



Researcher Development Team

Introduction to Research Data Management 2025-26

Welcome to Introduction to Research Data Management Training, a course designed to save you time and stress down the line by helping you to understand and implement best practice in managing your data.

INTRODUCTION

≡ Welcome to the course

≡ What is research data?

BEFORE YOU BEGIN

≡ 1. Before You Begin

≡ 1.1 Data Management Plans

≡ 1.2 Ethics, data protection and research data

RESEARCH IN PROGRESS

≡ 2.1 Research in progress

≡ 2.2 Storing data while you work on it

≡ 2.3 File names, folder structure and formats

APPROACHING COMPLETION

≡ 3.1 Approaching completion: Depositing your data



3.2 Choosing a repository and sharing data

CONCLUSION



4. Summary and quiz

Welcome to the course

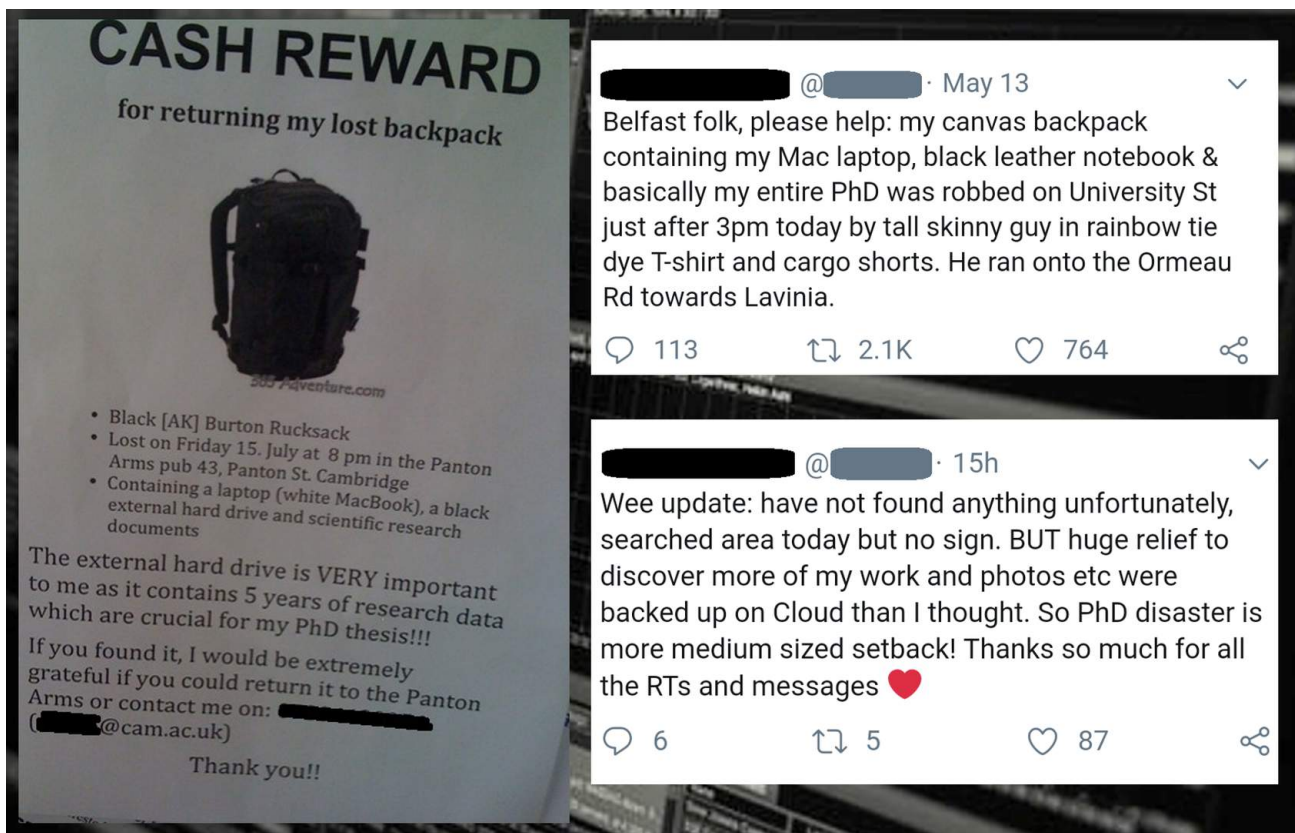
Welcome to **Introduction to Research Data Management**. This course aims to provide you with a basic introduction to best practice in research data management (RDM). We will also highlight the support that is available for research data activity within the University of Glasgow, and signpost external resources that you might find useful. This course is provided by the [Research Information Management team](#).

What is research data management?

Research data management is the **effective handling of information produced in the course of research**. It is the practice of collecting, organising and storing information to maximise its usefulness to you and to others who might reuse it.

The University of Glasgow has a range of resources to support researchers seeking to manage their research data, including the Research Data Management Service, these [Research Data Management webpages](#) and [Enlighten: Research Data](#), the University's research data repository.

A lot of this advice is common sense, so while you may not have thought of your existing practices in terms of research data management, you will probably find that you're already doing a lot of the things that this course is going to recommend.



Examples of catastrophic (and not-so-catastrophic) failures of data management. CC-BY

Avoiding disasters

The image above shows some of the worst examples of research data disasters – scenarios where researchers had only one copy of their data, which they have kept insecurely; the data is then either lost or stolen. We're interested in helping you avoid catastrophes like these –but we're also interested in preventing a problem that we think affects a much higher number of researchers.

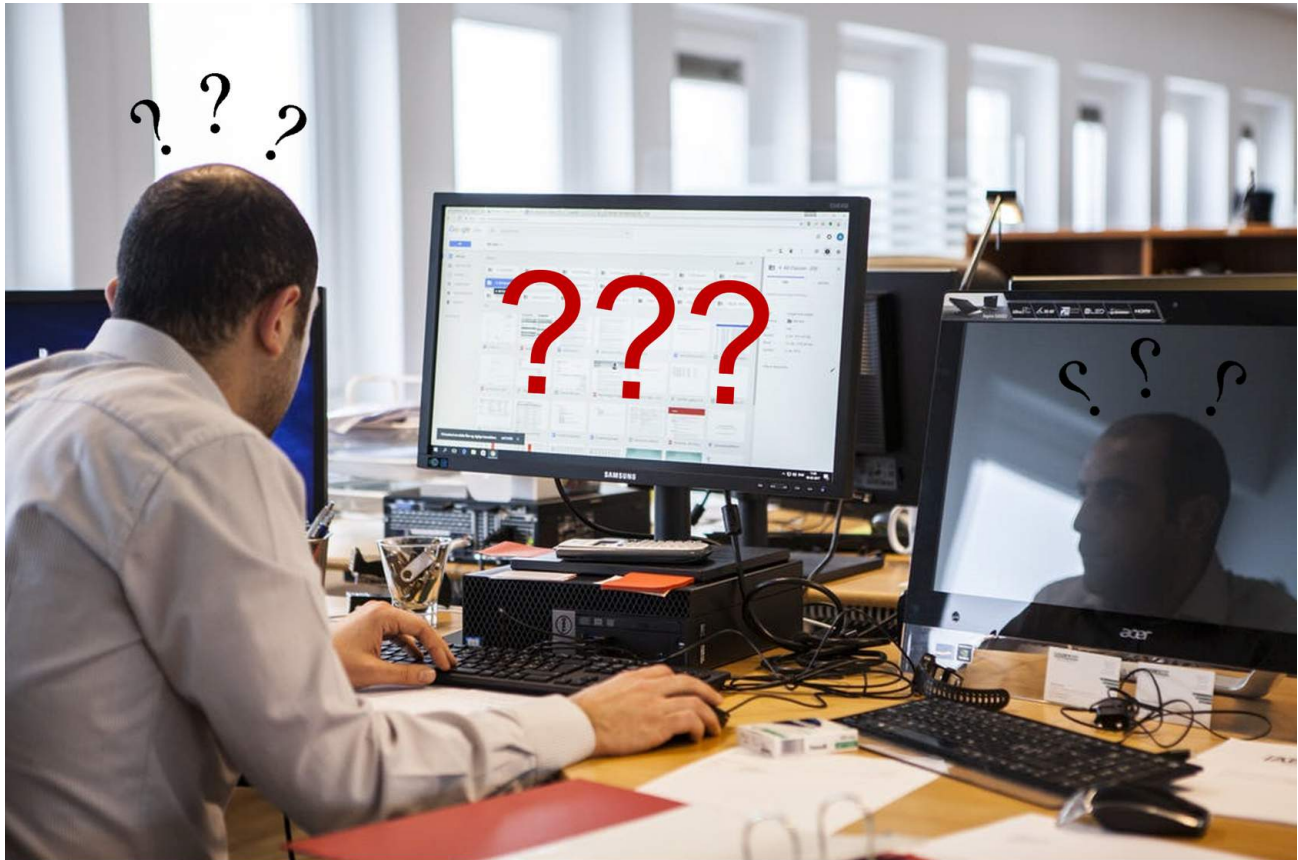
This is the long-term problem of **badly managed data**. If you don't organise and document your data appropriately, you can end up in a situation where although you have a copy of everything, you're not sure exactly which is the final version, you're not sure exactly what you did to produce it, and so on. This can be disastrous if you need to produce a definitive copy of your data for a reviewer or an examiner, or (in rare cases) if you need to defend the [integrity](#) of your research. It also goes against the University's policies, shown below.

At the University of Glasgow:

- We encourage transparency, openness, verification, and reproducibility by facilitating early and open sharing of research data, software, code, methods, preprints, open educational resources and materials with a wide range of

audiences.

- We place value on a wide range of different research output types across the whole lifecycle of research. This means that even if you can't make the data themselves openly available, you might identify other outputs like methods, protocols and so on which you can share.



Badly organised data. Composite, CC-BY

Benefits of good research data management

Good research data helps the research community:

- It improves the integrity and reproducibility of your research
- It facilitates sharing and reuse of data
- It strengthens the research environment, partly by enabling a better use of research funding.

