Introduction to Research Data Management 2022-23

Introduction to Research Data Management

INTRODUCTION

4. Summary and quiz

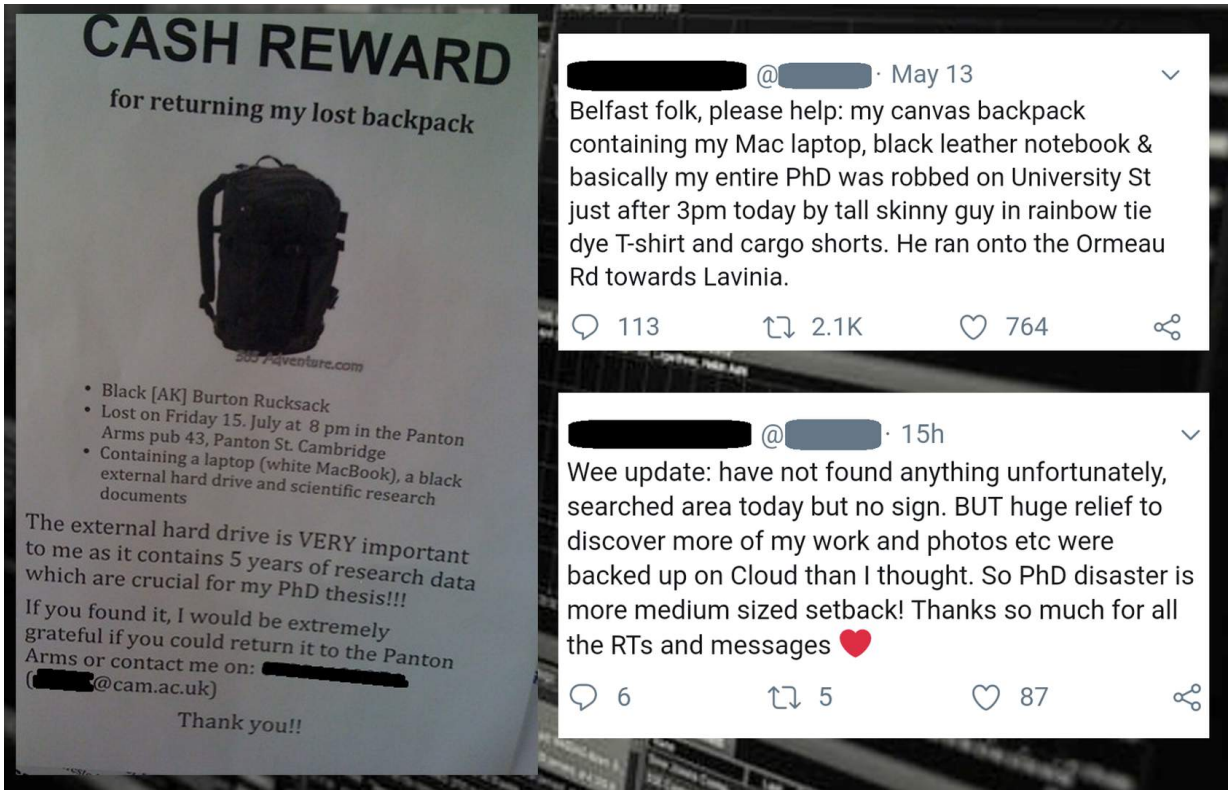
Welcome to the course

Welcome to the I**ntroduction 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 point to external resources that you can read if you want. 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.

A lot of this advice is common sense, so while you may not have thought of your 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.

