

Introduction to Research Data Management

Welcome to this introduction to Research Data Management. Each unit should take you no more than 10 minutes to complete.

We hope you find the course interesting and informative.

If you have any questions after completing the course, please email us at rdm@sheffield.ac.uk.

Learning outcomes

After completing this course you will be able to:

- relate the benefits of research data management to your own research;
- identify the key challenges you will face in managing your data;
- create a simple Data Management Plan;
- understand ethical issues around managing data;
- assess what further information and support you may require;
- source additional help and guidance.

Target Audience

These skills will be valuable for researchers at all stages of their careers; the course has been designed with postgraduate research students and early career researchers in mind.

There are no pre-requisites.

Unit 1: Looking after your data (8 sections)

1. Benefits of Research Data Management

Following good practice in Research Data Management will help you to maximise your data's reusability, accessibility, and discoverability. What do you think:

Could maximising your data's reusability, accessibility, and discoverability help you when consulting your data in future, by making it easier for you and saving you time?

... And by allowing a greater number of people to access and reuse your data, over a greater period of time?

... And by increasing your research profile and the number of potential collaborators and audiences?

2. The value of good Research Data Management and data sharing

Listen to these researchers and others explain why they believe research data management and research data sharing are of great value: [Vox Pop Research Data Management and Sharing \(https://youtu.be/yhVqImna7cU\)](https://youtu.be/yhVqImna7cU)

3. What do we mean by research data?

Research data are collected, observed, or created, for the purposes of analysis to produce and validate original research results.

It is vital you ensure your research data are accompanied by detailed documentation recording how they were generated or gathered, so as to maximise their reusability.

Therefore, you should seek to apply good practice in managing both your research data and any documentation necessary for interpretation and reuse of those data.

4. Some examples of research data & documentation

- an image of cells from a microscope
- a graph summarising gathered data
- a biochemist's Lab Book
- a photograph of a building for an architectural research project
- Matlab scripts used to process data
- a document describing a mathematical model
- a transcript of an interview with a survey subject
- the recording of the interview
- the survey questionnaire
- the survey codebook

5. Data Lifecycle

When considering what you'll need to do to manage your data well, it can be helpful to think of your research data as going through a lifecycle:

- Create
- Document
- Use
- Store
- Share
- Preserve

6. Requirements

Many research stakeholders now have data policies requiring researchers to demonstrate good practice in research data management and data sharing.

Q: What data management policies should you be aware of when considering your own research?

A: You should check the data policy of any journal you are hoping to publish in, your institution, and all your funders.

7. Services

- Research Data Management

The University of Sheffield Library Research Services Unit is here to help you with data management plans and depositing your research data: <https://www.sheffield.ac.uk/library/rdm/index>

- Online Research Data (ORDA)

Online Research Data (ORDA) is the hub for managing and sharing research data at The University of Sheffield: <https://orda.shef.ac.uk/>

8. Online resources

- MANTRA

A free, open educational resource to give researchers a grounding in Research Data Management: <http://mantra.edina.ac.uk/>

- Digital Curation Centre (DCC)

For detailed information and how-to guides on preserving and maintaining research data: <http://www.dcc.ac.uk/resources>

- UK Data Service

Advice on a wide range of aspects of data management, aimed particularly at social scientists: <https://www.ukdataservice.ac.uk/manage-data>

Unit 2: Finding and reusing data (7 sections)

1. Data discovery

It is a good idea to be open to the possibilities of using research data created by other researchers in your own work. This is called secondary use or reuse of data.

When you reuse data, it is very important to gather detailed information about the provenance of the data.

2. Licensing - 1

When you are thinking of reusing someone else's data, it is essential you find out whether it has been released under a licence which allows you to do so. You should consider the possibility that you may wish to analyse, reproduce, amend, and distribute new versions of their data, be they text, audio, video, or images.