There are also significant benefits to you as individual researchers:

- It reduces the risk of data loss
- It helps you to ensure compliance with funder and University requirements
- It can encourage new collaborations between creators and users of data
- But most importantly (for now) it **makes it easier to write up your work!**

Research Culture

The University's five Research Culture priorities are a statements of intention for the way all research should be done at the University of Glasgow:

- All colleagues are valued for their varied contributions to a diverse range of research activities and outputs (Research Recognition).
- People actively work together to support each other to succeed (Collegiality).
- We engage with and produce research that meets the highest standards of integrity (Research Integrity).
- We are committed to openness, transparency, rigour, and reproducibility (Open Research).
- Through the concepts above, and through a comprehensive framework of bespoke initiatives, we support all members of our research ecology to advance in their chosen career path (Career Development).

You can find out more about research culture at the University of Glasgow and how it applies to and benefits you here: [University of Glasgow - MyGlasgow - Research and Innovation Services - Research Culture](#)

Does this apply to me?

Historically, the vocabularies and concepts used to discuss research data management developed in relation to STEM research. However, the principles of good research data management will apply to all researchers. If your research generates documents, databases, data, formulas, interviews, recordings, or physical data then this will help! While you might not work with research data now, later in your career you may also find yourself using forms of data you didn't anticipate, so it is important that you understand the principles of good data management regardless of your current research practice.

Good research data management can support us in improving the openness of our research, and is a major component in our research culture and integrity. Properly managing your research data makes you a better researcher, and open research practices are rewarded at the University of Glasgow and beyond.

**At any point in the research process, if you have questions about your research data practice you can contact us at:
research-datamanagement@glasgow.ac.uk**

CONTINUE

What is research data?

RD Researcher Development Team



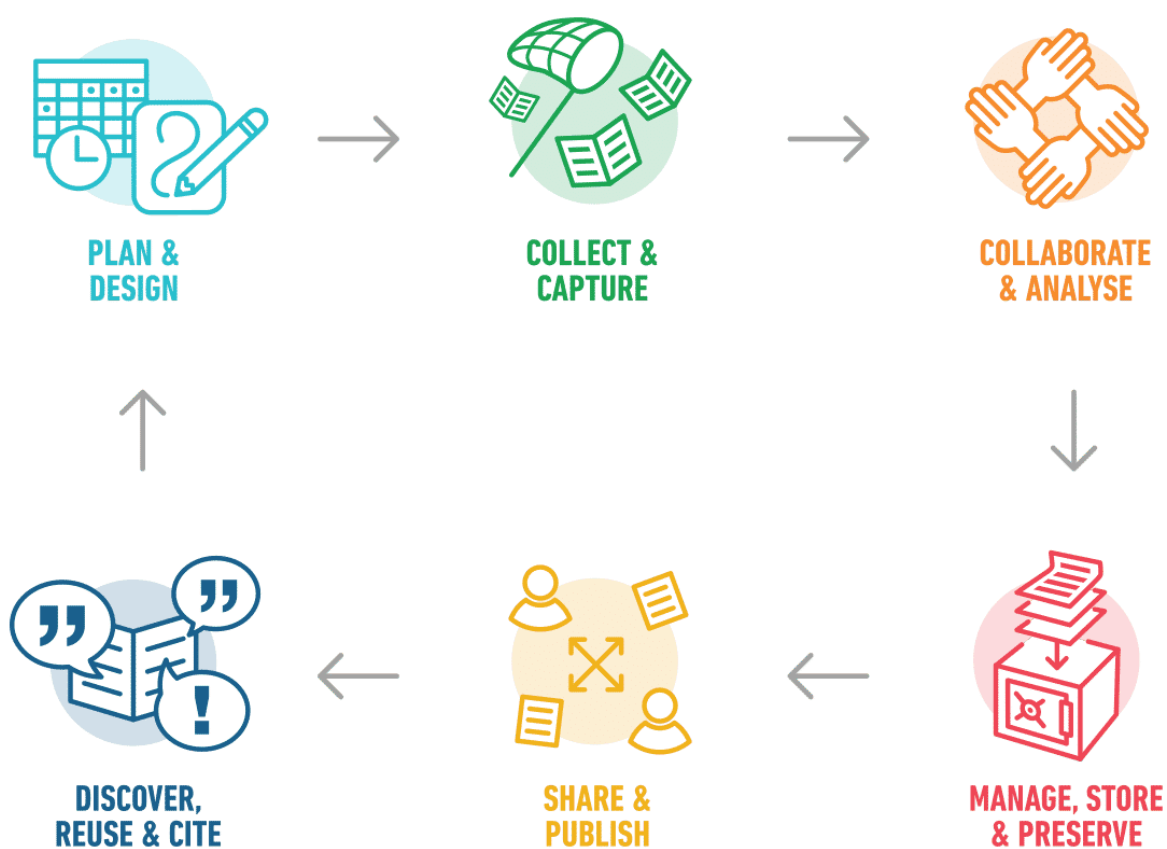
Types of research data.

Research data is **any information, whether digital or physical, required to underpin research**. It can be difficult to define in general what will constitute the data for any given project, and it will vary by discipline or approach. As an expert in your field, you have a lot of leeway to determine what data you think are essential for others to understand and build upon

your research. A good starting point is to think: *what information would someone need to see to understand how I reached my conclusions?*

The research data lifecycle

We think of research data management as taking place in a lifecycle – **planning and design**, then **collecting or creating data** and **organising and storing it**. Finally, **depositing, sharing, publishing and reusing data**.



(Image by Gaelen Pinnock reused from [Wikimedia Commons](#)^(opens in a new tab) under CC-BY-SA-4.0)

We have divided this course into three sections:

1

Before you begin your research. This section will cover the University's and funders' requirements, data management planning, and planning for ethics and data protection and other intellectual property concerns.

2

While your research is in progress. This section covers file formats for research data, documenting your data, folder structure, and using research notebooks. Here we will also discuss how to securely store research data while you work on it.

3

Approaching the completion of your research. This section will cover the process for selecting which data to deposit, depositing your data in an appropriate repository, choosing a licence and access conditions for your data, and linking your datasets to publications and other outputs.



In order to complete this course, you will need to work through all the sections and complete a short quiz.

CONTINUE

1. Before You Begin

RD Researcher Development Team



Researchers looking at an object.

This section will cover:

- **Understanding the University of Glasgow's requirements for research data**
- **Understanding your funders' requirements**
- **Writing a Data Management Plan**

It is important to plan for good data management and to meet the requirements of any relevant policies before the start of your project, as it will save you a lot of time and stress later on!

PGR CODE OF PRACTICE

OCTOBER 2023



The PGR Code of Practice

The University of Glasgow's policies

The University of Glasgow's policies relating to research data appear in:

The [Code of Good Practice in Research](#);

and the [Postgraduate Research Code of Practice](#).

As well as explaining what is required of you, these documents also contain lots of useful advice to help you conduct your research.

You can find these documents on the University's [Policies page](#).

The Code of Good Practice in Research

Section 7 of the [Code of Good Practice in Research](#) explains the University's expectations regarding your data. You should:

1. **Keep clear and accurate records of research**
2. **Prepare a data management plan for projects which will generate data**
3. **Keep sufficient metadata to put the data into context**

This means that you need to produce documentation enabling other researchers in your field to discover the data, understand what you did and if permissible, re-use it.



Metadata is additional information that accompanies the digital resource to ensure that it can be accessed and understood over time, such as keywords used as tags to enable accurate searching and retrieval.

Store data of long-term value for at least 10 years post-project

Data of long-term value is that which underpins a PhD thesis or a publication, or which you will use as the basis of a future funding application. Regardless of whether it can be shared, data of long-term value should be deposited somewhere suitable.

Deposit a copy of data generated using University resources when a member of staff/student leaves

Deposit data underpinning publications in a trusted repository

You don't have to put your data in the University's repository if it's deposited somewhere else that is more appropriate. If you deposit it somewhere else, let us know so we can make a record and link it to your thesis.

When possible, make deposited data openly available

This means that you should plan to make your data openly available unless you have a compelling reason to restrict access, for example, that it is copyrighted, confidential, or contains personal information that you can't disclose. You should only restrict data that needs to be restricted – any suitable data should be made available with an Open Access licence.



Our ['Open Research for restricted, sensitive or confidential data'](#) course will help you to understand the requirements and processes around working with data that should be restricted while you work on it, and will help you to understand which parts of these types of data can be deposited or shared.



UK Research
and Innovation



Science and
Technology
Facilities Council



Arts and
Humanities
Research Council



Engineering and
Physical Sciences
Research Council



Biotechnology and
Biological Sciences
Research Council



Economic
and Social
Research Council



Research
England



Natural
Environment
Research Council



Innovate
UK



Medical
Research
Council

UK Research and Innovation Logos.

Funders' data requirements

Some funders, notably UK Research and Innovation (UKRI) funders and charities like the Wellcome Trust and Cancer Research UK, have expectations regarding the research data produced as a result of their funding. These

expectations tend to fall in line with the UKRI Common Principles on Data policy, which says:

‘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.’ [UKRI ‘Making your research data open’](#)

Including Repository deposit costs in Funding Applications

Some funders provide their own data centres (eg [ESRC’s UK Data Service](#) and [NERC](#)) which are free for award holders to use. Usually, funders who do not provide a data centre but still have an expectation that you will deposit research data will permit data storage costs to be included in grant applications. Costs relating to RDM are subject to the normal rules of grant funding: for example, costs must be fully justified, be spent before the end date of the project and may only be charged for data produced by the project.

Activity

Familiarise yourself with the research data policies of a couple of funders. A good place to start is the Digital Curation Centre’s [list of policies](#). If you are not funded for your current research, look at the policies of funders who support research in your area.

While you are looking at funders’ policies, consider: **will you need to do anything in addition to the University’s data requirements to satisfy the funder?**



The University’s research data policies are designed to be compatible with most funders’ requirements. If you’re confused about what your funder expects, you can contact the [Research Information Management team](#) for advice

CONTINUE

1.1 Data Management Plans

RD Researcher Development Team



Every project which generates or processes data needs to have a data management plan (DMP). A DMP is a document which describes everything that will happen to your data during your project. The first draft of a DMP is usually quite straightforward, and as your practices develop the plan becomes more detailed, ideally until it can function as a handbook for recording your data management practices. DMPs usually cover the following areas:

- What data will you collect?
- How will it be documented?
- Ethical, legal and intellectual property considerations in the data
- Storage and organisation of the data
- Deposit and long-term preservation
- Plans for data sharing
- Responsibility for implementation



A messy archive box

Unlike ethical approval, Data Management Plans are not formally signed off by anyone in the University. **You should continue to update your DMP when your project is underway, to make sure that it still makes sense in the context of your research.** You should review your DMP after any milestone in your project that might affect how you

manage your data – for example, if you pilot data collection, when you first analyse your data and when you first write up your findings...



