WORKSHOP 1: 
Open Research Data in Social Sciences and Humanities (ADP)

June 18th 2014, Social Science Data Archives, Faculty of Social Sciences, University of Ljubljana
Introduction to RDM and emerging good RDM practice in UK universities

Joy Davidson, Digital Curation Centre
The DCC Mission

“Helping to build capacity, capability and skills in data management and curation across the UK’s higher education research community”

Phase 3 Business Plan

www.dcc.ac.uk
What is research data management and curation?

• Active management, adding value and maintaining access to research data over the course of its scholarly lifecycle.

• Part of good research practice
UK Context
What has been driving this agenda in the UK?

• Several factors making RDM a key issue for many UK HEIs
• These factors are international and are affecting research practice in many countries (US, Canada, EU)
“Data sets are becoming the new instruments of science”

Dan Atkins, University of Michigan
Digital data as the new special collections?

Sayeed Choudhury, Johns Hopkins
Research data: institutional crown jewels?

Give us back our crown jewels

Our taxes fund the collection of public data - yet we pay again to access it. Make the data freely available to stimulate innovation, argue Charles Arthur and Michael Cross.

Charles Arthur and Michael Cross
The Guardian, Thursday 9 March 2006
Article history

http://www.flickr.com/photos/lifes_too_short_to_drink_cheap_wine/4754234186/
What do UK funders expect?
Ultimately funders expect:

• timely release of data
  - once patents are filed or on (acceptance for) publication

• open data sharing
  - minimal or no restrictions if possible

• preservation of data
  - typically 5-10+ years if of long-term value

See the RCUK Common Principles on Data Policy:
www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
<table>
<thead>
<tr>
<th>Research Funders</th>
<th>Published outputs</th>
<th>Data</th>
<th>Time limits</th>
<th>Data plan</th>
<th>Access/sharing</th>
<th>Long-term curation</th>
<th>Monitoring</th>
<th>Guidance</th>
<th>Repository</th>
<th>Data centre</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>BBSRC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>CRUK</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>EPSRC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>ESRC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>MRC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>NERC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>STFC</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
<td>🟢</td>
</tr>
</tbody>
</table>

MRC - Medical Research Council

The MRC position statement supports open and unrestricted access to published research.

It also has a policy on research data sharing. This policy was revised in September 2011.

POLICY STIPULATIONS

Time Limits

The MRC expects publications to be made openly accessible as soon as possible and in any event within six months of the journal publisher's official date of final publication.

Data should be made available in a timely and responsible manner. A limited defined period of exclusive use of data for primary research is reasonable.

Data must be securely maintained for a minimum of ten years after completion of the research.

Data Plan

Applicants are expected to submit a data management plan at the grant proposal stage.
Working with UK Universities
Institutional Engagement programme

- Pilot phase ran from Spring 2011 to Spring 2013
- Funded by HEFCE under Universities Modernisation Fund
- Provided up to 60 days of resource to 21 institutions
- Aimed to:
  - Increase capacity and capability in Research Data Management
  - Drive efficiencies in the HE sector by sharing models and lessons
The engagement model

- Efficient information sharing across HE sector
- Redesign of DCC support
Who we worked with
DCC tailored support

www.dcc.ac.uk/tailored-support
Roles of main participants

50% of the engagements have senior academic champions

Team leaders
- Library – 11
- Research office – 8
- IT services – 2

Major support
- Library – 7
- Research office – 2
- IT services – 6
- Records management – 2
- Repository management – 1
Examples of support: CARDIO at Queen Mary

- Advised on the mini quiz and full CARDIO
- Ran a workshop to reach consensus on results
- Wrote up findings to feed into strategy
- ‘heatmap’ approach being used to guide roadmap development

DCC support fits into the curation strand of the IT Transformations Project at QMUL
Fostering skills development
Guidance webpages

Data management support for researchers

Research data takes many forms, ranging from measurements, numbers and images to documents and publications. These webpages will help you plan, create, organise and look after your data.

Creating your data

Do you know what formats to use or how to manage IPR and ethics? Who can help you explain this in your data plan?

Organising your data

Can you name your files, folders and control the way that makes data easy to use?

Accessing your data

Can you and your collaborators access all of the files you need? How does the storage you choose affect options for remote access?