3. Licensing - 2

If you can find no statement relating to rights over the data, you should work on the assumption that implicit copyright applies. In other words, if it doesn't say you can copy or reuse the data, then you may not do so without the permission of the data creator. Therefore, you should contact the data creator before copying or analysing their data.

4. Licensing - 3

Ask yourself: is it likely you will wish to share the results of your work with others in years to come through academic journals, posters at conferences, blogs, public engagement events, and so on? Will you need permission to do so?

5. Licensing - 4

You have found a photo on a website. You would like to annotate the photo and include it in a conference poster. Which of the following licence statements will allow you to do so?

- Creative Commons No-Derivatives (CC-ND)
Incorrect
Feedback: 'No-derivatives' means the data may not be altered.
- No statement
Incorrect
Feedback: When there is no statement, implicit copyright applies, so you may not copy the image, unless you get permission from the copyright owner.
- All rights reserved
Incorrect
Feedback: "All rights reserved" means the data creator has withheld permission for the image to be re-used.
- Creative Commons Attribution ('CC-BY')
Correct
Feedback: "CC-BY" means you may amend and share the data but you must cite the data creator.
- Open Data Commons Public Domain Dedication License
Correct
Feedback: If the data creator has released this image into the Public Domain then anyone may adapt and share it without restriction.

6. Citing data you have used as a source

When you reuse data created by others, make sure you cite the data properly. If a persistent identifier such as a DOI (Digital Object Identifier) or a handle is available, be sure to include it in your citation. A hyperlink containing a persistent identifier is more reliable than other URLs.

For example:

- Harding, J., Hall, S., Freeman, C., (2017). Simulations on calcite-protein interactions [Dataset]. Sheffield: ORDA. [Accessed 20 October 2017]. Available from: <https://doi.org/10.15131/shef.data.3491387.v1>
- Halcrow, M.A., Burrows, K.E., McGrath, S.E., Kulmaczewski, R., Cespedes, O., Barrett, S.A., (2017). Data to support study of Spin States of Homochiral and Heterochiral Isomers of [Fe(PyBox)2]2+ Derivatives [Dataset]. University of Leeds. [Accessed 20 Oct 2017]. Available from: <https://doi.org/10.5518/184>

7. Online resources

- re3data
The Registry of Research data Repositories: <http://www.re3data.org/>
- Finding Data (from the University of Edinburgh)
A searchable listing of online resources where researchers may find data to use in their research. N.B. a small number of the listed resources will be accessible only by subscription:

<http://www.ed.ac.uk/information-services/research-support/research-data-service/working-with-data/finding-and-using-research-data/finding-data>

- UK Data Service: Use Data
Guidance and tutorials on using data from the UK Data Service:
<https://www.ukdataservice.ac.uk/use-data>

Unit 3: Working with research data (13 sections)

1. File naming and version control - 1

Consider carefully the importance of being able to navigate your own data in the future. You, your colleagues, external collaborators, and other researchers who have read your publications should be able to find their way around the data, understand how it was created, and figure out the meaning of any abbreviations and variable names used.

2. File naming and version control - 2

You can help yourself achieve this by agreeing with the other members of your research group on a common approach to naming files, folders, and version numbers, and then documenting that approach for future users of your data.

3. Choosing the best formats for accessibility and sustainability – 1

Q: Will prospective users be able to open your files in ten years' time?

A: If you have generated your data using commercial proprietary software, there is a significant risk the answer will be 'no'!

4. Choosing the best formats for accessibility and sustainability – 2

You can maximise the chances that researchers will be able to use your data in years to come by using standards agreed by the researchers in your field, and by exporting or converting your data into file formats which can be read using open-source software.

[MANTRA \(https://mantra.edina.ac.uk/\)](https://mantra.edina.ac.uk/), the [UK Data Archive \(https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats\)](https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats), and the [Library of Congress \(http://www.loc.gov/preservation/digital/formats/\)](http://www.loc.gov/preservation/digital/formats/) provide good advice on selecting sustainable and accessible formats for your data.