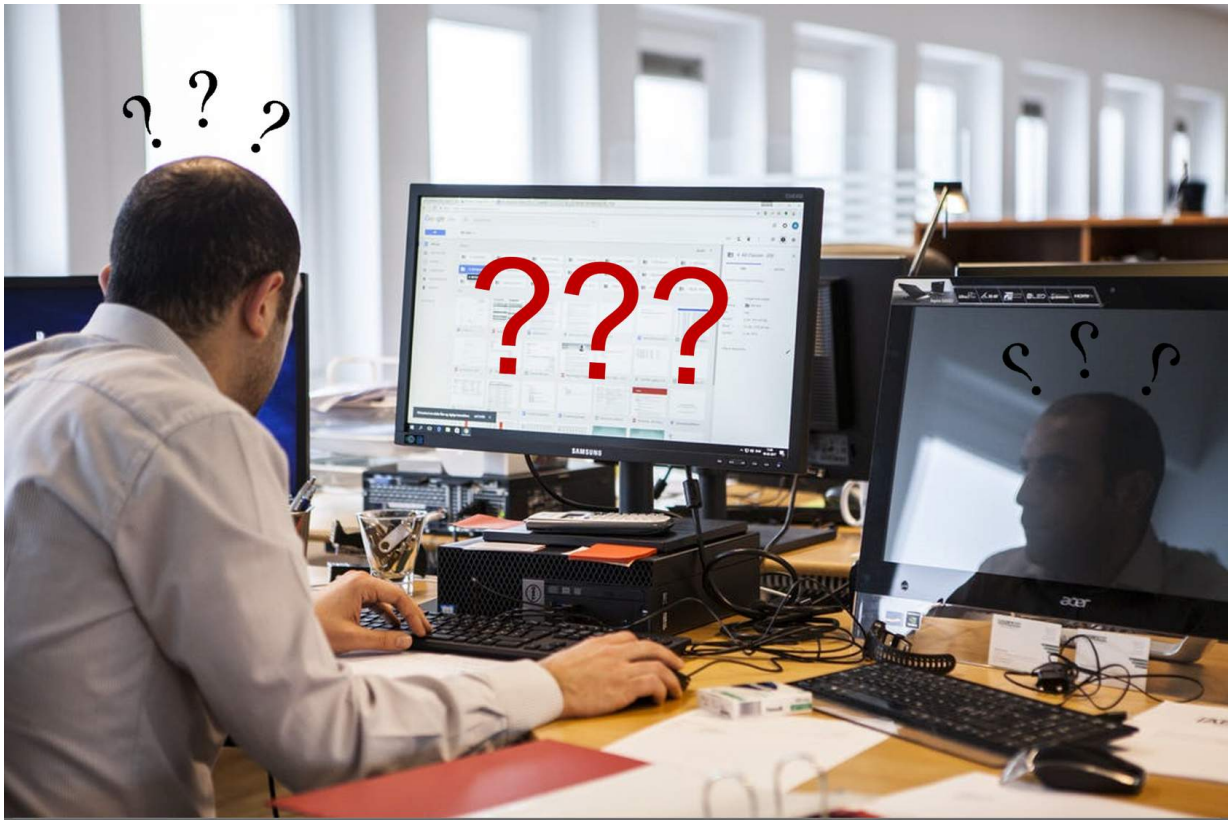


Examples of catastrophic (and not-so-catastrophic) failures of data management. CC-BY

Avoiding disasters

The image above shows some of the worst examples of research data disasters - researchers who have 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.



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**!

CONTINUE

What is research data?

Types of research data.

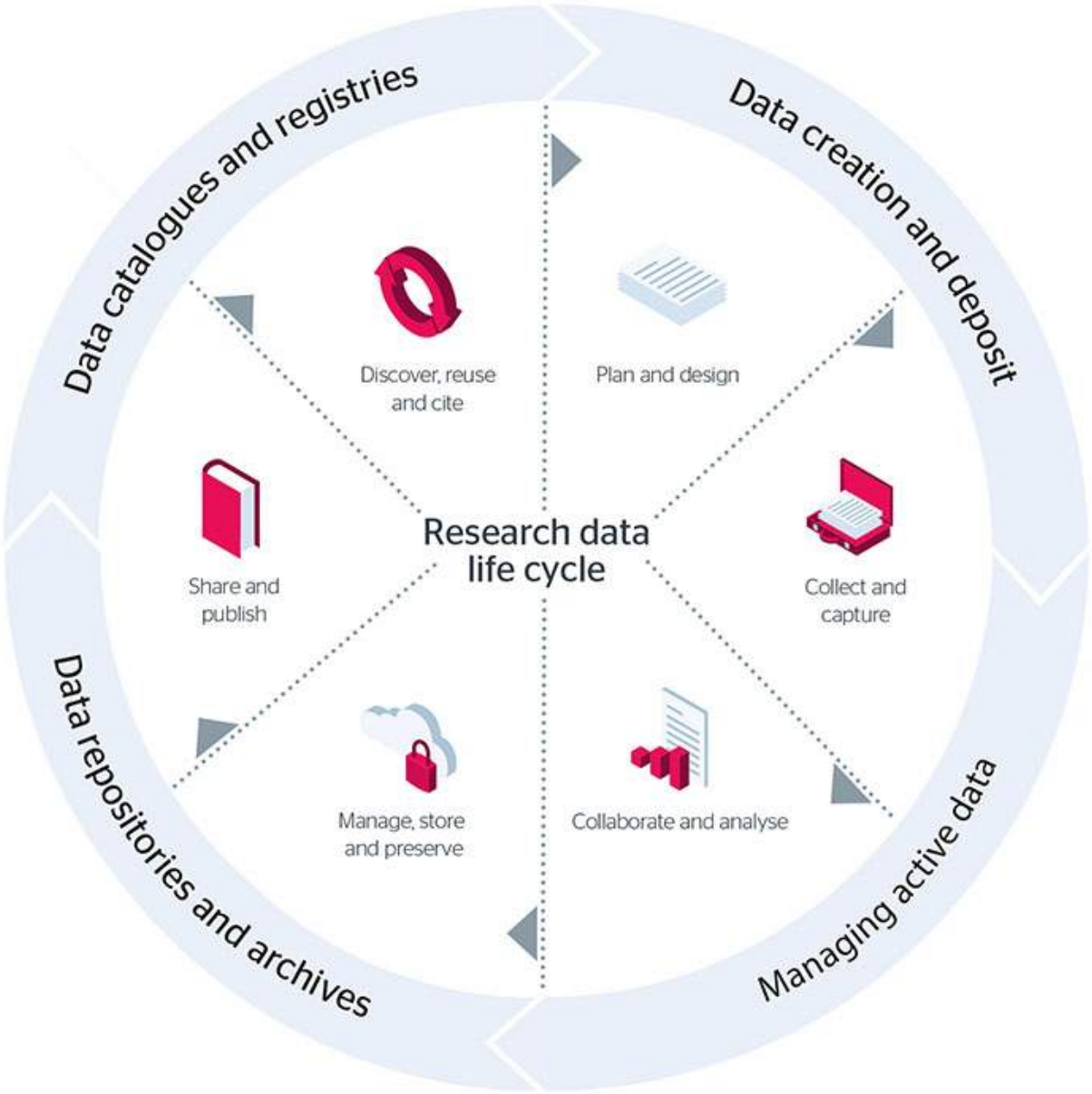
Research data is **any information, whether digital or physical, required to underpin research**. It is difficult to define what constitutes data for any given project, and it will vary by discipline or approach. As an expert in your field, you have a lot of leeway to determine what data you think are 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?

