



Citation for published version:

Ball, A 2013, 'Tackling Challenges in Research Data Management' Research Data Management User Group Launch Event, Leicester, UK United Kingdom, 21/01/13, .

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Publisher Rights
CC BY

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



because good research needs good data

Tackling Challenges in Research Data Management

Alex Ball

DCC/UKOLN, University of Bath

21 January 2013

Charles Wilson Building, University of Leicester



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

Funded by **JISC**



Who we are

The Digital Curation Centre (DCC) is a collaboration between

- ▶ University of Edinburgh
- ▶ HATII, University of Glasgow
- ▶ UKOLN, University of Bath

Key facts

- ▶ Funded by JISC
- ▶ Started in March 2004
- ▶ Hub of expertise in curating digital research data
- ▶ Observe, reach out, innovate, support JISC



Outline

Planning institutional readiness for research data management

Planning for research data management at the project level

Monitoring data-related activity

Licensing data

Making data citable

More Information

Planning institutional readiness for research data management



EPSRC Policy Framework on Research Data

The logo for EPSRC, consisting of the letters 'EPSRC' in a bold, purple, sans-serif font. The letters are underlined by two horizontal teal lines, one above and one below the text.

Engineering and Physical Sciences
Research Council

- EPSRC expects all those [research organisations] it funds to have developed a **clear roadmap** to align their policies and processes with EPSRC's expectations by 1st May 2012, and to be **fully compliant** with these expectations by 1st May 2015. •

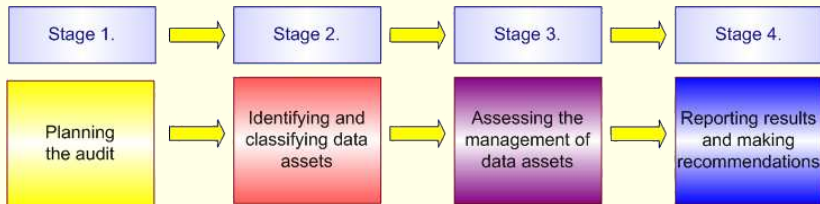
EPSRC Expectations

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.

Data Asset Framework



<http://data-audit.eu/>



CARDIO Pulse Check



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

CARDIO Pulse Check

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.

Next

Reset Quiz

-  Register for free
-  Find out more

Back to top 



CARDIO Process

Organisation	Technology	Resources
<ol style="list-style-type: none">1. Data Ownership and Management2. Data Policies and Procedures3. Data Policy Review4. Sharing of Research Data/Access to Research Data5. Preservation and Continuity of Research6. Internal Audit of Research Activities7. Monitoring and Feedback of Publication8. Metadata Management9. Legal Compliance10. Intellectual Property Rights and Rights Management11. Disaster Planning and Continuity of Research	<ol style="list-style-type: none">1. Technological Infrastructure2. Appropriate Technologies3. Ensuring Availability4. Managing data integrity5. Obsolescence6. Managing technological change7. Security Provisions8. Security Processes9. Metadata tools10. Institutional Repository	<ol style="list-style-type: none">1. Data Management Costs and Sustainability2. Business Planning3. Technological Resources Allocation4. Risk Management5. Transparency of Resource Allocation6. Sustainability of Funding for Data Management and Preservation7. Data Management Skills8. Number of Staff for Data Management9. Staff Development Opportunities

<http://cardio.dcc.ac.uk/>

Planning for research data management at the project level

Data Management Plans

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)



Dealing with Data: Roles, Rights, Responsibilities and Relationships

Consultancy Report

Document details

Author:	Dr Liz Lyon, UKOLN, University of Bath
Date:	19 th June 2007
Version:	V1.0
Document Name:	data-consultancy-report-final.doc
Notes:	

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



DMP Online allows researchers to

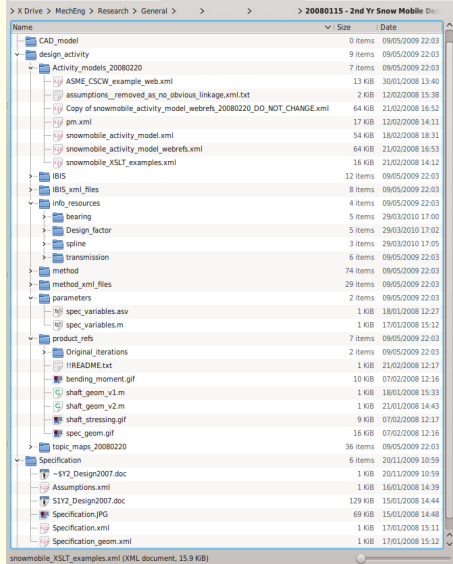
1. create, store and update Data Management Plans
2. meet both institutional and funders' data-related requirements
3. receive specific guidance from funders and institutions
4. export Data Management Plans in various formats

Monitoring data-related activity

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 Record Manifest template

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 reviewed. 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 a pair, and should be co-located.

