Research Data Management
Policy, Planning and Practice

Alex Ball

DCC/UKOLN, University of Bath

22 May 2013

University of Surrey

Except where otherwise stated, this work is licensed under
Creative Commons Attribution 2.5 Scotland: http://creativecommons.org/licenses/by/2.5/scotland/
Outline

Research Data

Management

Policy

Planning

Practice

Acknowledgement. Some of the content of this talk derives from ‘Research Data Management for Librarians’ by Sarah Jones, Marieke Guy and Miggie Pickton: http://www.dcc.ac.uk/training/rdm-librarians
Research Data
What is research data?

According to the MANTRA project . . .

“Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results.”
What is research data?

According to the MANTRA project . . .

‘Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results.’

According to the Scientific Data Application Profile Scoping Study . . .

‘The evidence base on which academic researchers build their analytic or other work, where this evidence base is typically gathered, collated and structured according to declared and accepted protocols.’
What is research data?

According to the MANTRA project . . .

“Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results.”

According to the Scientific Data Application Profile Scoping Study . . .

“The evidence base on which academic researchers build their analytic or other work, where this evidence base is typically gathered, collated and structured according to declared and accepted protocols.”

According to EPSRC . . .

“Recorded, factual material commonly retained by and accepted in the [research] community as necessary to validate research findings.”
What is research data, really?

So many things can be research data:

- Simulation data, models and software
- Survey results and interview transcripts
- Instrument measurements
- Machining/measurement parameter data
- Still images, video and audio
- Experimental observations and field notes
- Text documents, spreadsheets, databases
- System log files
- Specimens, samples, slides, artefacts,
- Sketches, diaries, lab notebooks . . .

Research data is not a type of thing, it is a type of use.
ERIM Data Record Types

Data Object (physical or digital)

- Data Record
  - Research Object Data Record
  - Experimental Apparatus Data Record
  - Research Data Record
  - Context Data Record
    - Associative Data Record
    - ...
- Context Data Object
  - Unintentional Context Data Object
  - Intentional Context Data Object (e.g. Context Data Record)
Importance of research data

Intelligently open data is

➤ **Accessible**: easy to find

➤ **Intelligible**: others should be able to understand it

➤ **Assessable**: both on its own merits and in context of authors’ interests

➤ **Usable**: in a sensible format, properly documented

“Public communication of scientific knowledge should not simply disclose conclusions but also communicate the reasoning and evidence that underlie them.”
Reinhart and Rogoff (2010) said public debt at 90% GDP was tipping point for recession. BUT

- contrary data excluded
- contrary data given less weight
- coding error

Source: Mike Konczal
2002: Jan Hendrick Schön found to have falsifying/fabricating results in 17 papers in solid state physics.

2006: Woo Suk Hwang found to have faked data relating to cloning and stem-cell research.

2011: Diederik Stapel found to have fabricated data underlying 30 peer-reviewed papers in clinical psychology.
Research Data Management
Timeline

- **mid-1990s**: NERC requires sharing of data funded by its grants
- **2000**: ESRC requires sharing of data funded by its grants
- **2004**: OECD *Declaration on Access to Data from Public Funding* – UK signs up; UNESCO *Policy Guidelines for the Development and Promotion of Governmental Public Domain Information*
- **2005–6**: RCUK issues a position statement
- **2007**: BBSRC, MRC, Wellcome require sharing of data funded by its grants
- **2008**: RIN *Stewardship of Digital Research Data*; AHRC requires sharing of data funded by its grants
- **2011**: STFC requires sharing of data funded by its grants; EPSRC requires institutions to perform RDM
What is research data management?

"the active management and appraisal of data over the lifecycle of scholarly and scientific interest."