Further reading

Definitions of research data from Leicester University.

Guide to research data in the Arts, Humanities and Social Sciences from RDMToolkit



Research data lifecycle diagram, ©Jisc, CC BY-NC-ND

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**.

Further reading

University of Glasgow RDM: Before you Begin

The Jisc Research Data Lifecycle from Jisc RDMToolkit

We have divided the research data lifecycle 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 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 depositing your data in an appropriate repository, selecting which data to deposit, choosing a licence 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

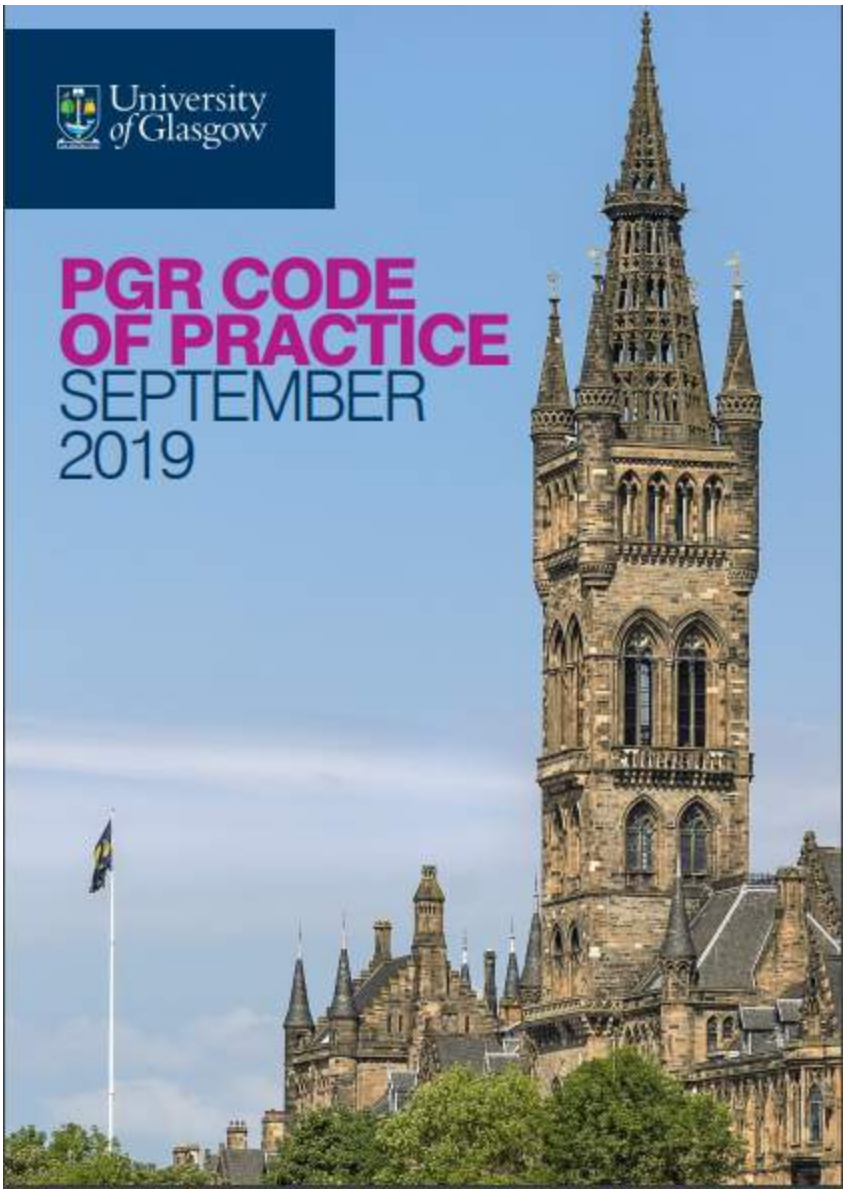
1. Before You Begin

Taco Ekkel https://flic.kr/p/5JGDNw

This section will cover:

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

- Working with sensitive data



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**;

the **Postgraduate Research Code of Practice;**

and the **Good Management of Research Data Policy**.