If you keep your data management plan up to date, you'll find it very easy to modify it for your next project, if you're working with similar data. It's also useful to have a current DMP if you need to check anything with your supervisor, or with an ethics committee, with the Research Information Management team, or with a funder.

Research Data Management costs

Some activities relating to research data management will incur costs. These might include deposit or storage costs for the data, costs relating to tools or processes for analysing or preparing data. You should plan for this, and allow enough time in your project plan for data management. More guidance on costs relating to RDM is available in our guide to [Anticipating the Costs of Research Data Management](#).

DMPs as part of PGR students' progression

All PhD students must prepare a data management plan for their research projects as part of their submission for progression from year 1 to year 2 of their doctoral studies.

Students should use the [University of Glasgow PGR template](#) regardless of their funder.

Key Links:

UofG Data Management Plan PGR template:

<http://eprints.gla.ac.uk/179057/>

UofG Data Management Plan guidance:

<http://eprints.gla.ac.uk/179058/>

UofG Data Management Plan training:
<https://www.gla.ac.uk/myglasgow/openresearch/training/>

If you will not be generating data, using secondary data, developing software or undertaking qualitative or quantitative analysis (eg. pure mathematics) then you can make a statement to this effect in the first part of section 2 of the template and leave the remainder blank.

PGR templates, and templates provided by funders, are also available in [DMPOnline](#), an online tool to help you write and review data management plans. We also offer a workshop on using the DMPOnline tool. You can find more details on our [training page](#).



Where does a DMP fit with other processes?

Every project collecting data should have a Data Management Plan, and if you're collecting personal data you need to undertake some additional steps (for other considerations which might affect your data please see our 'Open research for restricted, sensitive or confidential data' [training](#)). Below is a workflow for initiating projects involving **personal data**.

DMPs, ethics and data protection

Data Management Plans are a very useful first step in preparing an application for ethical approval, and in helping you to plan for compliance with data protection (GDPR) requirements.

Step 2

First prepare a Data Management Plan

Draft a DMP to plan what you want to do with the data (while satisfying University and funder requirements). We can review your DMP if you would like; just email us at research-datamanagement@glasgow.ac.uk.

Step 3

Complete the appropriate data protection paperwork

By completing the appropriate Data Protection paperwork you will establish what you are legally allowed to do with the data.

All projects collecting or processing personal data need to complete a **Data Protection Impact Assessment** form, and participants should be provided with a **Privacy Notice**. If you are sharing personal data outside the University, you need to arrange a **Data Sharing Agreement** with the parties receiving the data.

Step 4

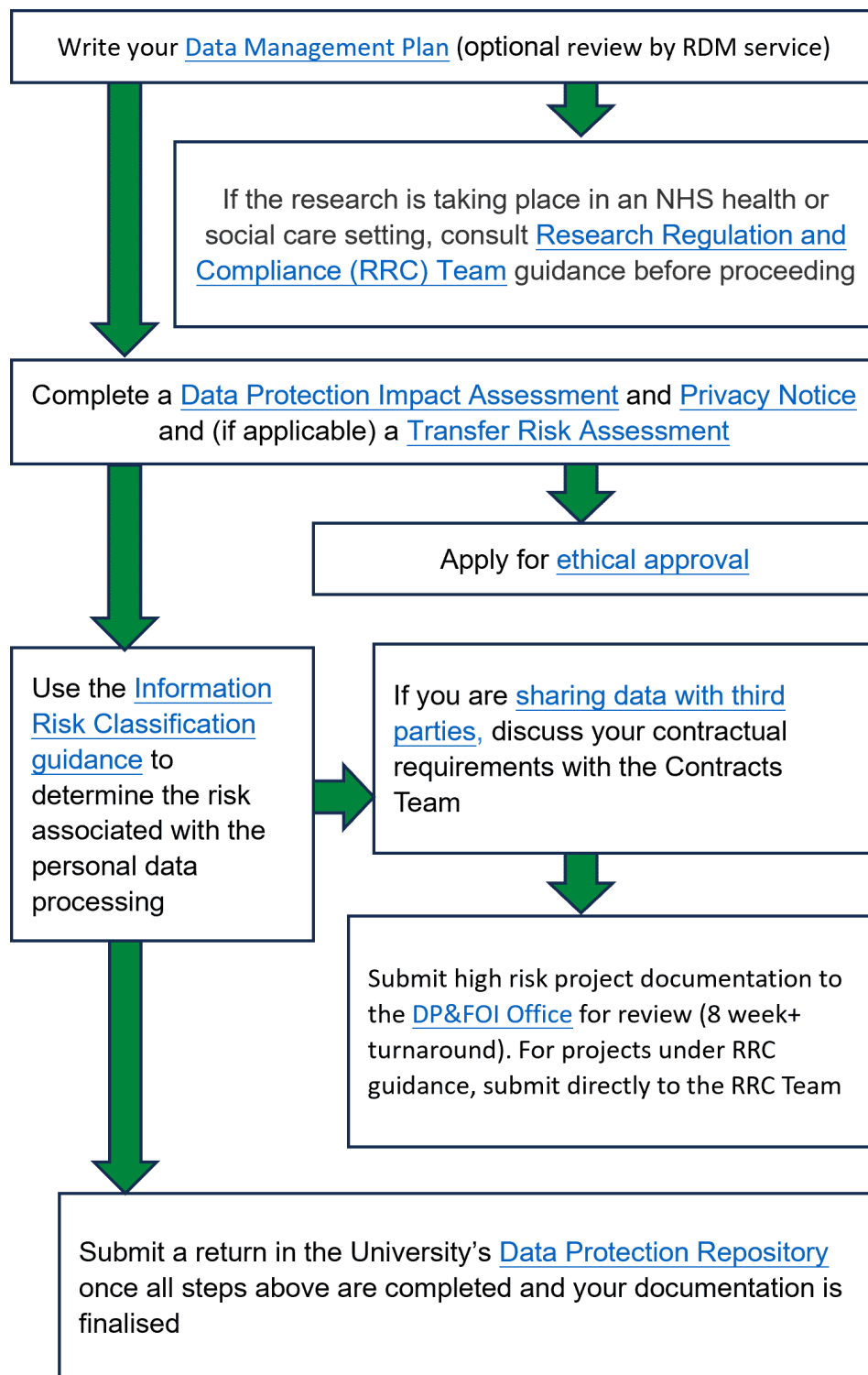
Seek ethical approval for your project

Having determined what you want to do with the data in the DMP, and what you legally can do with the data in the GDPR paperwork, you can now seek ethical approval for all the planned uses of the data, including depositing data of long-term value and sharing it where appropriate.

Summary

Bear in mind that the review processes for data protection and ethical approval can be lengthy. An expanded version of this workflow, including links to relevant documents and support, is available from our repository:

[10.36399/gla.pubs.202746](https://doi.org/10.36399/gla.pubs.202746).



A full version of this diagram with clickable links can be found at: <https://doi.org/10.36399/gla.pubs.202746>

Next we will explore ethics and data protection in a bit more detail.

CONTINUE

1.2 Ethics, data protection and research data

RD Researcher Development Team



Examples of personal data.



This section gives a very brief overview of data management considerations relating to ethics and data protection. Our [training course](#) titled 'Open Research for restricted, sensitive or confidential data' provides much more detail that is essential when working with research information and data concerning

humans, animals, clinical data, third-party data, has commercialisation potential or risk, or is subject to ethical approval and/or Trusted Research legislation.

Any research involving **human participants, material or data** requires approval from an appropriate **Ethics Committee** (College, University or NHS).

Consent from participants is required for ethical purposes, and GDPR requirements should be dealt with separately. You should seek clear and unambiguous ethical approval and consent for all planned uses of your data, including the long-term storage and sharing of the data.

For ethical advice, you should contact your College or School ethics officer in the first instance:

<https://www.gla.ac.uk/research/strategy/ourpolicies/ethics/>

All projects collecting or processing human data need to complete a Data Protection Impact Assessment (DPIA) form.

The Data Protection team will need to **review** your [DPIA](#) if the processing of the personal data is **high risk** (for instance if you have any [special category data](#) or a very large volume of personal data). You can use the [DPIA decision tool](#) and [Information Risk Classification](#) to establish whether your DPIA needs to be reviewed.

The online [Introduction to GDPR training course](#) is a requirement for all University of Glasgow staff and PGR students. For advice on GDPR, contact the Data Protection and FOI Office in the first instance:

Tel: 0141 330 6494, email: dp@gla.ac.uk, web: www.gla.ac.uk/services/dpfoioffice



You should not delete your data as soon as your project ends, unless you have a compelling reason to do so.

There is a special recognition in GDPR that research data should be kept for longer than other kinds of personal data.

The University requires research data of long-term value to be securely held for a period of ten years after the completion of a research project, or for longer if specified by the research funder or sponsor.

<https://www.gla.ac.uk/myglasgow/dpfoioffice/guidanceforstaffandstudents/research/retentionofresearchdata/>

Activity

When you first encounter advice on research data management it can seem quite abstract, especially if you haven't yet collected any data or started your data management planning. To help you, we have collected some case studies. These are examples of researchers who we think demonstrate good research data management practices.

Choose a case study to read. What data did they collect? How did they manage it, and what did they deposit and share? Can you apply any of this to your own research?

- [Research Data Management in Ethical Consumption](#)
- [Research Data Management in Natural History](#)
- [Open Research in Psychology](#)
- [Working in the Psychological Science Accelerator](#)



If you are interested in contributing to a case study, please get in touch at research-datamanagement@glasgow.ac.uk.

In the next section we will move on to look at documenting and organising your data.