– Digital Curation Centre
Data collection

- Gaining consent, contractual permissions, licences
- Collecting data
- Performing quality assurance and control
- Recording data in useful formats

Flowchart:
- Create/Receive
  - Document
  - Use
  - Store
  - Share
  - Preserve
  - Dispose
Documentation

Recording all information needed to understand and use the data

- at the time – it’s harder to do later
- according to metadata standards
- including local file naming, version control or structural conventions
Access and use

- Making analysis and processing reproducible: appropriate software, recorded steps
- Using secure, robust collaboration systems
- Version control: synchronisation, single working copy
Storage and backup

- Using managed storage rather than removable media
- Backing up data so there are
  - at least 3 copies of a file
  - on at least 2 different media
  - with 1 copy offsite
Appraisal and selection

Deciding what data should be published, what data should be in a dark archive, and what should be discarded

- What must be kept (legally, contractually, for scientific integrity)?
- What must be destroyed (legally, contractually)?
- What is of particular scientific, historical or economic value/interest?
- What is unique, or cannot be reproduced?
- Is there enough documentation to make keeping the data worthwhile?
Data deposition

- Selecting an appropriate repository: Databib, Re3data
- Packaging data in an appropriate submission format
- Obtaining a persistent ID (DOI, Handle) for the data
- Linking data to papers it supports
Research Data Management Policy
EPSRC Expectations

1. **Research organisations (ROs)** to raise awareness of data sharing responsibilities and issues.

2. Publications should link to underlying data.

3. **ROs** must keep track of their research datasets and requests for them.

4. Born-analogue data must also be shareable on request.

5. **ROs** must provide open, online catalogues of their data; digital data must be given a robust ID.

6. Access restrictions should be clear and justified.

7. **ROs** must provide access to data for 10 years from last access.

8. **ROs** must curate their research data.

9. **ROs** must pay for this from their existing public funding streams.
Statement of Commitment (2009)

"The University of Oxford is committed to supporting researchers in appropriate curation and preservation of their research data, and where applicable in accordance with the research funders’ requirements. It recognises that this must be achieved through the deployment of a federated institutional data repository."

http://www.ict.ox.ac.uk/odit/projects/datamanagement/

Policy on the Management of Research Data and Records (2012): 12 points, detailing with expectations for research data, retention periods, roles and responsibilities, etc.

http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf
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. ...
Data management is an essential and integral part of the responsible conduct of research. The University is responsible for:

1. ensuring effective data management to meet internal and external requirements, including enabling the re-use of research data and freely available public access to research data outputs in accordance with national and funding body policies;

2. retention of research data in sufficient detail for a defined period to enable appropriate responses to any questions about accuracy, authenticity, primacy and compliance with legal and regulatory requirements governing the conduct of research;

3. for supporting investigation into any allegations of misconduct or regulatory breach

http://sitem.herts.ac.uk/secreg/upr/IM12.htm
Data Asset Framework

http://data-audit.eu/

Stage 1.
Planning the audit

Stage 2.
Identifying and classifying data assets

Stage 3.
Assessing the management of data assets

Stage 4.
Reporting results and making recommendations
CARDIO Pulse Check

http://cardio.dcc.ac.uk/quiz

Take our quick survey to check whether your institution has its finger on the pulse of data management activity.

**Question 1 of 10**

What do you think are the risks associated with poor data management?

- A: Our institution is fully aware of the potential risks associated with poor research data management. We view research data management as part of good research practice which underpins our institutional reputation. We have taken concrete steps to provide our researchers and support staff with a working environment that fosters good research data management practice.

- B: We know that research data management is important for maintaining our institutional reputation and are keen to minimise any risks associated with poor data management. We are currently working to identify our main risks and to develop mitigation strategies.

- C: We'd like to understand more about the risks associated with poor data management so that we can assess our infrastructure and identify areas that could be improved.