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 Policies page.

The Code of Good Practice in Research

Section 3.6 of the Code of Good Practice in Research explains the University's expectations regarding your data. You should:

Keep clear and accurate records of research

Prepare a data management plan for projects which will generate data

Contextualise research data by keeping sufficient metadata

This means that you need to produce documentation that will enable other researchers in your field to understand what you did.

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.

Deposit a copy of data generated using University resources when 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 as long as it's deposited somewhere 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 in copyright, 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.

UK Research and Innovation logos.

Funders' data requirements

Some funders, notably the UK Research Councils (UKRI) and charities like Wellcome 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.'**

https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

Activity

Familiarise yourself with the research data policies of a couple of funders. A good place to start is the Digital Curation Coalition'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' polices, consider: will you need to do anything in addition to the University's data requirements in order 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

Data Management Plans

Writing a plan.

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 the course of your project. The first draft of a DMP is usually quite straightforward, and as your practices develop the plan becomes more detailed. DMPs usually cover the following areas:

- What data will you collect?
- How will it be documented?
- Ethics and Intellectual Property
- Storage and organisation of the data

- Deposit and long-term preservation
- Plans for data sharing
- Responsibility for implementation

A messy archive box

Data Management Plans are not 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, 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 an up-to-date 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 PGR 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 PGR template regardless of their funder.

UofG Data Management Plan template:
http://eprints.gla.ac.uk/179057/

UofG Data Management Plan guidance:
http://eprints.gla.ac.uk/179058/

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 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 in 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. Below is a workflow for initiating projects involving **personal data**.

DMPs, Ethics and GDPR

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.

1

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.

2

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.