The PDRM should be 'read-only', editing rights being limited to members of the originating research project team and by other nominated individuals such as the data manager. A versioning system must be in force.

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 spaces or within the research project wiki or other web space. If in doubt, the advice of the data manager (or failing that, the project PI) should be sought.

Summary of Research Activity

Project name
e.g. Long And Technical Textual Evaluation (LATTE)
Period of Project
e.g. October 2009 – March 2011
Lead and partner organizations
e.g. University of Bath (lead), University of Cambridge, University of Leeds
Principal Investigator (name and contact details)
Name: Contact details:
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 OAMP associated with this PDRM, for details of confidentiality status of individual records see the Project Data Record List below. As a guide, data access should be either consistent with or more restrictive than the confidentiality status.
Retrieving repository
e.g. The data from this Research Activity will be deposited according to the IdMRC OAMP (see below), or The data from this research activity will be deposited in.....
Related documentation
<ul style="list-style-type: none">• 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]
- Confidentiality agreement with (name) [wiki link: note if this agreement is itself confidential it should be placed in an appropriately protected location]
- Participant consent forms [wiki link] [physical location/contact name/contact details]
- Ethics form(s) [wiki link] [physical location/contact name/contact details]
- IP/3rd Streamline [wiki link] [physical location/contact name/contact details]
- UK Data Archive deposit requirements [wiki link]

Project Data Management Documentation

- Project Data Management Plan [wiki link] (this will be a reciprocal association, since the PDMP will identify the Project Data Record Manifest)
- RAD record(s) [wiki link] or
- Other data record associative documents [wiki link]

Project Data Record List

Every project data record should be listed in the table below in the form: Title, file name, record type, location, owner and contact details; confidentiality status

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 role. 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 to 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.

Record Title	File Name	Owner	Contact Details	Data Record Type	Confidentiality Status
Example:					
IdMRC Research Project Data Record Manifest	idmrcmanifest/102/7/mf	Manzur Darlington	amanzd@bath.ac.uk	associative data record	public domain

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

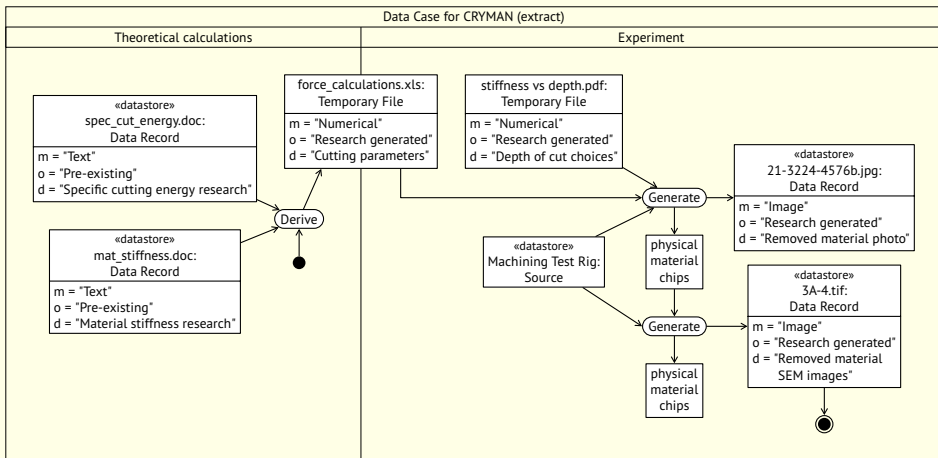
The 'owner' is the person currently responsible for the management of the record, and who is in a position to consider matters such as shareability and security. Ownership does not imply any rights to use or disposal. During the period that the research project is under way it is likely that the owner 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.

Record Title	File Name	Owner	Contact Details	Data Record Type	Confidentiality Status
<i>Example:</i>					
<i>IdMRC Research Project Data Record Manifest</i>	<i>erim6man110217mjd</i>	<i>Mansur Darlington</i>	<i>ensmd@bath.ac.uk</i>	<i>associative data record</i>	<i>public domain</i>

Example RAID diagram



RAIDmap

The screenshot displays the RAIDmap application window titled "RAIDMap: Derby \ Localhost \ Default \ ERIM". The interface includes a menu bar (File, Edit, View, Tools, Bookmarks, Workspaces, Window, Help), a toolbar, and a "Stencils" panel on the left with various icons and buttons like "Aggregate", "Migrate", "Collect", "Generate", "Annotate", "Populate", and "Aggregate".