Reset Quiz
- Register for free
- Find out more

Back to top
# CARDIO statements

## Organisation
1. Data Ownership and Management
2. Data Policies and Procedures
3. Data Policy Review
4. Sharing of Research Data/Access to Research Data
5. Preservation and Continuity of Research
6. Internal Audit of Research Activities
7. Monitoring and Feedback of Publication
8. Metadata Management
9. Legal Compliance
10. Intellectual Property Rights and Rights Management
11. Disaster Planning and Continuity of Research

## Technology
1. Technological Infrastructure
2. Appropriate Technologies
3. Ensuring Availability
4. Managing data integrity
5. Obsolescence
6. Managing technological change
8. Security Processes
9. Metadata tools
10. Institutional Repository

## Resources
1. Data Management Costs and Sustainability
2. Business Planning
3. Technological Resources Allocation
4. Risk Management
5. Transparency of Resource Allocation
6. Sustainability of Funding for Data Management and Preservation
7. Data Management Skills
8. Number of Staff for Data Management
9. Staff Development Opportunities

http://cardio.dcc.ac.uk/
Research Data Management Planning
**Recommendation 9.** Each funded research project, should submit a structured Data Management Plan for peer-review as an integral part of the application for funding.

Liz Lyon (2007), *Dealing with Data: Roles, Rights, Responsibilities and Relationships* (University of Bath)
Why?

Writing and using a Data Management Plan helps

- to co-ordinate the actions of data stakeholders
- to ensure all necessary tasks are accomplished
- to ensure data are properly curated
- with releasing data in a timely fashion
- with sharing data as openly as possible
- with preserving data for future use
DMPonline allows researchers to

1. create, store and update Data Management Plans
2. meet both institutional and funders’ data-related requirements
3. receive specific guidance from funders and institutions
4. export Data Management Plans in various formats
How DMPonline works

Create a plan based on relevant funder/institutional templates ...

...and then answer the questions using the guidance provided
Supporting researchers with DMPs

Various types of support could be provided by libraries:

- Guidelines and templates on what to include in plans
- Example answers, guidance and links to local support
- A library of exemplar DMPs
- Training courses and guidance websites
- Tailored consultancy services
- Online tools (e.g. customised DMPonline)
Research Data Management in Practice
Key documentation

Should be able to find quickly:

- Data management plan/record
- Project proposal (pre-award)
- Detailed project plan (post-award)
- Project record manifest
- Confidentiality agreements
- IPR statements, licences
**REDm-MED**

*Added by Alex Ball, last edited by Alex Ball on May 22, 2012 (view change) show comment*

**Principal investigator:** Chris McMahon

**Location on X drive:** file://localhost/X/MechEng/Research/General/CAM/CAM-0007/

**Data Management Plan:** n/a (but if we had one it would be attached)

**Links**
- Project web site
- Project blog

**Description**

The Research Data Management for Mechanical Engineering Departments (REDm-MED) Project is one of a number of projects being funded through Phase 2 of the JISC Managing Research Data Programme. The project will scope, specify, design and implement a research data management plan suited especially to the needs of the Department of Mechanical Engineering at the University of Bath. This work will build upon the team’s work investigating the research data management needs of the Innovative Design and Manufacturing Research Centre (IdMRC) at the University of Bath during the JISC-funded ERIM Project and other recent work by the Managing Research Data Community.

Work completed in the ERIM Project in specifying a Research Activity Information Development (RAID) Associational Tool will be continued by the development of a prototype tool for integration within the specified research data management infrastructure. The presumption is made that the research data management needs of the participating department will intersect closely with those of similar departments in other research institutions, not only of higher education but in the wider commercial world. Thus, the aim of the research will be to implement effective and practical research data management where there is little or none and that can be adopted easily elsewhere.

The work will take place in conjunction with the DCC at Bath, and with the support of the University of Bath's Department of Mechanical Engineering and the Bath University Computer Service (BUCS). The REDm-MED project will be complementary to the University of Bath’s JISC-funded Research 360 Project.
Engineering Data

- Different types of data each time
- Mix of common and obscure/proprietary formats
- Mix of confidential and unrestricted data
- Very hard to look at a directory of data files and know what it all means
# Project Data Record Manifest Template for IdMRC Projects

The Project Data Record Manifest (PDRM) constitutes the principal conduit through which the records relating to a research project may be identified and retrieved. It must be located in a publicly accessible and searchable place. The default location is an anonymous log-in page of the research project wiki.

The Project Data Management Plan and the Project Data Record Manifest should be considered as a pair, and should be co-located.