CONTINUE

2.1 Research in progress



This section will cover:

- **Documentation and research notebooks**
- **Safely and securely storing your data while you work**
- **File naming and storage**

Earlier we learned that the University's Code of Good Practice in Research asks you to **produce sufficient documentation that would enable someone in your field to understand what you did to produce your data**. Producing good documentation is also a condition of funding in many cases. In this section, we will explore organising and documenting your data, and present a simple way to produce a minimally acceptable standard of documentation.



2.1 Documenting your data

When you start to plan your documentation, bear in mind that you are the person most likely to re-use your own data. We want to avoid the problem of data loss through poor management, so you should always consider:

What information might you need to make sense of the data in a few years?

For example, you might include:

- details of sample preparation
- experimental protocol followed
- machine settings
- criteria for inclusion/exclusion
- key themes
- guide to field names

- details of software needed
- list of abbreviations used

The possibilities are endless and are specific to your research. Try to think of a few examples that apply to your project and jot them down to remember to write them up later.



CC-BY.

Always consider:

Are your files clearly described?

Is your data clearly labelled with versions and dates?

Is your data logically structured and named?

Where available, are you using standard **metadata and naming conventions**?

You should organise and describe your files as you create them, and consider keeping a research notebook to keep all your notes in one place. Consider using an electronic notebook like OneNote or RSpace. For more information, please read our [guidance on research notebooks](#) including paper notebooks, electronic notebooks, OneNote, Endnote, and Rspace.

Wherever possible you should follow the best practice in your discipline or research community. **Describe how you're going to document your data in your Data Management Plan.**

CONTINUE

2.2 Storing data while you work on it

RD Researcher Development Team



Sean Ellis <https://flic.kr/p/55RpEe>

In this section we will look at where to store your data while you are working on it. We will cover secure storage and backup, restricting access to your data, and then look briefly at choosing which data to keep after your project.

Any security measures you put in place for your data should be proportionate to the risks in the data.

- Consider your ethical approval and data protection requirements: don't share any personal data without permission and adequate safeguarding!
- Who will need to access the data during your project? You should plan for your data to be accessible to at least yourself (and your supervisor if you're a student), but there may be others who need to access it.
- Think about what will happen to your data in transit – is it secure?



<https://zenodo.org/doi/10.5281/zenodo.11147886>

Where should the data reside?

We recommend that you always use the University's authenticated storage spaces to store your research data while you work on it.

Remember that these storage locations are not a final resting place for your data – you need to plan to move the data of long-term value to an appropriate repository. All student accounts (UG, PGT, PGR) are terminated three months after graduation, and staff lose access to their accounts at the end of their contract of employment.

1

Microsoft OneDrive/Teams. Every member of the University has 1TB of storage space in their OneDrive account. OneDrive is suitable for solo research, while Teams should be used for projects with multiple UofG collaborators. Both OneDrive and Teams use Microsoft Sharepoint storage, contracted by the University. Log in through <https://office365.gla.ac.uk>.





































2

The shared drive (usually J:\). The University's shared drive is also suitable for research data. Some departments have their own local servers that are also suitable. Contact your local IT support for information on setting up a shared folder that is accessible only to those who need to have access to your data.





3

NextCloud. This is an [encrypted transfer and storage tool](#) that is installed on University servers. IT Services can provide an on-campus cloud service for people who cannot use Teams, or have local data protection requirements. It provides a simple cloud file-sharing system, accessible through your web browser or through a desktop sync app. Students will need their supervisors to arrange access via the IT Helpdesk.

More information on the University's storage is available in this [IT Services guide](#). The table below shows some considerations when choosing an appropriate storage area for your data. If none of these options look suitable to you, or if you expect to have a dataset over 1TB, contact your local IT support or the [Research Computing as a Service](#) team.

				
Data needs to be restricted to a small group				
Share with colleagues at the University of Glasgow				
Share with collaborators at another university				
Share with collaborators without university logins				
Data must stay in the EU				
Data must stay in the UK				
Data must stay on campus				
Storage managed or contracted by the University				








				
Data needs to be restricted to a small group	✓	✓	✓	✗
Share with colleagues at the University of Glasgow	✓	✓	✓	✗
Share with collaborators at another university	✗	✓	✓	✗
Share with collaborators without university logins	✗	✓	✓	✗
Data must stay in the EU	✓	✓	✓	✗
Data must stay in the UK	✓	✓	✗	✗
Data must stay on campus	✓	✓	✗	✗
Storage managed or contracted by the University	✓	✓	✓	✗








Free cloud storage tools

We don't recommend that you use free cloud storage tools for your research data. It is not appropriate to put personal data on any storage outside the EU unless you have taken appropriate steps to comply with data protection requirements.

				
Data needs to be restricted to a small group	✓	✓	✓	✗
Share with colleagues at the University of Glasgow	✓	✓	✓	✗
Share with collaborators at another university	✗	✓	✓	✗
Share with collaborators without university logins	✗	✓	✓	✗
Data must stay in the EU	✓	✓	✓	✗
Data must stay in the UK	✓	✓	✗	✗
Data must stay on campus	✓	✓	✗	✗
Storage managed or contracted by the University 	✓	✓	✓	✗

Storage managed or contracted by the University

The University's storage is provided to support you as researchers, whereas free cloud storage is there to tempt you to pay for a premium account to access extra features.

				
Data needs to be restricted to a small group	✓	✓	✓	✗
Share with colleagues at the University of Glasgow	✓	✓	✓	✗
Share with collaborators at another university	✗	✓	✓	✗
Share with collaborators without university logins	✗	✓	✓	✗
Data must stay  the EU	✓	✓	✓	✗
Data must stay in the UK	✓	✓	✗	✗
Data must stay on campus	✓	✓	✗	✗
Storage managed or contracted by the University	✓	✓	✓	✗

Personal data

Any personal data will have to remain within the EU unless you have explicit permission from your participants and adequate safeguarding measures are in place. All the University's storage areas are within the EU, and the shared drive and NextCloud are on campus.

Free personal cloud storage must not be used to store personal data. Paid personal cloud storage is discouraged as the location of the servers and the ownership conditions vary between providers and are not always obvious.



General Data Protection Regulation Article 45 says: **Personal data should not be shared outside the EU without explicit consent and adequate safeguarding.**

This means that you should not store any personal data somewhere if you're not sure what territory the data will be kept in. If you're not sure whether a storage area or tool is compliant with data protection legislation, you can check with the Data Protection and FOI Office at dp@gla.ac.uk.

<https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/>

Researchers travelling outwith the European Union should take care not to download any personal data to the local drives of their computer, as this would constitute taking data outside of the EU.

Handling confidential data

The University Information Security team has provided extensive guidance on storing confidential data.

If your data are explicitly subject to confidentiality restrictions, either because they are considered confidential information or because they are classified as medium or high risk, they are subject to the [University of Glasgow Policy on Confidential Data](#).

- Use central filestores on secure servers maintained in secure physical environments
- Confidential data should not be held on local disk storage (eg C:\)
- Confidential data should not be stored or accessed on mobile devices unless adequate security measures are in place.

Confidential data must be encrypted when:

- Stored on a laptop, phone or tablet
- Stored on portable media (eg memory stick, external hard drive)
- Exchanged with external organisations and individuals



Our [training course](#) titled 'Open Research for restricted, sensitive or confidential data' provides much more detail that is essential when working with research information and data concerning humans, animals, clinical data, third-party data, has commercialisation potential or risk, or is subject to ethical approval and/or Trusted Research legislation.

Choosing which data to retain after the project



<https://zenodo.org/doi/10.5281/zenodo.11147886>

It is not usually necessary to keep all the versions of data that you produce during a project. You should consider your obligations for data retention – these might come from University or funder policies, from a collaboration agreement between partners, or from your ethical approval. You may also have statutory (legal) obligations – for example, certain medical data needs to be retained for a statutory period.

Take a look through our 'Open research with restricted, sensitive or confidential data' [training](#) and handbook for more information on choosing what data to keep or destroy.



If you have any legal or ethical reasons to destroy all or part of your data, you need to make sure you do so properly. Having your data well organised and documented will help you to identify the parts of your data that you cannot keep.

For more information on data appraisal and selecting what to keep, see our guide: <https://edshare.gla.ac.uk/1385/>

CONTINUE

2.3 File names, folder structure and formats

RD Researcher Development Team

File names

You should choose a format for naming your files and use it consistently. Good file names are relatively short but rich in information, and they consist of information elements separated by underscores ‘_’.

Examples of elements	Top tips
<ul style="list-style-type: none">•Project or experiment name or acronym•Location / spatial coordinates•Researcher name/initials•Date or date range of data acquisition•Type of data•Conditions•Version number of file•File extension	<ul style="list-style-type: none">•Avoid special characters•Do not use spaces•Use an underscore as an element delimiter•Use a hyphen or capital to delimit words within an element•Think about the order of elements•Use naming standards wherever possible (eg use YYYYMMDD for dates)•Abbreviate where possible

Commonly seen filenames with no convention:
PlanFinalFINAL.doc
Test_data_2013
Second_test
Meeting Notes Oct 23

Filenames with a naming convention:
20130503_DOEProject_DesignDocument_Smith_v2-01.docx
20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx
20130825_DOEProject_Ex1Test1_Documentation_Gonzalez_v3-03.xlsx
20131002_DOEProject_Ex1Test2_Data_Gonzalez_v1-01.xlsx
20141023_DOEProject_ProjectMeetingNotes_Kramer_v1-00.docx

File name corrections.

In the image above we have taken some commonly used filenames and moved them to a standard naming convention. There are significant advantages to naming your files in this way; it makes it easier to identify duplicate

files or files which contain sensitive information, and you can be more certain that you should keep, deposit or share any given file.