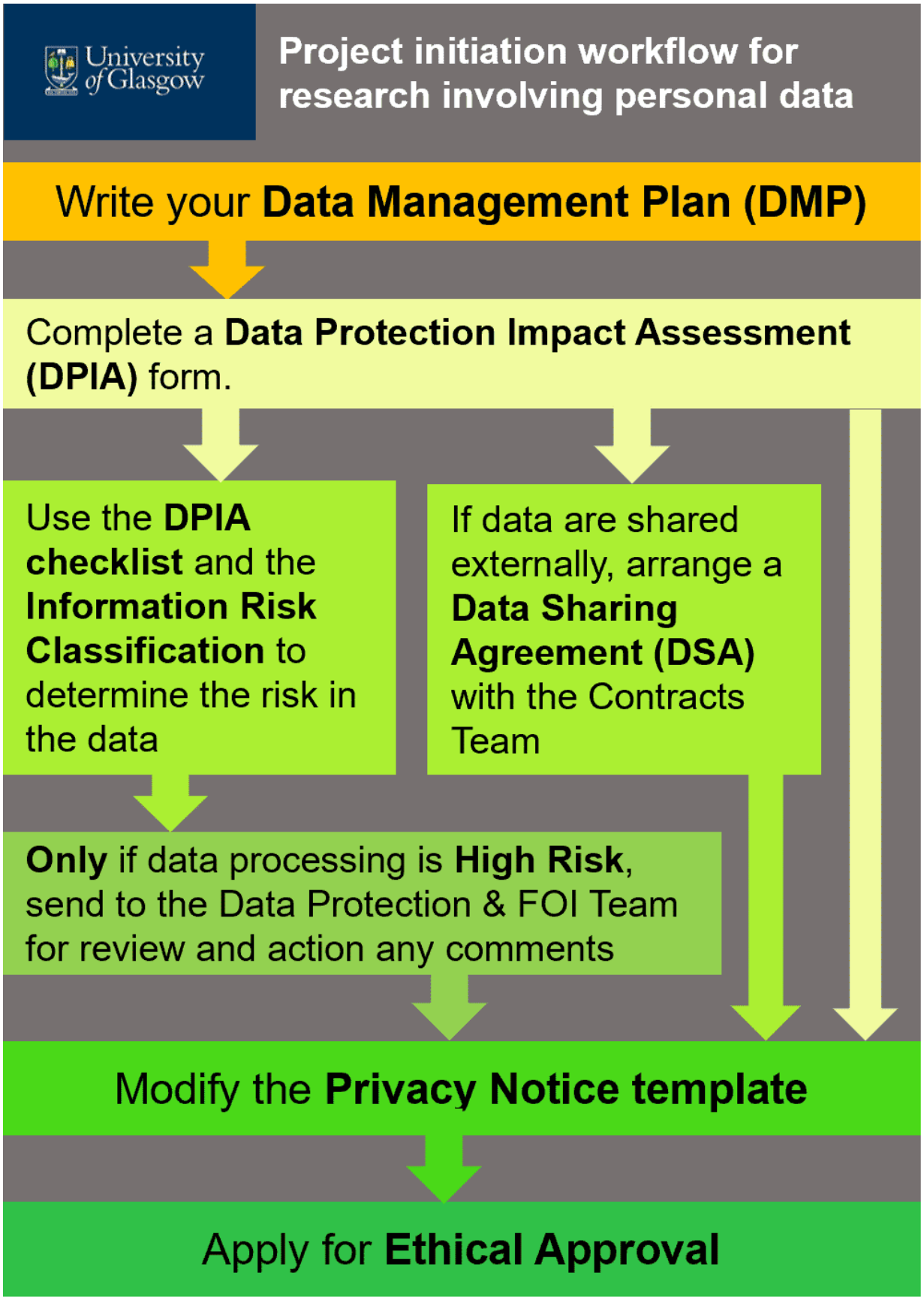
3

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

An expanded version of this workflow, including links to relevant documents and support, is available from our repository: 10.36399/gla.pubs.202746.


University of Glasgow
Project initiation workflow for research involving personal data
Write your Data Management Plan (DMP)
Complete a Data Protection Impact Assessment (DPIA) form.
Use the DPIA checklist and the Information Risk Classification to determine the risk in the data
If data are shared externally, arrange a Data Sharing Agreement (DSA) with the Contracts Team
Only if data processing is High Risk, send to the Data Protection & FOI Team for review and action any comments
Modify the Privacy Notice template
Apply for Ethical Approval

https://doi.org/10.36399/gla.pubs.202746

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

CONTINUE

Ethics and data protection

Examples of personal data.

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.

http://www.gla.ac.uk/services/rsio/researchstrategypolicies/ourpolicies/committeestructure/

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

Every project collecting or processing human data need a DPIA. 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 checklist and Information Risk Classification to work out 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 data 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.

Working with sensitive data

This guide from the UK Data Archive is full of valuable best practice guidance, including information on handling sensitive data. Our best practice advice is:

Aggregate your results wherever possible

Avoid collecting personal data if you don't need it for your research question

De-identify your data by removing personal information, and keep a log of what types of personal information you have removed.

Assess whether you have sufficiently de-identified your data that it can be considered anonymous.

If your data cannot be effectively anonymised, restrict access.

Intellectual Property

You may find that during the course of your research you produce some intellectual property that you would like to secure. Our advice for protecting your intellectual property rights is to:

- Keep good records and gather evidence and a paper trail. Record the IPR in detail at the time it is created.
- Don’t disclose your intellectual property (e.g. in a publication or seminar) until it is evaluated. Use confidentiality agreements if talking with external parties.
- Act quickly and decisively, and consult a specialist immediately. They will help you to consider the options and cost.

Consult the IP and Commercialisation office:
https://www.gla.ac.uk/myglasgow/ris/contact/ipcommercialisation/.

Further reading

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?

https://www.gla.ac.uk/myglasgow/openresearch/openresearchresources/#datamanagementcasestudies

If you are engaged in research and creating or working with open data and would be 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. Research in progress

Jenni Waterloo https://flic.kr/p/6dCbZY

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 minimal acceptable standard of documentation.

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 badly managed data, so you should always consider:

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

For example, you might consider including:

- 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!

Always consider:

Are your files clearly described?

Are your data clearly labelled with versions and dates?

Are your data logically structured and named?

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.

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.

Next we will look at file naming and folder structure.

CONTINUE

File names, folder structure and formats

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
·Project or experiment name of acronym ·Location / spatial coordinates ·Researcher name / initials ·Date or date range of data acquisition ·Type of data ·Conditions ·Version number of file ·File extension	·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

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.