Whilst the PDRM will be globally available, there will be some records associated with the research project which are confidential or sensitive. Access to records of this nature must be limited by placing the records in appropriately password-protected locations, this could be BUCS file space or within the research project wiki or other web spaces. If in doubt, the advice of the data manager (or failing that, the project PI) should be sought.

## Summary of Research Activity

<table>
<thead>
<tr>
<th>Project name</th>
<th>e.g. Long And Technical Textual Evaluation (LATTE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period of Project</td>
<td>e.g. October 2009 – March 2011</td>
</tr>
<tr>
<td>Lead and partner organizations</td>
<td>e.g. University of Bath (lead), University of Cambridge, University of Leeds</td>
</tr>
<tr>
<td>Principal Investigator (name and contact details)</td>
<td></td>
</tr>
<tr>
<td>Name:</td>
<td></td>
</tr>
<tr>
<td>Contact details:</td>
<td></td>
</tr>
</tbody>
</table>

## Data access summary

Data access refers to the physical means by which access to records is constrained. The overarching data access provisions for this research project are recorded in the DMP associated with this PDRM for details of confidentiality status and data access provisions.

Data access refers to the physical means by which access to records is constrained. As a guide, data access should be either consistent with or more restrictive than the confidentiality status.

## Receiving repository

- e.g. The data from this Research Activity will be deposited according to the IdMRC DMP (see below).
- or
- The data from this research activity will be deposited in …...

## Related documentation

- RCUK Policy and Code of Conduct on the Governance of Good Research Conduct
- The University of Bath Good Practice Guide for Research
- Engineering Research Data Management Plan Specification
- IdMRC Projects Data Management Plan

## Project Management Documentation

Note that some of these records may need to be placed in a password-protected storage area.

- Project Data Record Manifest [wiki link]
- Project Proposal [wiki link]
- Project Plan [wiki link]
- Participant consent forms [wiki link], (physical location/contact name/contact details)
- Ethical forms [wiki link], (physical location/contact name/contact details)
- IPR Statement [wiki link], (physical location/contact name/contact details)
- UK Data Archive deposit requirements [wiki link]

## Project Data Record List

Every project data record should be listed in the table below in the form:

<table>
<thead>
<tr>
<th>Record Title</th>
<th>File Name</th>
<th>Owner</th>
<th>Contact Details</th>
<th>Data Record Type</th>
<th>Confidentiality Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example:</td>
<td>IdMRC Project Data Record Manifest</td>
<td>erim6man110217mjd</td>
<td>Mansur Darlington</td>
<td><a href="mailto:assnsd@bath.ac.uk">assnsd@bath.ac.uk</a></td>
<td>IdMRC data record</td>
</tr>
</tbody>
</table>

## History of this PDRM
Project Record Manifest template

Record Type (for both electronic and physical records)

Every data record will be one of the following: research data record, context data record, associative data record, research object data record, experimental apparatus data record.

Location

If all the files are archived in a single, central location, the location need be identified for the set of records (the Data Case) only. For electronic records it is expected that a hyperlink or filepath to the location is recorded. For physical records the location should be described.

Owner

The ‘owner’ is the person currently responsible for the management of the record, and who is in a position to consider matters such as shareability and security. Ownership does not imply any rights to use or disposal. During the period that the research project is under way it is likely that the owner will be a research officer or an individual in a supervisory rôle. At project end the ownership should be transferred to an appropriate individual, such as the project PI or the data manager responsible. In many cases it will be appropriate for a research officer to retain ownership.

Confidentiality Status

Confidentiality status indicates what classes of people and what automated information-gathering systems may have sight of the data record; it does not provide information about how such records are protected. It is likely that the confidentiality status will change during the life-cycle of the data record, in which case the status must be updated. Access is either free or limited. If access is free, then the term ‘public domain’ should be used. If the access is limited, then the entities who are permitted to see this data should be identified either by naming groups or individuals.