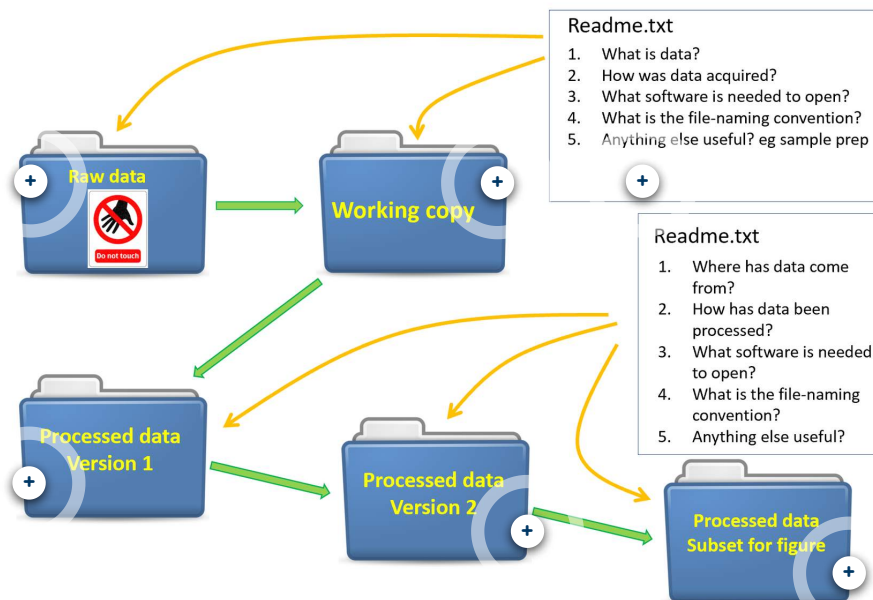
Including a 'status' field in your filenames is a good way of identifying the purpose of a given file within your project. A status field may indicate that a file is raw data, a working copy, or a version cleared for sharing.

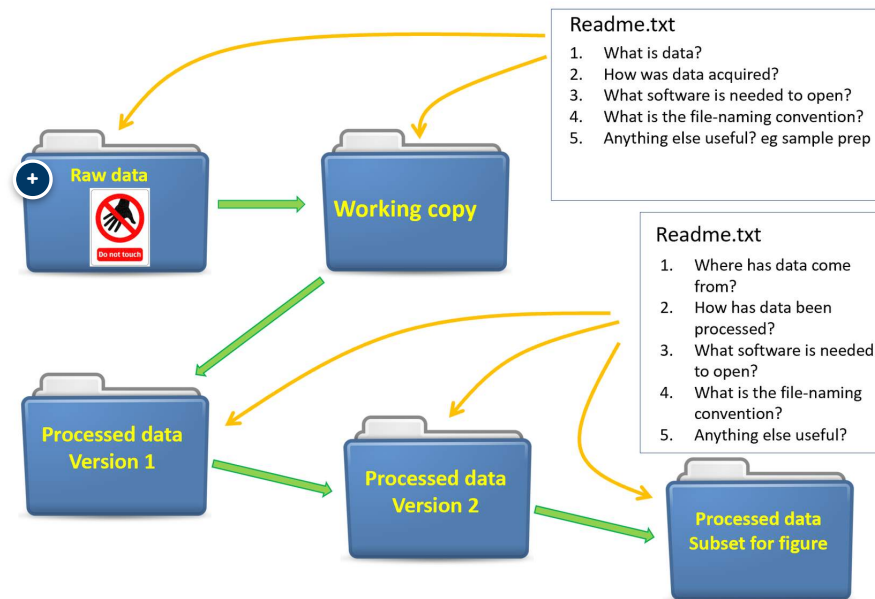
Further reading

For more information on file naming, see our [file naming guide](#).

Folder structure

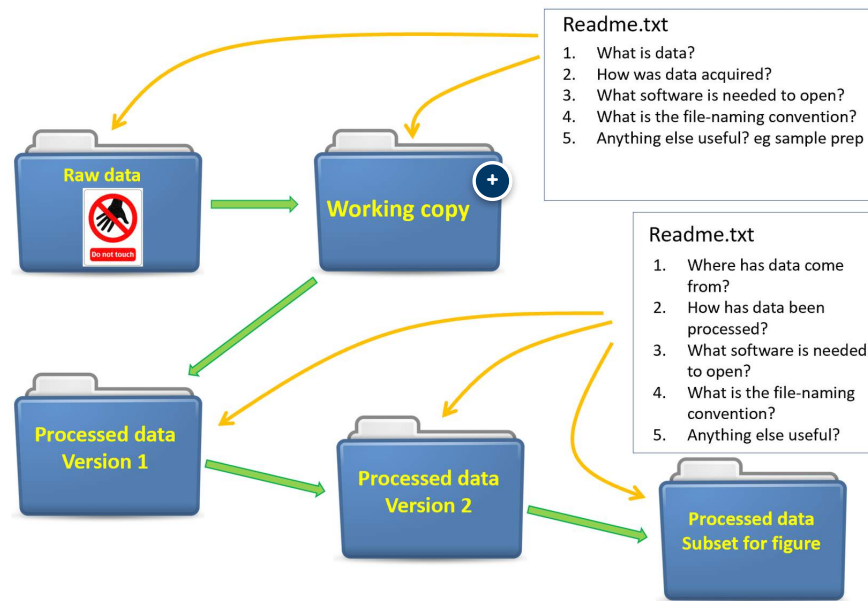
The diagram below illustrates our recommended folder structure for research data.





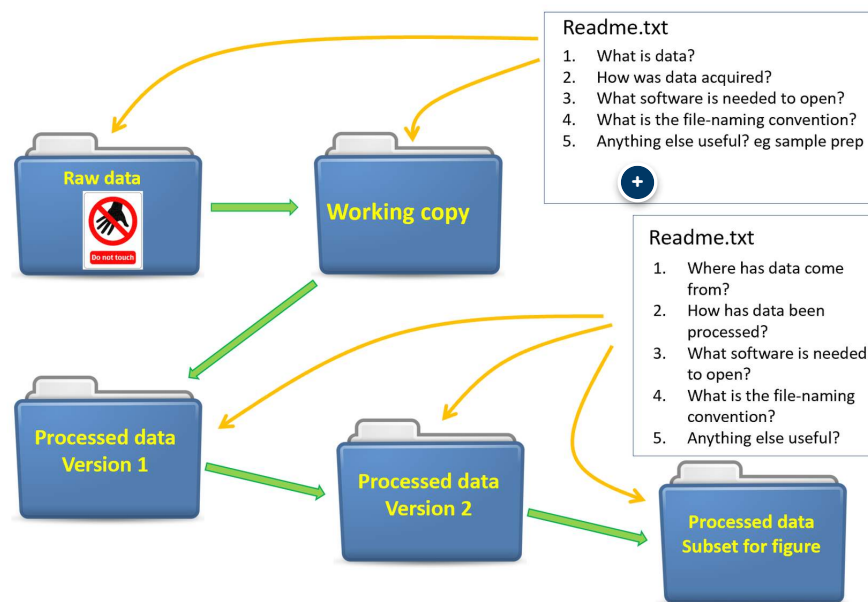
Raw data

Take a copy of your raw data and don't touch it for the duration of your project. You should always have a clean copy of the raw data you can return to in the event that you adulterate your working copy.



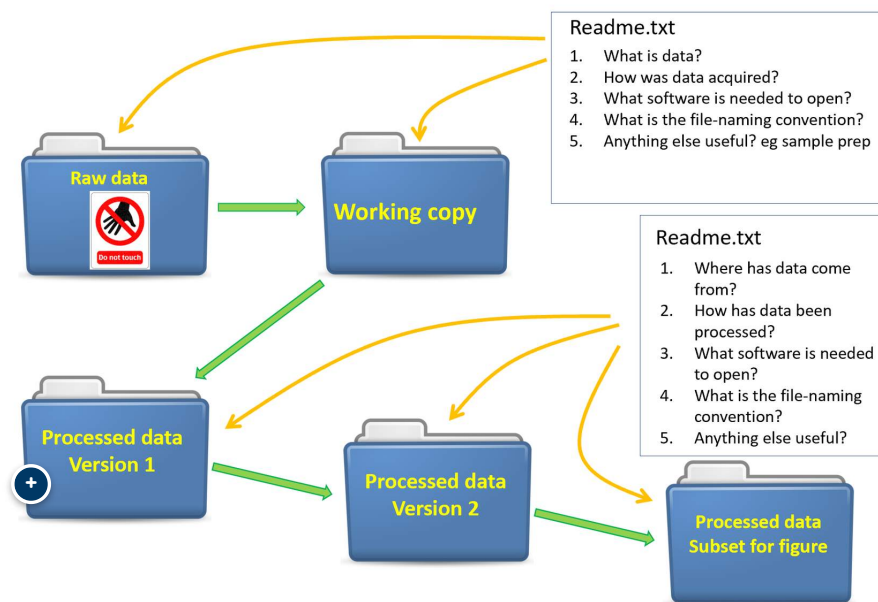
Working copy

The working copy should be the same as the raw data, but it is from this folder that you will derive all the processed versions of your data.



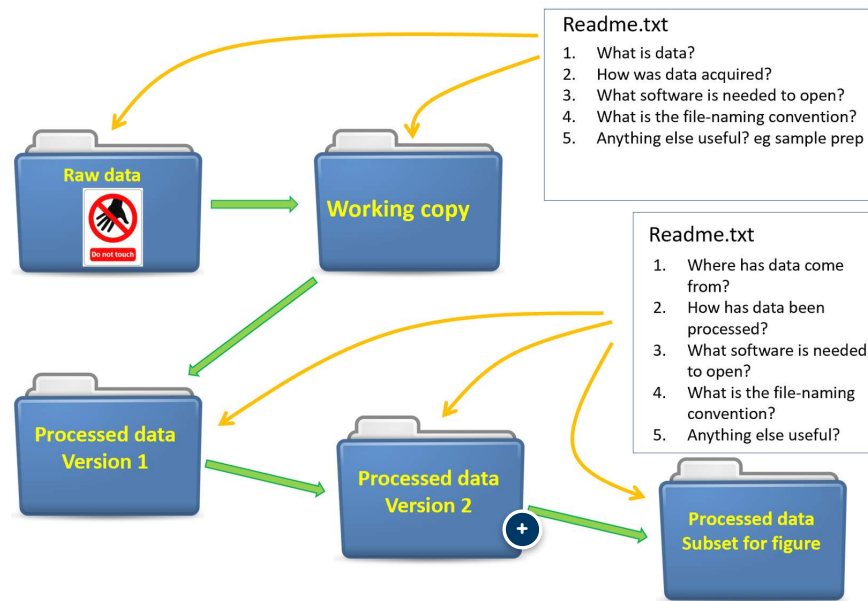
Readme.txt

In all folders containing data, you should include a Readme file. This is a simple text file which explains basic information about the data contained in the folder. For processed data, indicate where it has come from and how it was processed.



