Policy & Guidance for Research Data - An Edinburgh View

Professor Jeff Haywood
Vice Principal Knowledge Management, Chief Information Officer & Librarian to the University

jeff.haywood@ed.ac.uk

http://homepages.ed.ac.uk/jhaywood
Research data – the opportunity

• To manage the University of Edinburgh’s research data better for present and future researchers, scholars, students & the wider public

Our mission
The mission of our University is the creation, dissemination and curation of knowledge. As a world-leading centre of academic excellence we aim to:
• enhance our position as one of the world’s leading research and teaching universities and to measure our performance against the highest international standards
Research data – the scale

2008-09 financial year...

- Awards = 1064
  - *a proxy for no. of PIs to be engaged for data from these grants*
  - Science & engineering = 410
  - Medicine & Vet Medicine = 411
  - Humanities & Social Science = 235
  - *applications = 2512, a proxy for no. of PIs to be engaged re data plans*
- Value = £249M
- Sources = many!!
  - charities (324)
  - EU (112)
  - UK govt (154)
  - Industry (75)
  - Research Council (304)
  - .....
Research data – the enablers

- In general, our research communities are interested in good data management
- They press us for solutions
- It is clear from funders that it can be done (eg ESRC, NERC)
- There is widespread acceptance of the ‘open everything’ agenda
- There is a senior management recognition that ‘one can’t not do this’
- fEC funding is substantial, growing & strictly reserved for sustaining research infrastructure
- We have JISC’s Digital Curation Centre based in Edinburgh
- We have an integrated Information Service with library, e-learning & (academic+MIS) IT in the same unit
Research data – the inhibitors

- Not clear to us how to produce practical solutions to all the steps along the digital curation lifecycle – storage, curation, access, preservation, IP ownership...
- Central’s £££s
- Additional fEC income mostly returned to academic units
- Researchers want data management but don’t want to do it
- Some research councils still trying to make up their minds & acting as excuse to put off action
Research data – progress to date

- We have a research computing strategy 2008
- We have an institutional repository service for publications, inc theses, which mandates deposit
- We have learned from our data management projects
  - Data Audit Framework (DAF)
  - DataShare
- We have developed a draft policy for research data management
- We have twinned Working Groups planning for RDS and RDM
- We have maintained contact with UKRDS & intend to be a second stage Pathfinder Project (*despite the fact that these don’t exist in Scotland!!*)
- We are bringing together DataShare & ERA to give direct access to datasets from publications, inc theses (DRIL)
- We have created integrated web pages within the IS site to bring research support, including data management into one place
- We have started work to integrate data management training into PhD and ECR training thru our Transkills Unit
1.13 The Report from the recent IT & Computing Review (‘Kenway Report’) also contained a comment about research computing support, stating:

“Research Computing is inefficiently and insufficiently supported. The University is not fully exploiting its strengths in this area,.... “

The Panel suggested a recommendation to ensure better management of this area:

“**Recommendation:** The CIO should take ownership from the outset of major research and other academic computing projects...adopting best practice project methodologies and involving all relevant stakeholders from an early stage. IS should provide support for planning, procurement, implementation, operational management and, where appropriate, transition into an in-house or well managed external service... “
EXTRACT FROM RESEARCH COMPUTING STRATEGY

2 Vision
2.1 The University of Edinburgh will offer an excellent research computing infrastructure that contains world-class facilities and services, including
   • a high bandwidth network
   • access to high performance computing
   • data storage and services
   • advanced and standard software
   • support for all researchers to make the most of these services
The infrastructure will be robust and resilient and will keep pace with changing needs and technologies. Business continuity will be assured.

3 Strategy
3.1 Researchers will have access to world-class data services which will include storage, backup, sharing and access facilities to enable re-use, curation, and archive of data that they obtain through experimentation, observation and simulation or that is purchased or procured for use in research. It will be possible share data with groups both within and beyond the University.

3.3 Researchers will have the skills and knowledge to make best use of the computational facilities available to them. Training will be available in order to ensure this is the case.

3.5 Flexible and timely support will be available for all researchers to help them to make the most of these services.

3.6 The needs of individual researchers will be addressed, taking account of the different needs of large well endowd research groups compared with those of the lone scholar.

3.7 There will be partnership at all levels. Research facilities and support will be part of a robust core infrastructure, and where possible will be shared rather than owned by individual research groups. Common solutions for provision of data and computational services will be agreed between all stakeholders, so that research groups, Schools, Colleges, and Support Groups can work together.
Recommendations for Policy for RDM

- Ownership and intellectual property rights of research data assets produced by research staff and students should be clarified, including multi-institutional collaborations.
- Development and compliance with data management plans and procedures should be implemented at college, school, research unit and individual project level. These should include:
  a) the allocation of appropriate roles and responsibilities
  b) documentation/metadata to an identified minimum standard
  c) arrangements for access and re-use
  d) legal compliance.
  e) storage and backup procedures including provision for business continuity arrangements.
- Data upon which research outputs are published should be retained by the institution for sufficient time to allow reference.
- Guidance on the assignment of retention periods for research data should be made available by the University.
- Support and advice should be provided for researchers who wish to have their research data curated either after the recommended retention period, after the close of the research project, or when the researcher leaves the institution.
- A formal procedure for data transfer should be developed for when staff and students leave the institution.
Overview
Guidance for planning, documenting, securely storing and backing up research data.