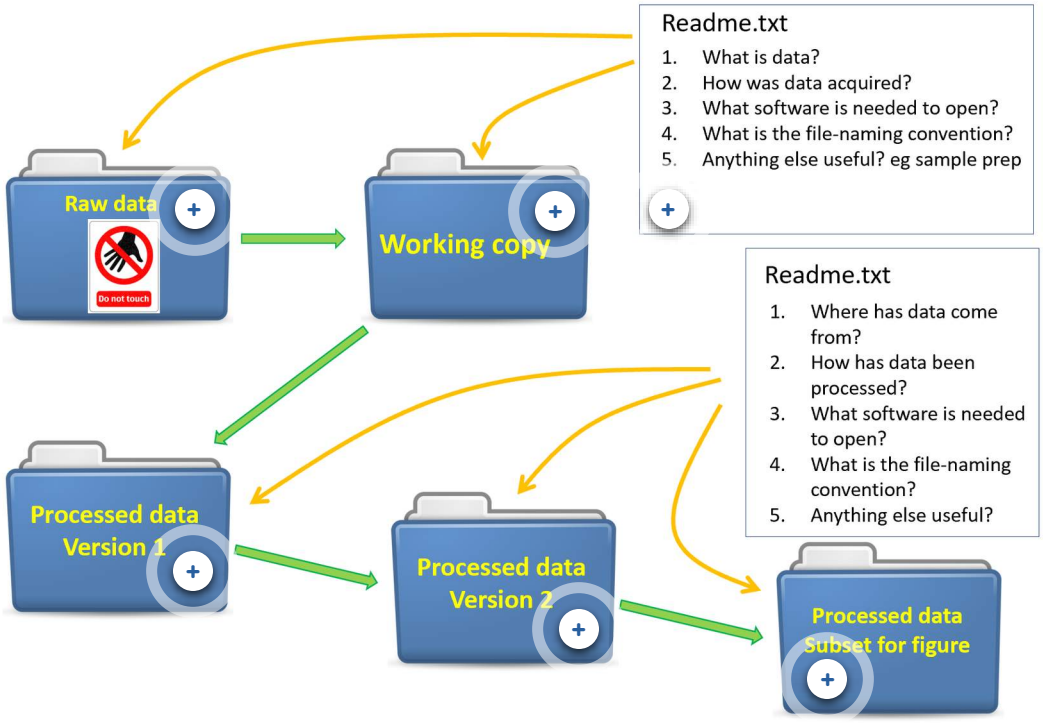
Further reading

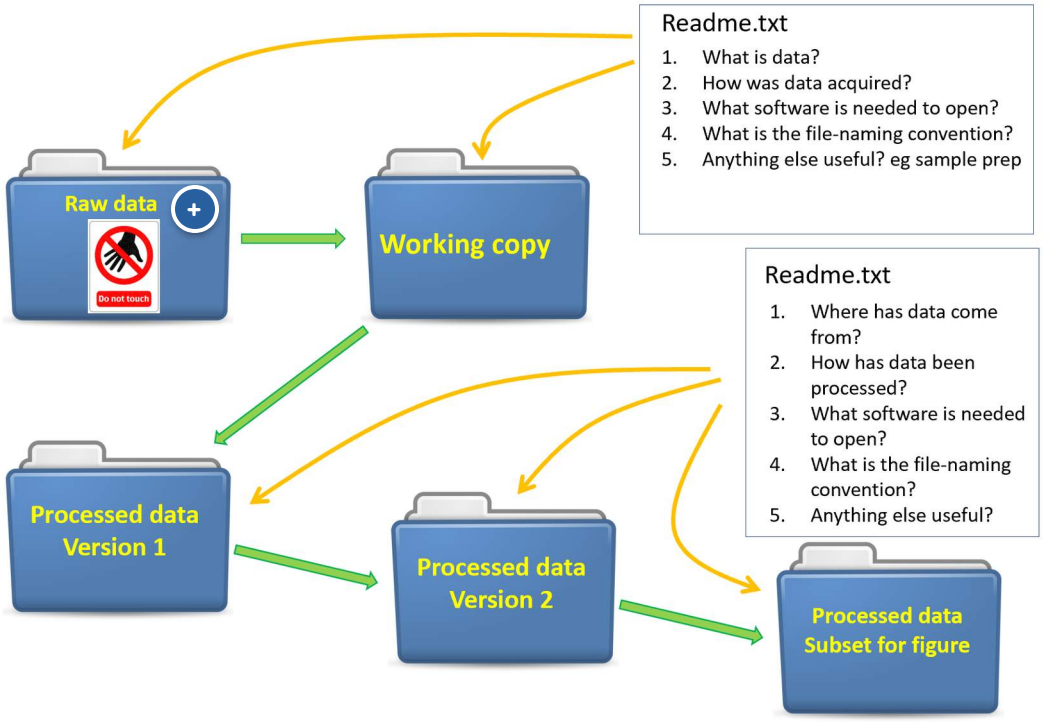
For more information on file naming, see our file naming guide.

