How Do I Manage My Research Data?

Dr. Chris Tibbs
Research Data Officer
rdm@exeter.ac.uk
Open Data

• **Open Data** = a new paradigm across all disciplines, providing benefits to individual researchers, institutions, and society

• Good research data management habits are essential to creating data that are suitable for sharing
What is Research Data Management?

- **Research data** = all information collected/created and used in research
  - structured data (e.g., databases, tables, etc.)
  - unstructured data (e.g., textual sources, images, audio recordings, personal notes, emails, etc.)

- **Data management** = effective management of the information lifecycle
  - planning for data
  - working with data
  - preserving and sharing data
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
Exeter 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
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/
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.

Data Management Planning
Research Data Lifecycle

Image: CC BY-ND 4.0 JISC
What is a Data Management Plan?

A DMP is a formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data in a data repository for long-term management and preservation.

- What data will be collected/created (format, types, and size) and how?
- How will the data be documented and described?
- How will data ethics and Intellectual Property be managed?
- What are the plans for data sharing and access?
- What is the strategy for long-term data preservation?
DMP Online is an online web-based tool, developed by the Digital Curation Centre (DCC), to help write data management plans.

https://dmponline.exeter.ac.uk

rdm@exeter.ac.uk
Funders That Use DMP Online

https://dmponline.exeter.ac.uk/public_templates
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. Request feedback on your DMP

vi. Customise, export, and share DMPs in a variety of formats in order to facilitate communications within and beyond research projects
Working With Data
Research data files and folders need to be organised in a systematic way to be identifiable and accessible to both you and colleagues and potential future users.

- Use dates in YYYYMMDD or YYYY_MM_DD format at the beginning of the file name
- Use underscores instead of blank spaces between words
- Don't use special characters: & , * % # ; * ( ) ! @$ ^ ~ ' { } [ ] ? < >
- Keep the file name as brief as possible
- Use meaningful abbreviations
- Order the elements of the file name from general to specific
- Keep the file name independent of its location
- Use appropriate file extensions
Version Control

Use a consistent method to identify and distinguish versions of data files.

This is extremely useful if you need to go back and find data from an earlier phase of the project.

Suggested strategies:
• use a sequential numbered system e.g., v1, v2
• Use version control software e.g., SVN, CVS, Git

“Piled Higher and Deeper” by Jorge Cham www.phdcomics.com
Data documentation includes:

• A basic description of the project
• A description of the methodologies, protocols, sampling techniques used to produce the data
• An explanation of the content of all of the data files and how they are related to each other
• A list of any software that is required to access/analyse the data
• A detailed list of all of the variables

Electronic notebooks are:

• Searchable by any word, while physical notebooks are usually only searchable by date
• Able to support collaboration from multiple researchers
• Easier to backup and share notes with collaborators
• Accessible anywhere on multiple devices.
• Convenient to transport
Storing And Protecting Your Data

Where to store your data?
- University U:Drive
- University OneDrive for Business
- University shared network drive
- Portal storage (external hard drives, memory sticks)

Data backup strategy
- Ensure that you have at least 3 copies of your data
- Store these copies on at least 2 different media
- Keep at least 1 of these in a physically different location

Data security
- Access controls
- Encryption

“Piled Higher and Deeper” by Jorge Cham www.phdcomics.com
Preserving And Sharing Data
FAIR Data

To meet the funder requirements, you should make your data:

Findable  Accessible  Interoperable  Reusable

Image: CC BY-SA 4.0 Sangya Pundir

https://www.force11.org/group/fairgroup/fairprinciples
Digital research data are best preserved and published using a research data repository.

A repository is an online database service, an archive that manages the long-term storage and preservation of digital resources and provides a catalogue for discovery and access.

Most data repositories do not charge to deposit research data, though many require registration.

Registry of research data repositories: [http://www.re3data.org/](http://www.re3data.org/)
Open Research Exeter (ORE)

ORE is the University of Exeter's institutional repository, suitable for both publications and data.

ORE offers long-term data preservation and curation.

https://ore.exeter.ac.uk/repository/
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 can be deposited to ORE via Symplectic using the same procedure as the one used to deposit journal articles – simply select the type “dataset” rather than “journal article”

https://researchpubs.exeter.ac.uk/
Depositing Data In ORE (≥2GB)

The ORE Deposit Tool can be used to deposit large datasets into ORE

https://ore.exeter.ac.uk/oredeposit/

Uses Globus Connect to transfer the data in the background without having to store the data on an intermediate server.
Licensing Your Data

Licences to share, remix, and distribute legally, while ensuring authors get credit for their work

Terms
- **Attribution (BY)**
  - Allows others to copy, display, remix, and redistribute the work as long as they provide appropriate credit to the creator.

- **ShareAlike (SA)**
  - Allows others to copy, display, and remix the work as long as they distribute any modified work on the same terms.

- **NoDerivatives (ND)**
  - Allows others to copy, display, and distribute only original copies of the work, without modification.

- **NonCommercial (NC)**
  - Allows others to copy, display, remix, and redistribute the work for non-commercial purposes only.

Licences
- CC BY
- CC BY-SA
- CC BY-ND
- CC BY-NC
- CC BY-NC-SA
- CC BY-NC-ND

Least Open

Most Open

https://creativecommons.org/licenses/
Digital Object Identifiers (DOIs)

To effectively share research data, it must be uniquely identified with a persistent identifier

All datasets deposited in ORE are allocated a DataCite DOI

University of Exeter DOIs are of the format: https://doi.org/10.24378/exe.XXXXX
Data Access Statements

All research publications should include a statement on how the underlying data can be accessed (a "data access statement").

- **How the data can be accessed?**
  - Web link
  - Digital Object Identifier (DOI) or other persistent identifier

- **Who must be contacted to request access?**
  - Departmental/group email address (not a personal email address)

- **On what terms are the data available?**
  - A general licence
  - A data sharing agreement must be entered into before access to the data is granted
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”
Why Archive Research Data?

https://archive.stsci.edu/hst/bibliography/pubstat.html
Further help?

Contact:
rdm@exeter.ac.uk

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