Research Data Management

Dr. Chris Tibbs
Research Data Officer
rdm@exeter.ac.uk
Open Data is a new paradigm in which research data are freely and openly shared, with full re-use rights. Open data helps to maintain research integrity and enables validation of results. Additionally, open data allows the public to access publicly funded research.

The main point to remember about the open data environment is that data preservation is key, however, to ensure that data are suitable for sharing, good research data management practices are essential.
Research data includes *all* of the information that is used for research, irrespective of its format.

Research data management can be broken down into 3 key aspects:
- good research data management requires advanced planning prior to the start of a project
- you need to implement effective day-to-day data management practices during a project
- you need a long-term strategy for data preservation after the project is completed
Importance of Good Data Management

- Data are the cornerstone of research
- Good quality data leads to good quality research
- To protect data from loss, destruction, corruption
- Ensure that data remain accurate and reliable
- Increase research productively
- Enables compliance with ethical codes, data protection laws, journal requirements and funder/institutional policies
**UoE Requirements – Key Points**

- Research data management is the joint responsibility between PI(s) and researcher(s)

- All research proposals must include data management plans

- Research data management is a valid cost for research proposals

- Researchers are responsible for ensuring that data are deposited in an appropriate repository after project completion

- Publications should include a short data access statement

Full details: [http://hdl.handle.net/10871/26168](http://hdl.handle.net/10871/26168)
Funder Requirements – Key Points

Funders are increasingly requiring researchers to meet certain data management criteria

- Submission of a technical or data management plan when applying for funding

- Open data sharing after project completion
  - Deposit data in a data repository
  - Minimal or no access restrictions

- Long term preservation of the data
  - Most funders require 10+ years

Further details on specific funder policies: http://v2.sherpa.ac.uk/juliet/
This table displays the major research funders’ policy coverage for a range of different aspects of research data management.

<table>
<thead>
<tr>
<th>Research Funders</th>
<th>Published outputs</th>
<th>Data</th>
<th>Time limits</th>
<th>Data plan</th>
<th>Sharing/access</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>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>Cancer Research</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>European Commission</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>

[http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies]
UKRI have a set of principles to guide the data policies of the 7 UK Research Councils.

UKRI Common Principles

- 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.

- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.

- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

http://www.rcuk.ac.uk/research/datapolicy/
Data Management Planning
Research Data Lifecycle

Type, format, volume of data, chosen software for long-term access, secondary data, file naming, structure, versioning, quality assurance processes.

Select retention period, repository choice, costs involved in long-term storage?

Make data publicly available (where possible) at the end of a project, license data, any restrictions on sharing, access controls?

Information needed for the data to be understood in future, metadata standards, methodology, definition of variables, format & file type of data.

Access restrictions, risks to data security, appropriate methods to transfer/share data, encryption, legal, ethical issues.

© Stuart Macdonald/EDINA. Used with permission.
A data management plan (DMP) is a basic statement describing how the research data will be managed throughout a project and beyond.

A data management plan is a valuable tool for navigating a research project, but should not be viewed as set in stone. In fact, it’s good practice to review your plan periodically, and revise or update it as necessary.
DMP Online is a web-based tool that contains data management plan templates for all of the major research funders, and provides guidance and suggestions for what to include. It also contains a data management plan checklist and a range of example data management plans.
What does DMP Online do?

A web-based tool that enables users to

i. Create, store, and update multiple versions of DMPs throughout the research lifecycle

ii. Collaboratively work on the DMP with multiple people able to access the DMP

iii. Meet a variety of specific data-related requirements (from funders, institutions, etc.)

iv. Get tailored guidance on best practice and helpful contacts, at the point of need

v. Customise, export, and share DMPs in a variety of formats in order to facilitate communications within and beyond research projects
Preserving And Sharing Data
In general, your research data should be FAIR. Findable means that the data should be registered in a searchable repository and be assigned a persistent identifier. Accessible means that the data must be available via a standardised protocol e.g., http, and that the metadata is available, even if the data themselves are not. Interoperable means that the data must be able to be seamlessly integrated with existing data and tools, and Reusable means that the data must be clearly licensed.
Although most funders require data to be preserved in a repository, not all of them provide a specific repository, simply allowing the data to be deposited in any appropriate repository. Subject based repositories are available, and can be found using the registry of research data repositories.
ORE is a showcase of the research outputs of the University.
 Depositing Data In ORE

Is the size of your data <2GB?

Yes

Deposit your data in ORE via Symplectic

No

Deposit your data in ORE using the ORE deposit tool

ORE
Datasets <2GB in size can be deposited in ORE via Symplectic in the same manner that is used for journal articles. Note that the dataset and associated journal article require 2 separate Symplectic records.
The ORE Deposit Tool can be used to deposit large datasets into ORE. If you wish to use the tool for the first time, please contact us.
If you want to share your data and allow others to use it, you need to license your data. A licence determines how others can use, modify, and distribute your data.
The University has an agreement in place with DataCite, through the British Library, and since the start of the year, all datasets deposited to ORE are allocated a DOI.
A data access statement is a short statement that should describe where and how the underlying research data can be accessed, ideally including a link to the data using a persistent identifier. It should also explain on what terms the data are available.
Example Data Access Statements

**No new data generated**
e.g., “This study did not generate any new data.”

**Openly available data in a repository**
e.g., “The research data supporting this publication are openly available from the University of Exeter’s institutional repository at: https://doi.org/10.24378/exe.XXXXX”

**Sensitive data with restricted access**
e.g., “Due to ethical concerns, the research data supporting this publication can only be made available to bona fide researchers subject to a data access agreement. Details of how to request access are available from the University of Exeter’s institutional repository at: https://doi.org/10.24378/exe.XXXXX”
Practical Steps…

1. Write/update a data management plan
2. Deposit research data in a repository
3. Include a data access statement in publications
Need to know more?

Contact:

Chris Tibbs – Research Data Officer
Pat Liebetrau – Research Support Manager

rdm@exeter.ac.uk

http://www.exeter.ac.uk/research/researchdatamanagement/