CONTINUE

Folder structure

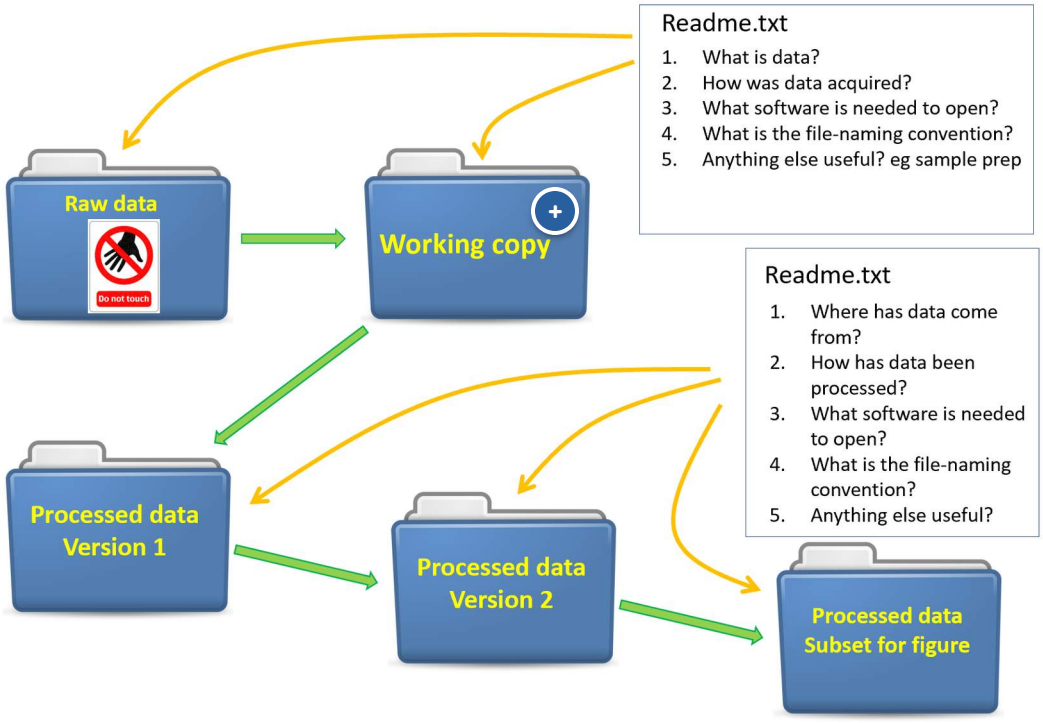
The diagram below illustrates our recommended folder structure for research data.


Readme.txt
1. What is data?
2. How was data acquired?
3. What software is needed to open?
4. What is the file-naming convention?
5. Anything else useful? eg sample prep
Raw data
Do not touch
Working copy
Readme.txt
1. Where has data come from?
2. How has data been processed?
3. What software is needed to open?
4. What is the file-naming convention?
5. Anything else useful?
Processed data Version 1
Processed data Version 2
Processed data Subset for figure



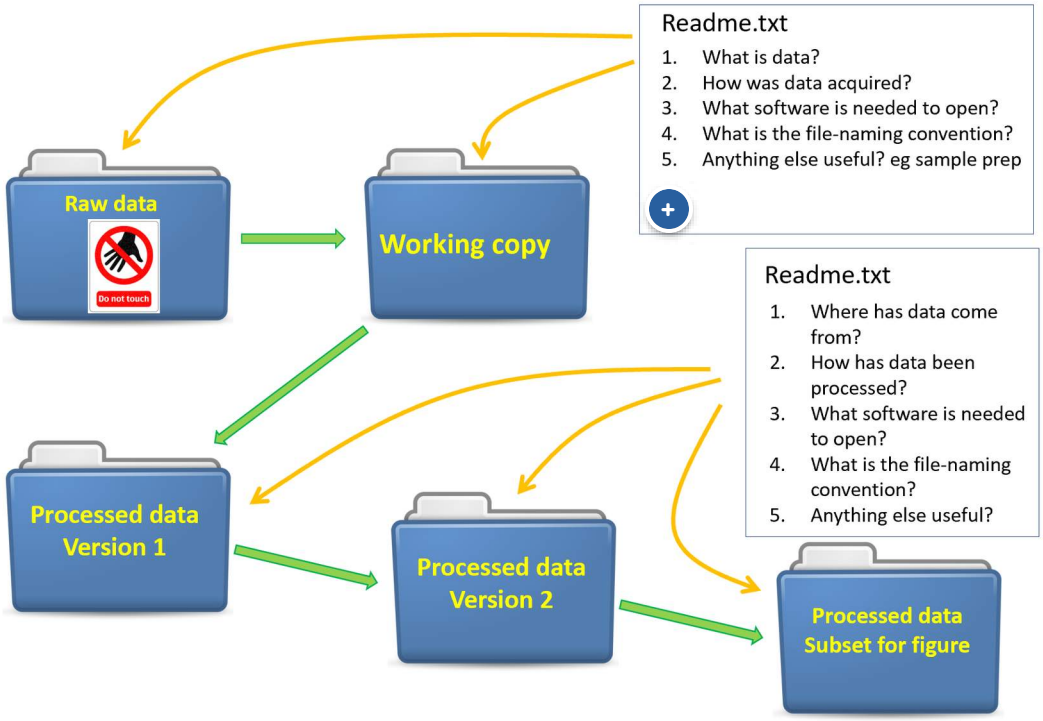
Raw data

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.