Looking after your data

How can you store and share data securely? Can the data be selected and used for the long-term?

www.gla.ac.uk/datamanagement

www.bath.ac.uk/research/data

POSSIBLE CONCERNS

 wondered how much detailed information should be included in a final report. The Act states that all information that is necessary for the implementation of the recommendations in the report should be included, and the constitution of the committee suggested that the Act should be amended to remove any requirement for a scientific report.
Online training for PhD students

Research Data MANTRA is a course designed for PhD students and others who are planning a research project using digital data.

The online learning units on the left cover a number of important topics. To begin, click on the first unit, Introduction to the course and page through.

After, work through one of our software-specific data handling practicals.

Feel free to view our project website, guidance for research staff, read some testimonials, find out who contributed content, or give us feedback on the course, using the menu along the top.

View an A5 leaflet in pdf or a conference poster about MANTRA.

This course is an Open Educational Resource that may be freely used by anyone. It is available through a Creative Commons open attribution license for re-using, rebranding, repurposing.

http://datalib.edina.ac.uk/mantra
Examples of support: Capacity building at UEL

• Co-designed a series of modules for training RDM support staff
• Developed online learning resource, presentations and exercises
• Delivered training with UEL colleagues
• Helped to build awareness and confidence in dealing with RDM

SupportDM: http://www.uel.ac.uk/trad/outputs/resources/

Review of the course: http://datamanagementuel.wordpress.com/2013/06/19/are-you-rdm-ready-from-zeroes-to-heroes/
Supporting data management planning
What is a DMP?

A short plan that outlines:

• what data will be created and how

• how it will be managed (storage, back-up, access...)

• plans for data sharing and preservation
Some other funders that require DMPs or equivalent
Five common themes

- Description of data to be collected / created
  (i.e. content, type, format, volume...)

- Standards / methodologies for data collection & management

- Ethics and Intellectual Property
  (highlight any restrictions on data sharing e.g. embargoes, confidentiality)

- Plans for data sharing and access
  (i.e. how, when, to whom)

  Strategy for long-term preservation
Main features of DMPonline

• Templates for different requirements (funder or institution)
• Tailored guidance (funder, institutional, discipline-specific etc)
• Ability to provide examples and boilerplate text
• Supports multiple phases (e.g. pre- / during / post-project)
• Granular read / write / share permissions
• Customised exports to a variety of formats
• API for systems interoperability
• Shibboleth authentication
Institutions can customise DMPonline

Profile local support via custom guidance and boilerplate text

Select / write desired questions

Add your logo, URL etc

www.dcc.ac.uk/news/customising-dmponline
Facilitating research data discovery
Archiving - external data centres

Research funders’ data centres...

Structured databases

Registries of international data centres

Disciplinary & community initiatives
Institutional data repositories

Not intended to replace national, subject or other established data repositories

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

www.dspace.cam.ac.uk

https://databank.ora.ox.ac.uk

Acknowledging hybrid environment

Research Data at Essex and DataPool at Southampton

https://databank.ora.ox.ac.uk
Data catalogues

• Oxford is developing its DataFinder tool 
  http://blogs.it.ox.ac.uk/damaro

• Research Data @ Essex has developed a profile based on DataCite, Inspire and DDI standards  
  www.data-archive.ac.uk/media/395364/rde_march2013_repositories.pdf

• C4D is developing a research data extension to the CERIF standard -  
  http://cerif4datasets.wordpress.com

• CKAN is being explored by various projects  
  http://ckan.org
Research Data Registry pilot

• aim is to improve data discovery and provide a national level service

• aggregate metadata relating to data collections or datasets held in UK HEIs and subject data centres

• have mapped metadata schema to develop crosswalks

• trialling ANDS Research Data Australia
Jisc RDRDS Pilot Partners

Data centres:
- UK Data Archive
- NERC Data Catalogue Service
  - BADC
  - BODC
  - EIDC
  - NEODC
  - NGDC
  - PDC
  - UKSSDC
  - ADS

Universities:
- Edinburgh
- Glasgow
- Hull
- Lincoln
- Leeds
- Oxford
- Oxford Brookes
- St Andrews
- Southampton
RDRDS Metadata Work

