Introduction to Data Management Planning and Data Sharing

Joy Davidson
Digital Curation Centre
About this course

- Short presentations with exercises and discussion
  - Introduction to Data Management Planning
  - DMPonline overview and demonstration
  - Data management plan exercise
  - Customising DMPonline for your institution
Research Data Management Landscape
“Data sets are becoming the new instruments of science”

Dan Atkins, University of Michigan
Exceptions to copyright

*Reforms to copyright law come into force on 1st June*

The exceptions coming into force today will bring a range of benefits to a wide range of groups:

• Researchers will benefit from the introduction of the new text and data mining exception for non-commercial research, as well as the reforms to existing

• Libraries, archives and museums will now be better able to protect our cultural heritage and preserve their collections. The existing preservation exception has been expanded to cover all types of copyright work, and now applies to museums and galleries as well as libraries and archives.
Data Management Planning and Sharing
### Research funder data policies

[Table showing coverage and stipulations for various research funders.]

Find more information at [www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies](http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies)
What is data sharing?

“… the practice of making data used for scholarly research available to others.” [Wikipedia]

Who’s involved?
- the data sharer
- the data repository
- the secondary data user
- support staff
- research participants
- commercial partners
BENEFITS

- Avoid duplication
- Scientific integrity
- More collaboration
- Better research
- Increased citation

9-30% increase shown in study
Managing restrictions on sharing

**Ethics**
Balance data protection with data sharing
- Informed consent – cover current *and* future use
- Confidentiality – is anonymisation appropriate?
- Access control – who, what, when?

**IPR**
- Clarify copyright before research starts
- Consider licensing options e.g. Creative Commons
How to share research data

- Use appropriate repositories and data catalogues
  - http://databib.org
  - http://www.re3data.org/
  - Jisc/DCC research data registry (coming soon!)

- License the data so it is clear how it can be reused
  - www.dcc.ac.uk/resources/how-guides/license-research-data

- Make sure it’s clear how to cite the data
  - http://www.dcc.ac.uk/resources/how-guides/cite-datasets

- Consider publishing a data paper based on your DMP
  - http://metajnl.com/
What is a DMP?

A short plan that outlines

- what data you will create and how
- how you will manage it (storage, back-up, access…)
- plans for data sharing and preservation

DMPs are often submitted as part of grant applications, but are useful whenever you’re creating data.
Why develop a DMP?

- to help you plan to share your data appropriately
- to anticipate and avoid problems e.g. data loss
- to comply with funders requirements...
- to provide guidelines for everyone working on the project
They typically want a short (c.1-2pp) statement covering:

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation
Data sharing plan
Covering:
- Dataset
- Standards
- Metadata
- Preservation
- Data sharing
  - method
  - timescale
  - restrictions
  - agreements

Data Mgmt Plan
Covering:
- Data
- Data sharing
  - when?
  - where?
  - how?
  - restrictions?
- Preservation
- Resources

Data Mgmt Plan
Covering:
- Data
- Data collection
- Management
- Security
- Data sharing
  - Responsibilities
  - Related policies
- Resources
- Admin details
Some funders don’t ask for a DMP but still have expectations

Researchers applying for an HTA grant should consider data sharing in their proposals

Don’t prescribe sharing but expect researchers to consider and plan for it as appropriate

Follow DOH Research Governance Framework and MRC guidelines for Good Research Practice

www.hta.ac.uk/funding/troubleshooting/index.html
Tips for writing DMPs

- Start early - don’t wait until the last minute to plan!
- Don’t write the plan in isolation - seek advice from colleagues, ethics, IT, library, DP/FoI...
- Be realistic - base plans on available skills & support
- The plan will - and should - change over life of project
- Use plan as a communication tool with partners, funders and yourself!
Funders’ expectations of public access

“Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.”

RCUK Common Principles on Data Policy
http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
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](http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx)
“Research organisations will ensure that effective data curation is provided throughout the full data lifecycle, with ‘data curation’ and ‘data lifecycle’ being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated...”

...institutional responsibility

www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx
Research Data Management Costs
## Research funders’ policies

<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>

www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies
Key differences in policies

- Length of retention periods range from 3 years to in perpetuity but most funders ask for 10+ years

- Some penalties may be imposed for non-compliance (e.g., ESRC and NERC may withhold the final grant payment if data aren’t offered for deposit)

- Cancer Research UK states explicitly that it will NOT provide additional funds for RDM
What RDM cost can be included?

- **In-project (direct) costs:**
  - covers hardware, staff, expenses, costs of preparing data & metadata...

- **Post project (largely indirect) costs:**
  - existing services should be used where possible
  - where an institution is going to provide a data repository, costs should be met through FEC
  - outsourcing to a third-party is also an option
How should costs be included?

- In-project costs should be included in the direct costs for a project.
- Post-project costs could be direct (e.g. charges levied by data centres) but typically fall into indirects as universities should provide infrastructure to support RDM.
- The Justifications of Resources should, where possible, separate out the following RDM cost elements:
  - cost of collecting data
  - the cost of curating data
  - the cost of analysing data
  - the cost of preservation and sharing
Key messages

- Research data management should not be regarded as optional.
- DMPs should make clear what is provided and what activities are being charged against a grant.
- The cost of RDM is project-specific and entirely depends on the type of work.
- Be creative - it may be possible to set up small research facilities to recover the cost of RDM (e.g. similar to provision of HPC), possibly as a cross-institutional service.
Thanks – any questions?

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

Follow us on twitter:
@digitalcuration and #ukdcc