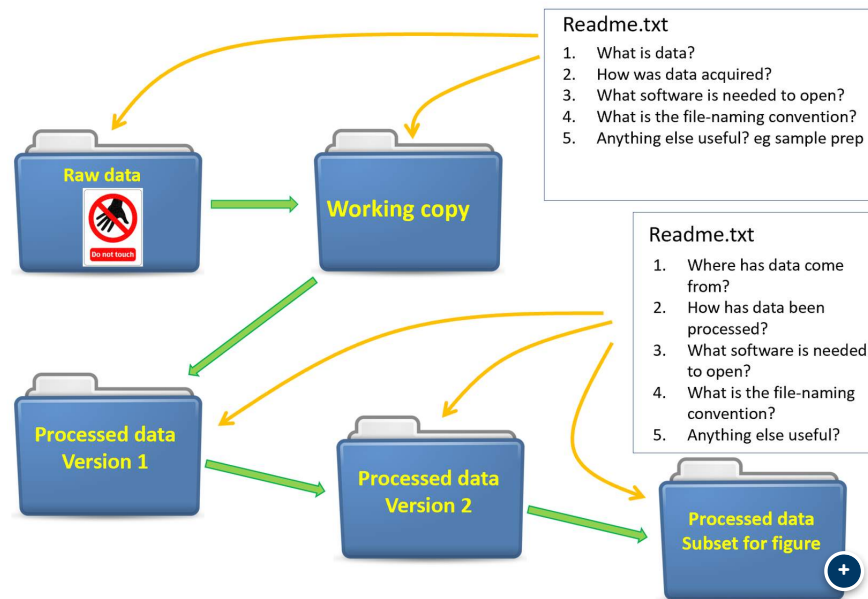
Processed data

When you process your data, wherever possible make a new folder that contains the outputs of the processing. This way you produce a chain of provenance from your raw data to the final processed data that will appear in your thesis or in an article.



Processing sensitive data

If you remove identifiers from your data, keep a record of which you have removed. If you have removed sufficient identifiers that your data is considered suitable for sharing, you can be confident that everything 'downstream' of this version can be shared too.



Data subsets for figures

When you are preparing figures, excerpts or illustrations of your data to include in a publication, wherever possible you should try and separate the subsets of the data that underpin them. You can usually release data subsets in order to satisfy funders' requirements to share data upon publication.

Of course, this is an idealised workflow and your actual research practice may differ from it – we know that research is rarely as linear as this! But if you take this approach as your starting point, you can diverge from it in planned fashion and still enjoy its benefits.

Activity

Drag and drop the cards below to sort types of data into the appropriate folders. Some are data in various stages of a project and some are additional documentation that you might also consider storing.

Raw data

**Audio recording of an
interview**

Reading from an instrument

**Spreadsheet downloaded
from government data
source**

Processed data

Transcript of an interview

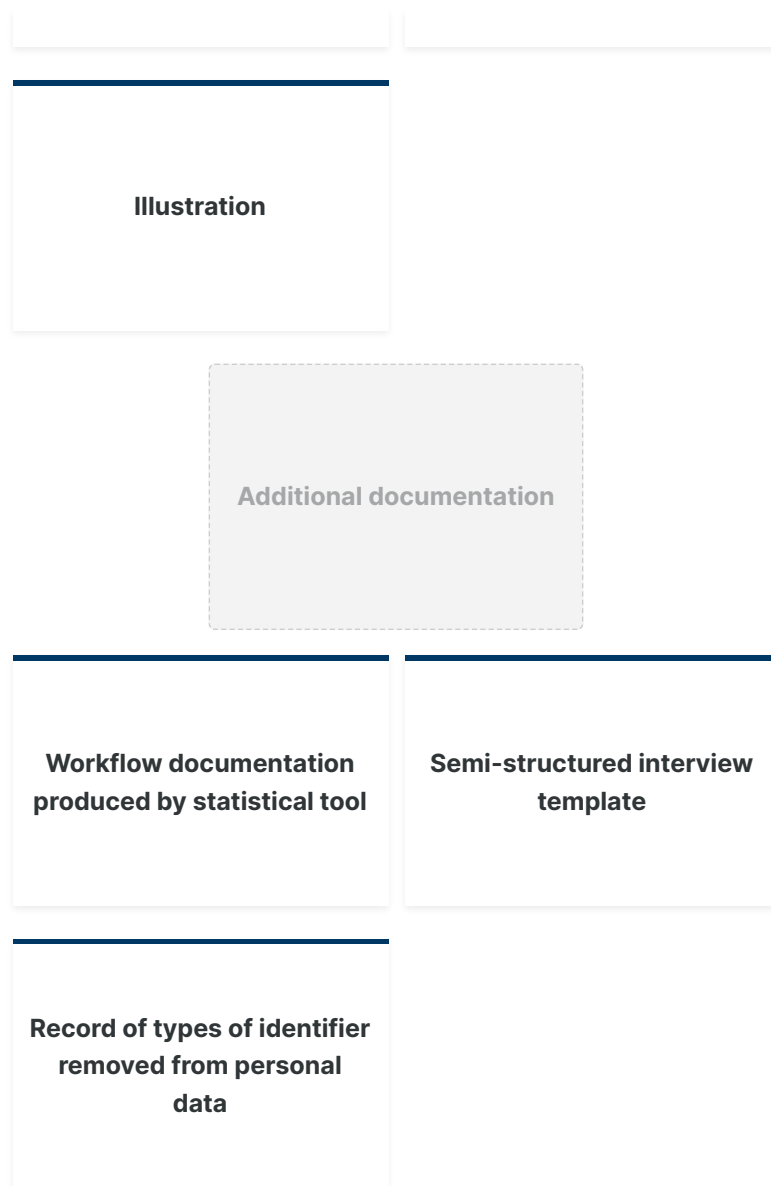
**Instrument readings
collated into a spreadsheet**

**Spreadsheet data that has
been modified with a
statistical tool**

**Data subset appearing in
publication**

**Anonymised extract from
interview**

**Subset of results used to
produce a visualisation**



Extra documentation

If you have a good file naming convention in use, and your data is organised into logical folders which contain Readme files, it should be obvious to you if there is any missing information that someone would need to make sense of your data. This might be a log from the tool you're using to process the data, or an extra document like a structured interview prompt. Or you might be given confidence that no extra documentation is needed.



Many researchers like to also keep a data handlist. This is a spreadsheet with a line for each data file, that brings together information about how it was collected, what it contains and so

on. Data handlists are very useful for managing your data, but can also be useful resources for someone who might reuse your data (provided they have been suitably anonymised).

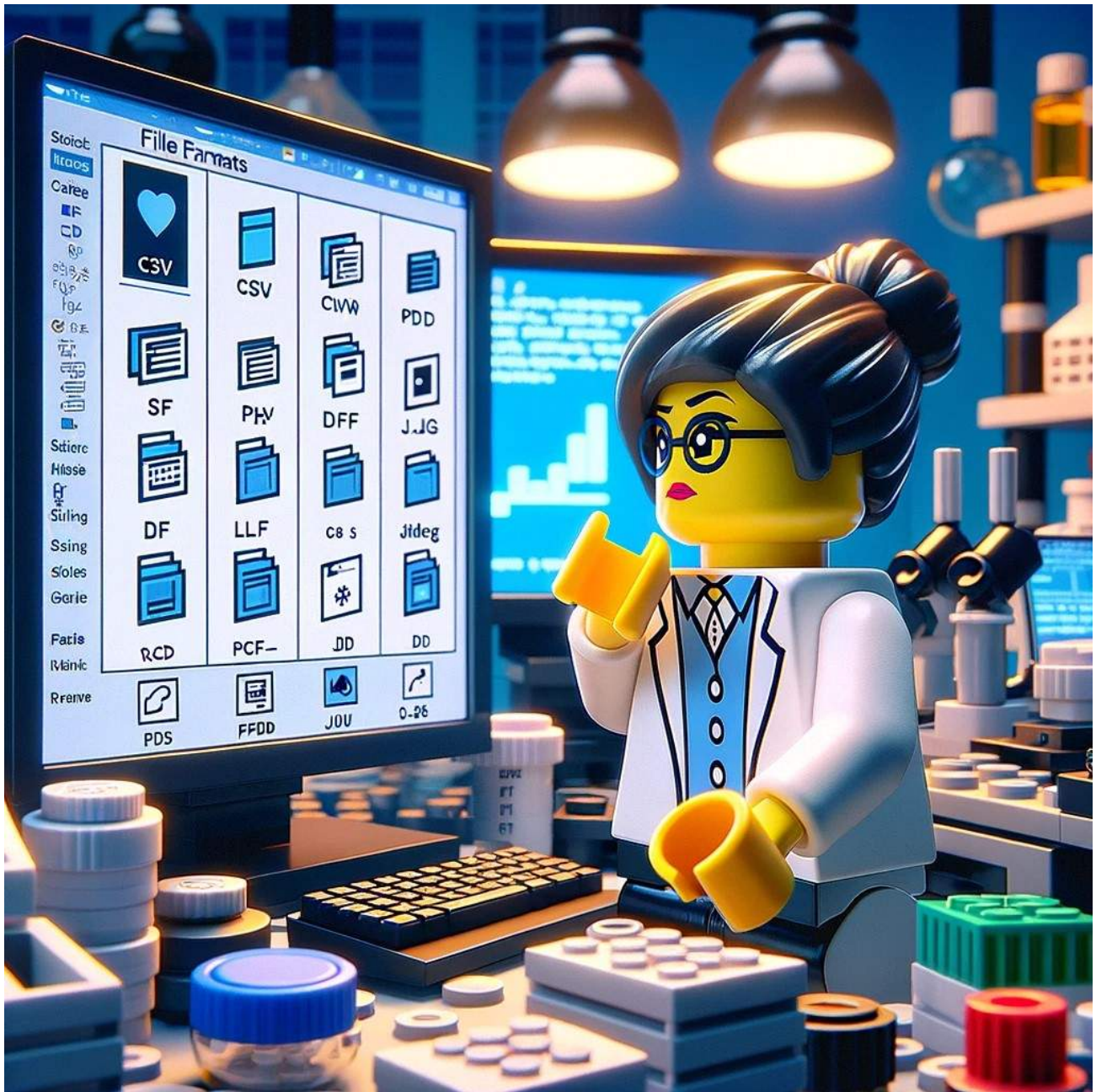
File formats

We would generally recommend that you should use file formats with open, well documented standards – these are often called *non-proprietary* formats as they do not belong to a particular company. You should try and make copies of your data in non proprietary formats, if you can do so without losing functionality.