Why manage research data?
Benefits of managing research data.

Defining research data
A definition for research data.

Funders’ policies and guidelines
Information about funding bodies’ policies and guidelines for research grants.

Data management planning
Data planning checklist and data management plan are the first two crucial steps of data management planning process.

Data documentation & metadata
Funders' policies and guidelines

In the UK, the following research funders have some data policies in place:

The Arts and Humanities Research Council's (AHRC) policy came into effect from April 2008. AHRC funding for the Arts and Humanities Data Service ceased on 31 March 2008.

The Biotechnology and Biological Sciences Research Council (BBSRC) takes a devolved approach and its policy came into effect in April 2007.

The Economic and Social Research Council (ESRC) requires data to be offered to its national data centres (UK Data Archive and the Economic and Social Data Service).

The Engineering and Physical Sciences Research Council (EPSRC) has no policy on data as yet.

The Medical Research Council (MRC) has no data centres but adopted a data sharing policy with effect from April 2006.

The Natural Environment Research Council (NERC), which has a detailed data policy handbook and guidelines for grant-holders, has seven designated data centres where grant-holders can deposit their data.

The Science and Technology Facilities Council (STFC), formed in 2007 by a merger of the Council for the Central Laboratory of the Research Councils (CCLRC) and Particle Physics and Astronomy Research Council (PPARC), has yet to develop its formal data sharing policy, though its facilities have well-developed individual policies.

The Wellcome Trust has both a statement on open access of research outputs and a brief policy regarding data management and sharing.

Research funders' policies - The Digital Curation Centre (DCC)
What you need to know about the data repository service if you are considering sharing your research data online.

Benefits | Policies | Definitions | Background

Go to Edinburgh DataShare
(the UoE data repository service)

Email datashare@ed.ac.uk for further information or assistance.

Benefits of depositing your dataset

Studies have shown that researchers consider access to a publication’s primary source data a significant advantage to their own research. Digital repositories facilitate the trend towards global research collaboration through the sharing of research outputs. Sharing data enables the exploration of topics not envisioned by the initial investigator. A dedicated data repository at Edinburgh ensures safe-keeping and continuing access for your dataset, while enhancing its availability and usability as a resource for research as well as learning and teaching.

As the placing of research papers in open access repositories is increasingly encouraged by funders and research institutions, so too is the requirement to manage and share data after the research project is finished. Edinburgh DataShare can assist you to meet these requirements by providing a permanent location for your completed and documented dataset, with an open access metadata record, whether or not the dataset itself is intended to be shared openly. The secure storage and continuing access which it offers can also prevent the accidental loss of invaluable data collected over years through file corruption, storage device failure or being at “orphaned” when funding money runs out or key staff move on.

Edinburgh DataShare policies

All of our service policies are available to read in full.
- Depositor Agreement (between the data creator or depositor and the service provider)
- Metadata policy (e.g. reuse)
- Data policy (e.g. re-use)
- Content policy (e.g. scope)
Welcome to Edinburgh DataShare

Edinburgh DataShare is a digital repository of multi-disciplinary research datasets produced at the University of Edinburgh, hosted by the Data Library. If you would like to deposit your data, or have staff deposit items on your behalf please send an email to datashare@ed.ac.uk.

Search

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

Communities in Edinburgh DataShare

Select a community to browse its collections:

- Information Services (IS)
- School of GeoSciences
- School of History, Classics and Archaeology
- School of Molecular and Clinical Medicine
- Scottish Alliance for Geoscience, Environment and Society (SAGES)
- test

Links

- About Edinburgh DataShare
- Data Library Home
- Edinburgh Research Archive
- DISC-UK DataShare Project

Latest Items

- Scottish Election Results 1997 - 2003 (06 Oct 2009)
- Refractive indices (500-3550 cm⁻¹) and emissivity (600-3350 cm⁻¹) of pure water and seawater (22 Sep 2008)
Digital curation training for all

Although our focus is currently on equipping data custodians with the necessary skills to share and preserve data effectively, the DCC mission doesn’t end there. To help even more data practitioners develop enhanced digital curation skills, the DCC will roll out a training programme to encourage the transfer of knowledge and best practice among data custodians, producers and users. In this way, you will be able to share your skills and the responsibility for data curation with others on your research team.

Watch this space for more information. Our specially devised training programme will be rolled out across individual institutions once we have first worked with data custodians to build firm foundations for good data curation practice across the HE sector.

Digital Curation 101
This intensive, three-day DCC course is designed to meet the specific needs of data custodians. An introductory course, DC 101 Lite, which distils the same content into half a day, is also available for complete beginners.

Tools of the Trade
Our half-day, hands-on courses for data curation practitioners focus on making use of popular curation tools such as DRAMBORA and DAF.

Train the Trainer
Each year, the DCC runs two Train the Trainer courses in digital curation
transkills

A PhD is about much more than writing a thesis. Researchers today are expected to have a wide range of professional and research skills, be adaptable, self aware and write for publication as well as for their thesis. This is a lot to achieve in a limited time.

Developing these broader professional skills and qualities mean that PhD graduates are in high demand. Many different employment sectors, including academia, business and commerce, teaching, the public and voluntary sectors, are actively recruiting PhDs.
Research data – a one sentence summary

Good progress but we still have a long way to go!!