5. Evaluating and mitigating risk – 1

Q: Have you considered the risks involved in your research?

A: Even well-resourced organisations like banks and IT companies can suffer data security breaches, as illustrated by this blogpost: "[6 Notorious Cases of Data Loss...](https://www.r1soft.com/blog/6-notorious-cases-of-data-loss-all-hosting-providers-can-learn-from)" (<https://www.r1soft.com/blog/6-notorious-cases-of-data-loss-all-hosting-providers-can-learn-from>).

6. Evaluating and mitigating risk – 2

Q: Are you working with human subjects? If so, have you considered the security of the data you gather about them?

A: And university buildings have been known to burn down, such as the University of Southampton's Mountbatten Building. See [Fire destroys top research centre \(http://news.bbc.co.uk/1/hi/england/hampshire/4390048.stm\)](http://news.bbc.co.uk/1/hi/england/hampshire/4390048.stm)

7. Data loss

Q: Your research data are a precious output of your research. It is vital you avoid losing them!

Q: What would you do if your laptop were stolen?

Q: What would you do if the hard drive on your office computer failed?

Q: Are your data backed up safely somewhere you can recover them?

8. Prepare for the worst, hope for the best

If you were planning to climb a mountain, it would be prudent to check the weather forecast and equip yourself adequately.

You can mitigate the legal, ethical, and data-loss risks associated with your project by planning your approach and equipping yourself with the requisite knowledge and skills.

"I am prepared for the worst, but hope for the best." Benjamin Disraeli (1883), *The Wondrous Tale of Alroy*, p. 61

9. Share your data to get more citations

The Digital Curation Centre, in conjunction with SPARC Europe, has gathered together the latest evidence demonstrating in a wide variety of academic fields that researchers who share their data get more citations of their publications:

[SPARC Europe - The Open Data Citation Advantage \(https://sparceurope.org/wp-content/uploads/dlm_uploads/2017/03/SPARCEurope_Briefing_OpenDataCitation.pdf\)](https://sparceurope.org/wp-content/uploads/dlm_uploads/2017/03/SPARCEurope_Briefing_OpenDataCitation.pdf)

10. Other benefits of sharing your data

Preservation for your own future use: by preparing and documenting your data for sharing with others, you will benefit by being better able to identify, retrieve, and understand it yourself after you have lost familiarity with it.

Discoverability: the metadata you add to your data in an online repository will be indexed by search engines to maximise discoverability.

Impact: Others who reuse your data and cite it in their own research help to raise interest in your research and increase your impact within your field and beyond.

Take-home message: Sharing your work online will allow more people to discover, read, and engage with it, and perhaps even enjoy it.

11. Licensing

You can enhance the reusability of your data by making it clear for others to understand what conditions you attach to its reuse.

This is done by attaching a licence to your data.

A [Creative Commons Attribution licence \(CC-BY\)](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>) places no restriction on reuse of the data except that you must be cited, and must be given credit as the creator of the data.

12. Metadata

This is an example [data deposit](https://figshare.com/articles/Simulations_on_calcite-protein_interactions/3491387/1) from figshare: https://figshare.com/articles/Simulations_on_calcite-protein_interactions/3491387/1

Q: What is a DOI?

A: Digital Object Identifier

Q: What is metadata?

A: Data about data

These are two common elements of a repository record.

13. Online resources

- [UKDA File Formats: http://www.data-archive.ac.uk/create-manage/format/formats-table](http://www.data-archive.ac.uk/create-manage/format/formats-table)
- [LoC – Library of Congress: http://www.loc.gov/preservation/digital/formats/](http://www.loc.gov/preservation/digital/formats/)
- [MANTRA: https://mantra.edina.ac.uk/](https://mantra.edina.ac.uk/)

Unit 4: Data Management Planning (10 sections)

1. Benefits of creating a Data Management Plan (DMP)

The process of creating a Data Management Plan will make you ask yourself the questions that will help you handle your data appropriately.