You might find that in order to conduct your research, you need to use some proprietary or home-made software and file formats. If this is the case, that's fine – but you should make sure to provide enough documentation that someone who wants to reuse your data or assess the quality of your work can do so.

For more information on file formats, please see our File Formats for Research Data guide:

<https://edshare.gla.ac.uk/1345/>



<https://zenodo.org/doi/10.5281/zenodo.11147886>

CONTINUE

3.1 Approaching completion: Depositing your data



Researcher Development Team

This section will cover:

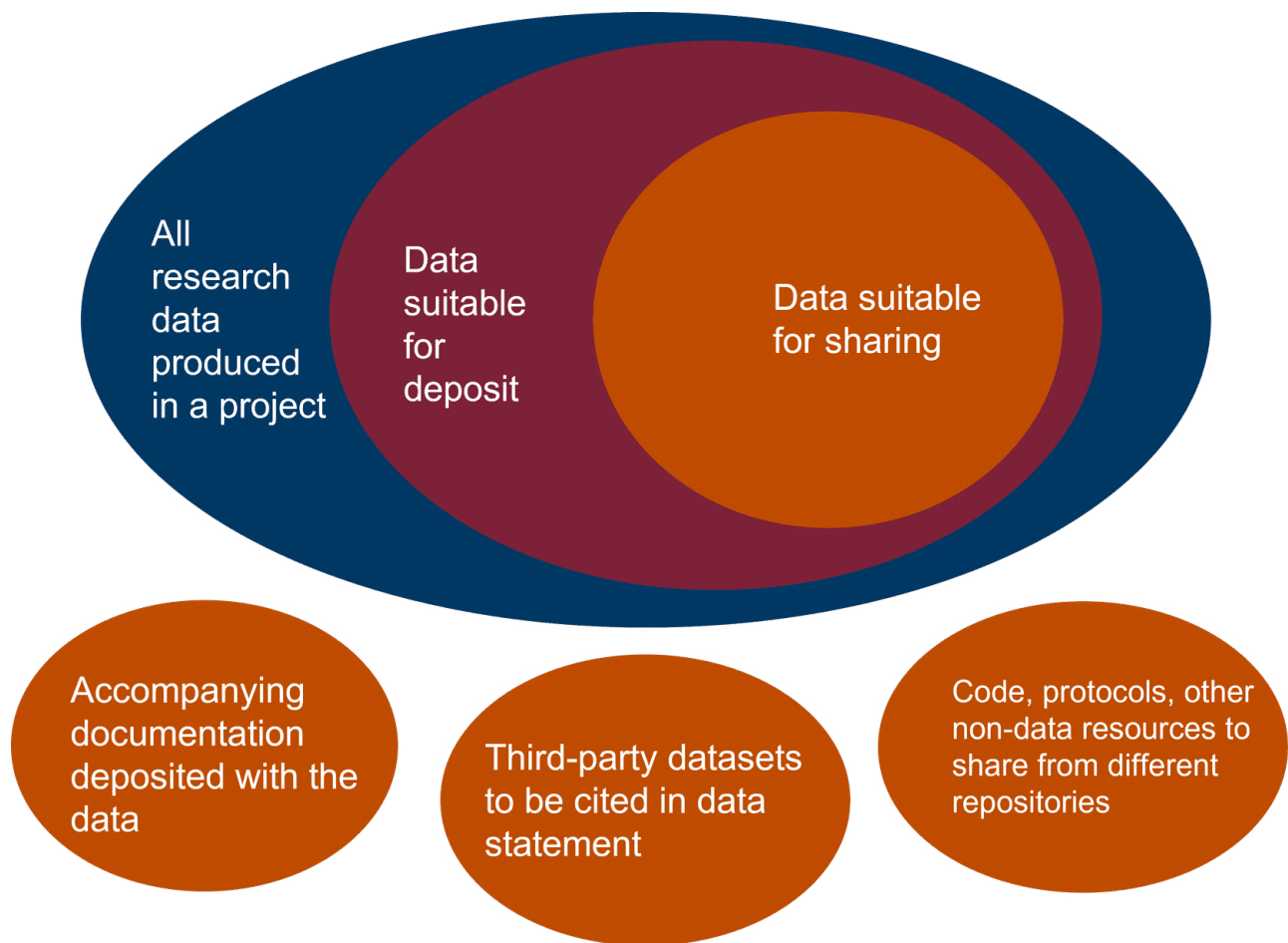
- **Depositing data at the end of a project**
- **Selecting a repository for your data**
- **Choosing a licence for your data**
- **Linking your datasets to publications and your thesis**
- **The University of Glasgow's data repository, [Enlighten](#)**

The University requires that all data of **long-term value** is deposited in an appropriate repository, where it will be stored for **at least ten years** after the end of your project. You should make your data available for sharing wherever it is appropriate to do so.

First decide what data you will deposit, and then decide which of that deposited data is suitable for reuse.

Deposit all data of long term value

We do this in order to ensure the [integrity](#) of our research by making sure a definitive copy of the data is deposited at the end of the project, and to comply with funder and University requirements.

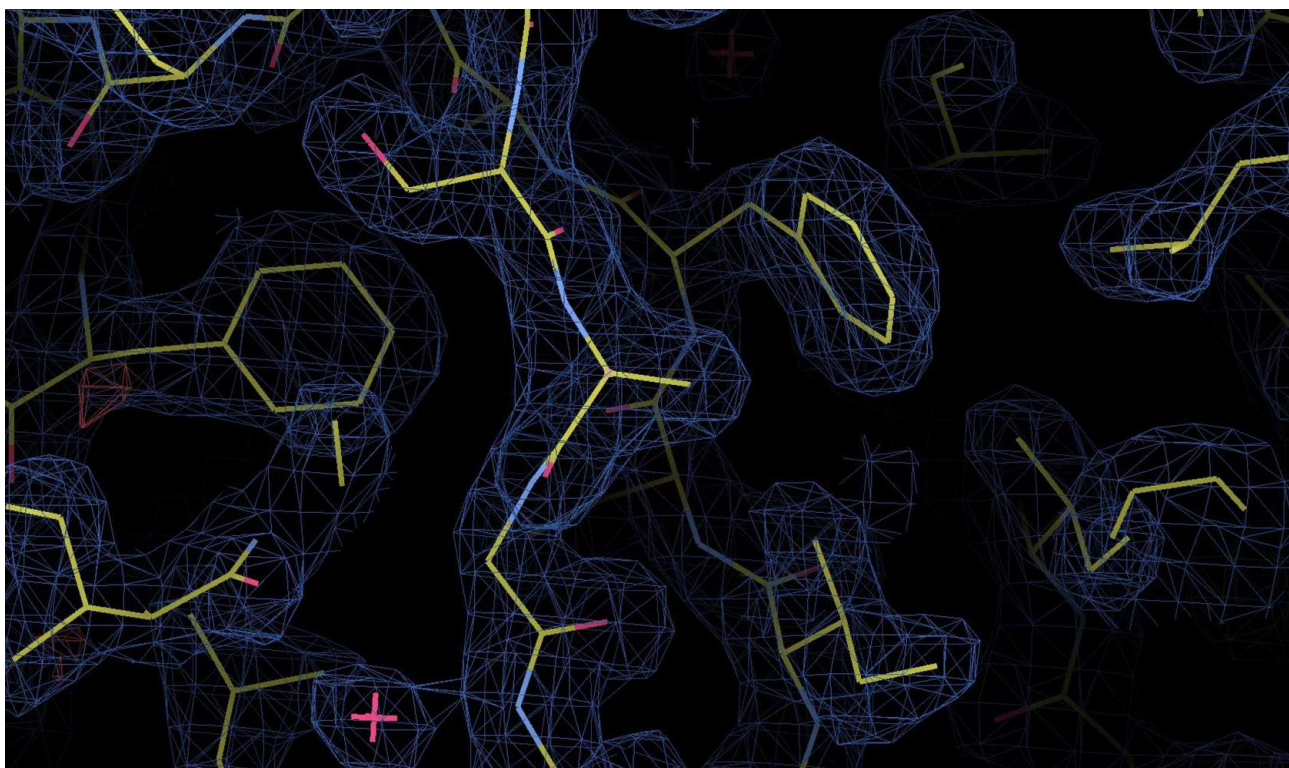


A venn diagram showing the types of data you might produce and deposit. Guidance on selection and appraisal of data for deposit is available here: <https://edshare.gla.ac.uk/1385/>

Make data available for reuse where possible

We do this in order to:

- Help others to avoid duplication of effort
- Encourage new collaborations between creators and users of data
- Increase the impact and visibility of our research (e.g. more citations)
- Meet the aims of national assessment exercises like the [Research Excellence Framework \(REF\)](#).



X-Ray Crystallography

Open data practices are increasingly common and even rewarded. For example, lots of crystallographic researchers conduct their research on open platforms like GitHub and deposit the data in the [CCDC](#).

Next we will look at depositing data in a repository.

CONTINUE

3.2 Choosing a repository and sharing data



What is a repository?

A **repository** is somewhere that stores and manages files. Repositories usually accept a **definitive version** of a dataset, and they are usually **open**, in the sense that they have a public list of their contents, even if the content can't all be freely accessed.

Repositories are where other people can find our files, and where they can potentially access and reuse them.

A repository performs a similar function to a records management service – someone hands over material, and specifies who can access it. The repository managers provide finding aids to help users see what has been deposited, preserve the material, and administrate access as appropriate.



An archive store.