The main workspace shows a diagram titled "[Map]: ERIM_Project" with the URL <http://sourceforge.net/p/raidmap> displayed at the top. The diagram illustrates four input nodes labeled "rig7_exp1b", "rig7_exp1a", "rig7_exp2", and "rig7_exp4", each with a red "RDR" icon. Arrows from these nodes point to a central "Aggregate" node labeled "Rig outputs 1a, 1b, 2 & 4". A separate window titled "[Map]: Rig outputs 1a, 1b, 2 & 4" is also visible, showing a similar diagram with the "Aggregate" node highlighted.

The bottom status bar shows "Data Connection: Derby: Default", "10 secs", and a "Start" button. The text "Rig outputs 1a, 1b, 2 & 4" is visible in the bottom left corner of the application window.

Smart Research Framework

The screenshot shows a web browser window with the following elements:

- Address Bar:** <http://blogs.blogmydata.org/feeds/higem/uid/588e7f94d05a8e04d2>
- Page Title:** HIGEM Sandpit
- Content Area:**
 - North Atlantic Data** (HIGEM)
 - Cameron's LaBlog**: The online open laboratory notebook of Cameron Noyles
 - Some analysis of reflection data from Crisp Experiment**
 - Procedure: Data_Analysis**
 - H-1 tails O2D**
 - D-1 tails NW**
 - D-1 tails 30% O2D**
 - I-1 tails NW**
 - I-1 tails NW**
 - Text:** 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 mN data the fits really want to minimize the headgroup thickness. Hydration of the headgroup doesn't seem to affect the fit much. Some issues with SLO 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 mN data, the predicted NR fits kind of ok by eye to data from the D-Tail O2D contrast with no further fitting (DH+2 around 70 for six data sets). Further fitting leads to a very thin (~1.5Å) layer with high hydration, suggesting a very minimal headgroup layer at this surface pressure.
 - Text:** Moved on the ~14 nm data and was getting very similar results until I realized I had the SLO for the heads fixed at the wrong value?!? On fixing this I can get reasonable fits with the correct SLO and a layer thickness of 2 Å
 - Text:** After playing around with some roughness it doesn't really help. Hydration tends to go to 100% with a thinner layer. Still not well determined by the data though.
- Map:** A map of the North Atlantic Ocean with a red dot indicating the location of the experiment.
- Table:**

Sample	Frame	Exposure time	I22 Run #	Data
				Post Id:
				Post Id:
				Post Id:
				Post Id:
				Post Id:
				Post Id:
				Post Id:

- Form:** A form to add a post from a template, with fields for Title (SAXS on I22 template) and Text (following samples were run in 1 mm capillary cells on I22. Detector was at 5m and q range standardised against silver behenate).

Images: Simon Coles

<http://www.mylabnotebook.ac.uk/>

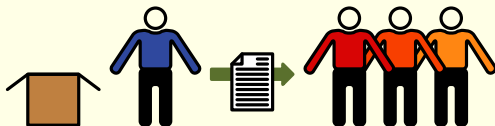
Licensing data

Types of licenses

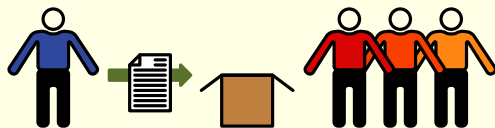
▶ Contracts



▶ Pure licences



▶ Waivers



Licensing questions

1. Do I need to make a choice?
 - ▶ Institutional policy
 - ▶ Data archive policy
2. If so, would a standard licence suffice?
3. If not, how do I write my own licence?
4. Do I need more than one licence?

Making data citable

Key citation elements

- ▶ Author
- ▶ Date made available
- ▶ Title
- ▶ Publisher/host
- ▶ Location (= identifier)

More Information

Literature

DCC How-to Guides:

<http://www.dcc.ac.uk/resources/how-guides>

- ▶ How to Cite Datasets and Link to Publications
- ▶ How to Develop a Data Management and Sharing Plan
- ▶ How to License Research Data
- ▶ How to Develop Research Data Management Services (in preparation)

ANDS guides: <http://ands.org.au/guides/#datamanagement>

- ▶ Creating a Data Management Framework
- ▶ Data Management Planning
- ▶ Ethics, Consent and Data Sharing
- ▶ Storage



because good research needs good data

Thank you for your attention

DCC Website: <http://www.dcc.ac.uk/>

Alex Ball: <http://www.ukoln.ac.uk/ukoln/staff/a.ball/>