The DMP will help you plan how you will gather your data and look after your data from the beginning to the end of the research project.

It will help you plan for the resources you'll need in order to store your data securely, preserve them, and share them.

Furthermore, it will encourage you to think through carefully what kind of consent you need from any human subjects.

Finally, it will help you to avoid future problems sharing your data.

2. Research Data Lifecycle and DMP

It's a good starting point for your Data Management Plan if you write at least one paragraph for each of the six stages of the Research Data Lifecycle:

- Create: What data will be collected or created?
- Document: How will the data be documented and described?
- Use: Who will be responsible for data security and backup?
- Store: Where will the data be stored during the project lifecycle?
- Share: How will the data be shared and with whom?
- Preserve: Which data will be preserved for the long term?

3. DMP Top Tips

A Data Management Plan should be a "living document"--something you amend throughout the life of your project as you get a clearer picture of what data you'll be managing, and what you'll need to do with them.

- Start early - don't put it off till the last minute.
- Keep your DMP as jargon-free as possible.
- Keep your DMP short!
- Contact your funder if you need further information or clarification.

4. DMP Help

Using [DMPonline](https://dmponline.dcc.ac.uk/) (<https://dmponline.dcc.ac.uk/>) will make it easy to create and maintain your DMP.

The University of Sheffield also has [guidance for DMP creation](https://www.sheffield.ac.uk/library/rdm/dmp) (<https://www.sheffield.ac.uk/library/rdm/dmp>)

5. Planning for sharing your data

These are questions your DMP should answer:

Q: Are there standard formats for data used in your field and beyond?

Q: Should you plan to gather or convert your data to these standards?

Q: Would that make it easier for you and others to analyse and reuse the data in future?

Q: Who will own the intellectual property in your data?

Q: Will intellectual property issues restrict the ways your data may be shared with others?

6. Planning for archiving your data

When you don't need access to your data regularly you will be able to store them in a long-term repository.

In your grant application you will need to include costs associated with long-term archiving of your data.

7. Planning for handling sensitive data

If your project involves gathering data from human subjects, you need to address issues of data protection:

- You should develop a clear understanding of the [General Data Protection Regulation](https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation) (<https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation>).
- You will need to give careful consideration ahead of time to how you will keep that data secure.
- You may need to establish how you will carry out de-identification to allow the data to be shared more widely.

8. Planning for getting consent

If you are planning to work with human subjects, you should plan carefully how you will gain consent appropriately from them for all the possible reuses of the data.

MANTRA (<https://mantra.edina.ac.uk/>) is a free online resource which provides helpful examples and advice from University of Edinburgh professors on obtaining consent from subjects.

9. Research Data Lifecycle: a puzzle

- Q: Bearing in mind the goal of maximising long-term access, which software will you use to create your data?
A: Create
- Q: What will you have to document, in order for the data to be read and interpreted in the future?
A: Document
- Q: Who will be responsible for data security and backup?
A: Use
- Q: Where will the data be stored during the project lifecycle?
A: Store
- Q: How will you make your data available for others to reuse, and ensure you meet your funder/institution's data policy?
A: Share
- Q: How long does your funder require you to preserve the data, and what costs will need to be met for long-term storage?
A: Preserve

10. Online resources

- [DMPonline](http://dmponline.dcc.ac.uk) allows you to create a Data Management Plan with the option of using a funder-specific template. It also allows you to export your plan in a format appropriate for submission to your funder or for sharing with colleagues (<http://dmponline.dcc.ac.uk>)
- [DCC - Policy resources](http://www.dcc.ac.uk/resources/policy-and-legal): Information on UK Funders' data policies from the Digital Curation Centre (<http://www.dcc.ac.uk/resources/policy-and-legal>)

Well done.

You have completed all the units of this course "Introduction to research data management".

If you have any question, please contact us at:
rdm@sheffield.ac.uk

Best of luck with your research.

For more information visit our webpages:
<https://www.sheffield.ac.uk/library/research>