share – Live Now
✓ DMPonline – Live Now
✓ Website – Live Now
✓ Training – Live now & ongoing
✓ Awareness Raising – Live now & ongoing

• Data Management Planning Support – Aug 2014
• Data Store – Roll-out completed by Dec 2014

• Data Asset Register – Dec 2014
• Data Vault – Dec 2014
Let’s not be too complacent…..

There is survey evidence that suggests that researchers are not yet aware of the need for RDM, nor know about policy and services in their own university

- EPSRC UK-wide survey showed: only 55% of researchers knew if their university has RDM Policy; 70% don’t know if they offer support or training

- Digital Curation Centre (DCC) UK-wide survey showed: 50% said limited progress to date; 60% said it was low priority for researchers; 71% said insufficient funding; 75% said sustainability had least progress

I suspect we are very little different in Edinburgh. We have ~5000 researchers & PhD students, in ~50 academic subjects at 6 major locations, busily writing grant applications and doing the work…..there is still lots to do.......
Thank you for your attention 😊