An ice core repository

Repositories don't just hold text-based and digital materials – there are also lots of repositories for physical data. This picture shows a repository for ice cores. If you have any kind of physical data that you think you'll need to keep, you can contact the Research Information Management team for advice.

If you are conducting practice research, you may want to look at the guidance from the [UK Practice Research Advisory Group](#).

You may want to look at this [registry of repositories](#) to find suitable options for your data.



A messy archive

The problem is that lots of websites (and hard drives and filing cabinets) look more like this – and they’re only useful to people who already know what is there. We’d like data to be discoverable, and accessible where possible, to a wider audience.

In order to achieve this, data needs to be **well described** (which you’ll be able to do if you follow the advice we gave in section two). It also needs **finding aids** (usually prepared as part of the deposit process) and **URLs that don’t break**.

Features of a repository

404 File not Found

Could not find the file: /6105/

The specified file could not be found on this server. If you reached this page by following a link within the repository, please contact the [Enlighten: Research Data administration](#). Otherwise, please check that you have typed the URL in correctly, or contact the person or site that supplied you with this URL.

A Chrome browser showing a 404 error message

Persistent URLs

Most researchers will be familiar with the experience of trying to follow a citation in a book or article by typing out a printed URL, only to find that it doesn't go anywhere. URLs change all the time, because people restructure or abandon websites or change details like the server name, causing links to break. This is an issue for citing data because we expect the files to be accessible for at least ten years after the project ends.

The solution is a **Digital Object Identifier (DOI)**. A DOI is a URL that will always resolve to the correct location, even if the location is moved or if the server name changes. All DOI numbers begin with a 10 and contain a prefix and a suffix separated by a slash. DOIs can be assigned to any piece of content, whether it is digital or physical.

A dataset in Enlighten: [10.5525/gla.researchdata.191](#)

A paper in the Journal of the National Cancer Institute: [10.1093/jnci/djv204](#)

You need a DOI to cite your data properly in publications and theses, so you should make sure you have deposited your data in a repository and received a DOI before you publish an article or submit a thesis. You should always use DOIs in your citations instead of URLs if they are available.



The **correct citation format** for data in a thesis or publication is:
“The data that underpin this research are available from [DOIs].“

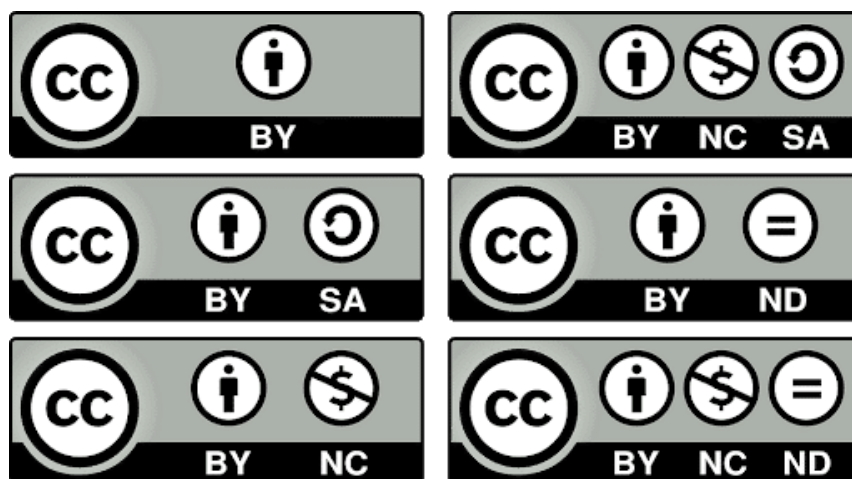
You can include this text in your **data accessibility statement**, which should appear as a note at the start of your thesis or article, or in a footnote or endnote in an article if the journal does not have a dedicated place for one.

For more detailed guidance on data accessibility statements, including examples, see our guide to Identifiers and Citation: <https://edshare.gla.ac.uk/1419/>

Licences and access conditions

When you put your data in a repository, you will need to give it a **licence**. A licence tells potential users of the data what they can do with it. When you're planning to deposit data you can start by dividing it into two categories: data that can be **open access**, and data that should be **restricted or closed**. You should take into account the preferences of your funders and any project partners when choosing licences for your data.

Open access data is that which is freely available for other people to reuse, with no barrier to entry. A user will not need to complete a licence, register an account or pay any fee to get open access data. **You should make your data open access unless you have a compelling reason for restricting it.**



The most common open access licences are [Creative Commons](https://creativecommons.org/), which are identified by their grey logos. The default licence for open access data in the University's Enlighten: Research Data repository is CC-BY, which means that users must attribute the original author of the material but are otherwise free to reuse it as they wish.

Restricted data is that which cannot be accessed without some additional procedures like completing a user licence or submitting a request for access. In some cases, users requesting access to restricted data will need to submit a new application for ethical approval.

Closed data is that which is deposited for long-term storage but cannot be shared with anyone.

If you think you have some data that needs to be restricted or closed, you should look at the policies of your chosen repository to see what options are available.

The [Enlighten: Research Data](#) repository has a range of access options for data, including open access, controlled, closed and dark data:

<https://www.gla.ac.uk/myglasgow/openresearch/researchdatamanagement/enlightenresearchdatainformation/#accessoptions>

Another good example of tiers of access can be found on the UK Data Archive: <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control.aspx>.

Choosing a repository

There are a number of generic and subject-specific repositories you could use. Below are just a few examples:

- Generic: [Figshare](#), [Zenodo](#)
- Funder-provided: [NERC data centres](#), [UK Data Service \(ESRC\)](#), [Wellcome Open Research](#)
- Subject specific: [Dryad](#) for Ecology, [ADS](#) for Archaeology
- Institutional: [Enlighten: Research Data](#)
- For code: [Github](#), [Bitbucket](#) etc. You can produce DOIs for code using GitHub: <https://guides.github.com/activities/citable-code/>.



Re3data.org

Enlighten: Research Data

Mixed methods survey of zoonotic disease awareness and practice among animal and human healthcare providers in Moshi, Tanzania

Zhang, H. and Mnzava, K. and Mitchell, S. and Melubo, M. and Kibona, T. and Cleaveland, S. and Kazwala, R. R. and Crump, J. A. and Sharp, J. and Halliday, J. E. B. (2016) *Mixed methods survey of zoonotic disease awareness and practice among animal and human healthcare providers in Moshi, Tanzania*. [Data Collection] (Unpublished)

Datacite DOI: [10.5525/gla.researchdata.262](https://doi.org/10.5525/gla.researchdata.262)

College / School:	College of Medical Veterinary and Life Sciences > Institute of Biodiversity Animal Health and Comparative Medicine
Date Deposited:	02 Feb 2016 09:58
Enlighten URL:	http://eprints.gla.ac.uk/107867/
Funder's Name:	Biotechnology and Biological Sciences Research Council (BBSRC), Biotechnology and Biological Sciences Research Council (BBSRC), Biotechnology and Biological Sciences Research Council (BBSRC)
URI:	http://researchdata.gla.ac.uk/id/eprint/262

Available Files

Data

 [SemiStructuredI ... angPLoSNTDs.pdf](#)

 [QuestionnaireDa ... ngPLoSNTDs.xlsx](#)

Additional details

Cite this record

Zhang, H. and Mnzava, K. and Mitchell, S. and Melubo, M. and Kibona, T. and Cleaveland, S. and Kazwala, R. R. and Crump, J. A. and Sharp, J. and Halliday, J. E. B. (2016); *Mixed methods survey of zoonotic disease awareness and practice among animal and human healthcare providers in Moshi, Tanzania*

University of Glasgow

[10.5525/gla.researchdata.262](https://doi.org/10.5525/gla.researchdata.262)

Retrieved: 2016-03-02

A dataset in Enlighten: Research Data

Our institutional repository is called [Enlighten: Research Data](#). All staff and students can deposit research data here. Enlighten: Research Data meets a high level of information security which means that we can store very sensitive data, provided you have permission to deposit it. We can store open access, controlled, close and dark data. Some [deposit charges may apply](#); contact us if you think that your dataset will be larger than 100GB.

The process for depositing in Enlighten: Research Data is as follows:

Preparing for deposit

Prepare your data for deposit by dividing your data into by [access condition](#). Make sure that you have permission from participants or rights holders to deposit all your data.

Step 2

Notify us

Let us know that you plan to deposit your dataset by emailing us at research-datamanagement@glasgow.ac.uk. We will ask you to complete a [deposit form](#) where you provide enough information for us to create a record for your data. At this point we can reserve a DOI for you to include in your thesis. If your dataset is likely to be over 100GB, get in touch as soon as possible, as some charges may apply.

Step 3

Send us the data

Send us your data in separate zipped folders for each access condition that applies. Send us Readme files and any other useful documentation (but don't zip these!). If you had ethical approval for your project, send us a **blank** consent form and information sheet.

You can send us the data by email or by [Transfer](#). If your dataset is particularly large we can arrange to collect it.

If your data is at all sensitive, you should not send it by email.

Depositing the data

We will deposit the data and activate your DOI - this could happen immediately or upon publication of your article.

If the dataset is open access, it will be available for users to download. If restrictions apply, we will pass access requests on to the appropriate role within the University or ask users to complete a licence before accessing the data.



You should also get an **ORCID - a persistent digital identifier** that distinguishes you from other researchers and helps you to aggregate your outputs (datasets, articles etc). Registration takes 30 seconds at <https://gla.ac.uk/orcid>.

CONTINUE

4. Summary and quiz



This concludes the Research Data Management course - thanks for your attention, and good luck with your research!

Remember that you can [contact the Research Information Management team](#) at any time with questions relating to your research data, open access for publications and other research outputs, and open research more broadly. The team are:

Gail Annan, Alastair Arthur, Jacqui Brannan, Mary Donaldson, Valerie McCutcheon, Matt Mahon and Lisa Turner Warnecke. You can reach us at research-datamanagement@glasgow.ac.uk.

Please return to Moodle to **complete a short quiz** so that your attendance can be recorded. If you've got any follow-up questions, please email us at the address above, and keep an eye out for our regular drop-in advice sessions, which will be advertised through your College, the [Research Staff Assembly](#) and the [Research Culture Commons](#).