**DDI Codebook 2.5**
- UK Data Archive

**DataCite 3**
- Edinburgh (TBC)
- Oxford (TBC)
- Hull (TBC)

**OAI-PMH Dublin Core**
- Oxford Brookes (TBC)

**UK Gemini 2.2**
- NERC Data Catalogue Service

**EPrints 3**
- Glasgow
- Leeds
- Lincoln (TBC)
- Southampton
Metadata catalogues: defining levels
RDRDS ongoing investigations

- Would another platform suit us better?
- Would another internal metadata scheme suit us better than RIF-CS?
- What use cases should the registry target?
- How can we add value to the registry’s records?
- Could the registry add value to other systems?
Assessing RDM costs and benefits
<table>
<thead>
<tr>
<th>Research Funders</th>
<th>Policy Coverage</th>
<th>Support Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Published outputs</td>
<td>Data</td>
</tr>
<tr>
<td>AHRC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBSRC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPSRC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESRC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NERC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STFC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wellcome Trust</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies
Collaboration to Clarify the Costs of Curation (4C)

http://4cproject.eu/
Collaboration to Clarify the Costs of Curation (4C)

Comparison of costs by activity with others

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Comparison with all other users

<table>
<thead>
<tr>
<th>Activity</th>
<th>Average (€/TB)</th>
<th>Yours (€/TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingest</td>
<td>2400</td>
<td>2520</td>
</tr>
<tr>
<td>Archival Storage</td>
<td>450</td>
<td>500</td>
</tr>
<tr>
<td>Access</td>
<td>260</td>
<td>220</td>
</tr>
</tbody>
</table>

Segmented comparison

See the comparison of your costs with other users:
- Of the same organization type as yours
- That have data volumes of the same scale as your organization
- That have the same policy for number of copies
- ...

http://4cproject.eu/
Lessons we can share
Components of RDM services

- RDM policy and strategy
- Business plans and sustainability
- Data Management Planning
- Managing active data
- Data selection and handover
- Data repositories
- Data catalogues

Guidance, training and support
UK research data policies

“Statement of commitment”
- Infrastructure policy

“10 commandments”
Mutual promises aspirational

Baseline of RCUK Code + procedures & support

Legal tone / language a section in uni DM policy useful guide as appendix

Based on Edin. with a few additions
Sample roadmaps: University of Bath

- Based on Monash University RDM strategy
- Identifies the current position and proposes activity
- Defines roles and responsibilities and timeframes

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Actions</th>
<th>Milestones</th>
<th>Roles &amp; Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Develop the data management skills and knowledge of Bath researchers</td>
<td>• Initiate and coordinate a program of research data management (RDM) skills development for Bath researchers&lt;br&gt;• Work with Faculties, DTCs and the Bath Graduate Schools to explore embedding research data management skills development in coursework curricula and professional development offerings</td>
<td>• Pilot RDM skills development to new Bath researchers from September 2013&lt;br&gt;• High quality training delivered (participant feedback, impact notes)</td>
<td>• Associate Deans of Research&lt;br&gt;• Doctoral Training Centres and Graduate Schools&lt;br&gt;• Library/UKOLN/BUCS&lt;br&gt;• Learning &amp; Teaching Enhancement Office &amp; Researcher Development Unit</td>
</tr>
<tr>
<td>1.2 Develop the skills and knowledge of key professional services staff&lt;br&gt;  • Library&lt;br&gt;  • BUCS&lt;br&gt;  • RDSO</td>
<td>• Initiate and coordinate a program of research data management skills development for key professional services staff at Bath&lt;br&gt;• Seek opportunities for professional services staff to network and participate in communities of practice – internally, nationally and internationally&lt;br&gt;• Promote research data management career paths within professional support services where appropriate</td>
<td>• Pilot RDM skills development to key professional services staff from September 2013&lt;br&gt;• High quality training delivered (participant feedback, impact notes)&lt;br&gt;• Support Bath staff participation in DCC meetings (RDMF, ICC etc.)</td>
<td>• Library/BUCS&lt;br&gt;• RDSO&lt;br&gt;• Staff Development Unit&lt;br&gt;• Data Scientist</td>
</tr>
</tbody>
</table>

http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf
Architecture models

Diagram courtesy of Sally Rumsey, University of Oxford
Thanks - any questions?

DCC guidance, tools and case studies: www.dcc.ac.uk/resources

Follow us on twitter: @digitalcuration and #ukdcc