<table>
<thead>
<tr>
<th>Record Title</th>
<th>File Name</th>
<th>Owner</th>
<th>Contact Details</th>
<th>Data Record Type</th>
<th>Confidentiality Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IdMRC Research Project Data Record Manifest</td>
<td>erim6man110217mjd</td>
<td>Mansur Darlington</td>
<td><a href="mailto:ensmd@bath.ac.uk">ensmd@bath.ac.uk</a></td>
<td>associative data record</td>
<td>public domain</td>
</tr>
</tbody>
</table>
Theoretical calculations

<table>
<thead>
<tr>
<th>Data Case for CRYMAN (extract)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical calculations</strong></td>
</tr>
<tr>
<td>force_calculations.xls:</td>
</tr>
</tbody>
</table>
| `m = "Numerical"
  o = "Research generated"
  d = "Cutting parameters" | `m = "Numerical"
                           o = "Research generated"
                           d = "Depth of cut choices" |

Derive

| «datastore» spec_cut_energy.doc: |
| Data Record |
| m = "Text"
  o = "Pre-existing"
  d = "Specific cutting energy research" |

| «datastore» mat_stiffness.doc: |
| Data Record |
| m = "Text"
  o = "Pre-existing"
  d = "Material stiffness research" |

Generate

| «datastore» 21-3224-4576b.jpg: |
| Data Record |
| m = "Image"
  o = "Research generated"
  d = "Removed material photo" |

| «datastore» 3A-4.tif: |
| Data Record |
| m = "Image"
  o = "Research generated"
  d = "Removed material SEM images" |
RAIDmap

http://sourceforge.net/p/raidmap
**Cameron's LabLog**

*The online open laboratory notebook of Cameron Naylor*

**Some analysis of reflection data from Crisp Experiment**

*10th December 2009 @ 11:36*

**Procedure: Data, Analysis**

I have previously done single layer fits of data from these experiments for the following contrasts:

- H-tails DDO
- D-tails NW
- D-tails 30% DDO
- 1:3 tails NW
- 1:1 tails NW

These show a reasonably reproducible thickening of the layer for the three surface pressure regimes.

On moving to a two layer model with the 7 mm data the fits really want to minimise the headgroup thickness of the headgroup doesn’t seem to effect the fit much. Some issues with SLD for the mixed tails, similar to what was seen previously with the single layer fits.

With the parameters taken from a two layer fit of the 7 mm data, the predicted NR fits kind of ok by eye to data from the D-Tail DDO contrast with no further fitting (chisq~2 around 7%) for six data sets). Further fitting leads to a very thin (~1.5 A) layer with high hydration suggesting a very minimal headgroup layer at this surface pressure.

Moved on to ~14 mm data and was getting very similar results until I realized I had the SLD for the heads fixed at the wrong value!?? On fixing this I can get reasonable fits with the correct SLD and a layer thickness of 2 A.

Images: Simon Coles

http://www.mylabnotebook.ac.uk/
RDM and libraries
How are libraries engaging in RDM?

- Making a business case for RDM to senior managers
- Defining institutional RDM strategy
- Developing institutional and departmental RDM policy
- Delivering training courses
- Helping researchers to write DMPs
- Advising on data sharing and citation
- Setting up data repositories and catalogues
- Auditing the institution’s data holdings
- ...
Librarians taking the lead

In the UK, libraries are taking the lead in institutional RDM initiatives, because

- they have a highly relevant skill set:
  - they have knowledge of information management, metadata, etc.
  - they have experience of teaching information literacy
  - they already run publication repositories
  - they have proven liaison and negotiation skills
- they have good relationships with researchers
- they have existing open access leadership roles
Next steps

Digital Curation Centre

- Briefing papers:
  http://www.dcc.ac.uk/resources/briefing-papers

- How-to guides:
  http://www.dcc.ac.uk/resources/how-guides

Training materials aimed at librarians

- MANTRA:
  http://datalib.edina.ac.uk/mantra/libtraining.html

- RDMRose Lite:
  http://rdmrose.group.shef.ac.uk/?page_id=364

- SupportDM:
  http://www.uel.ac.uk/trad/outputs/resources/
Thank you for your attention

DCC Website: http://www.dcc.ac.uk/
Alex Ball: http://www.ukoln.ac.uk/ukoln/staff/a.ball/