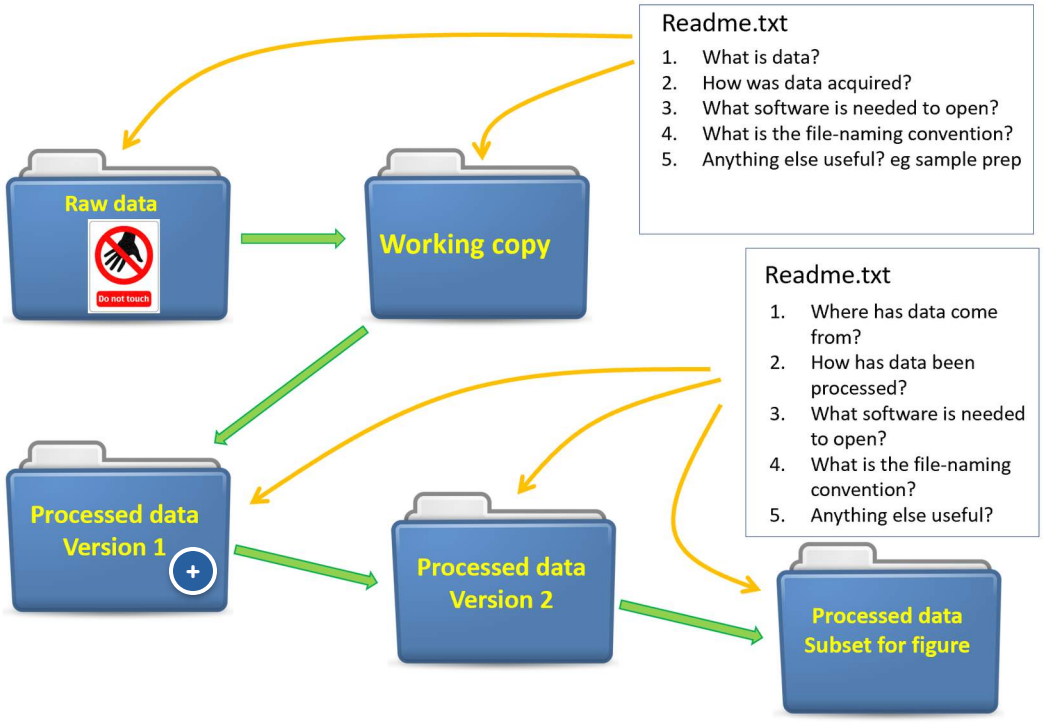
Working copy

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.



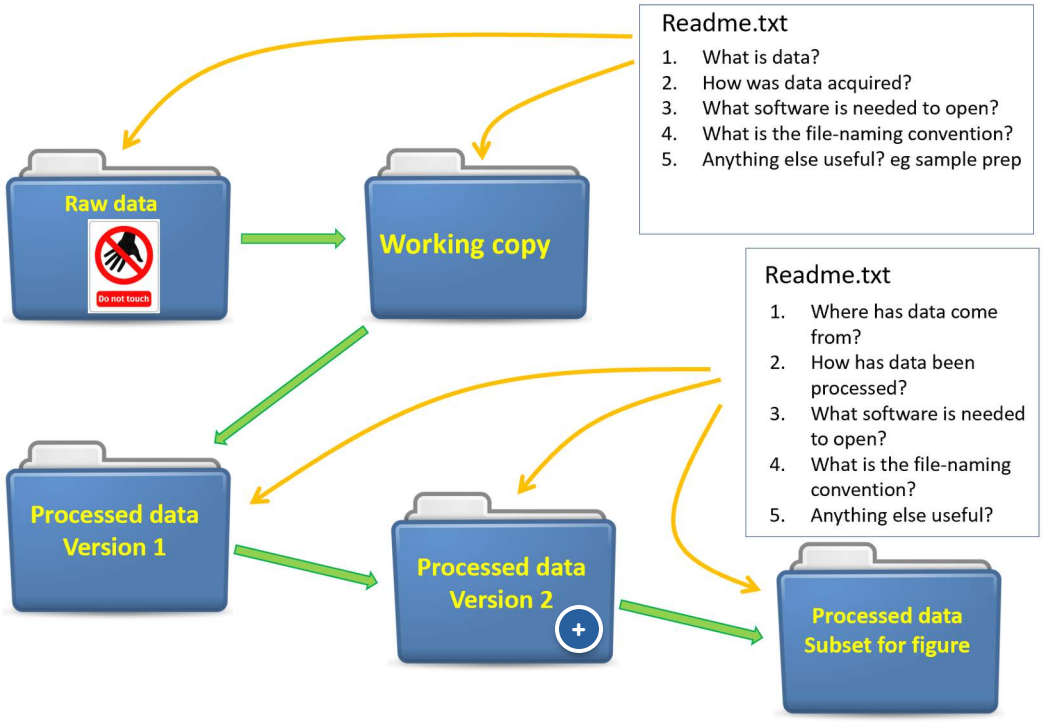
Readme.txt

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.



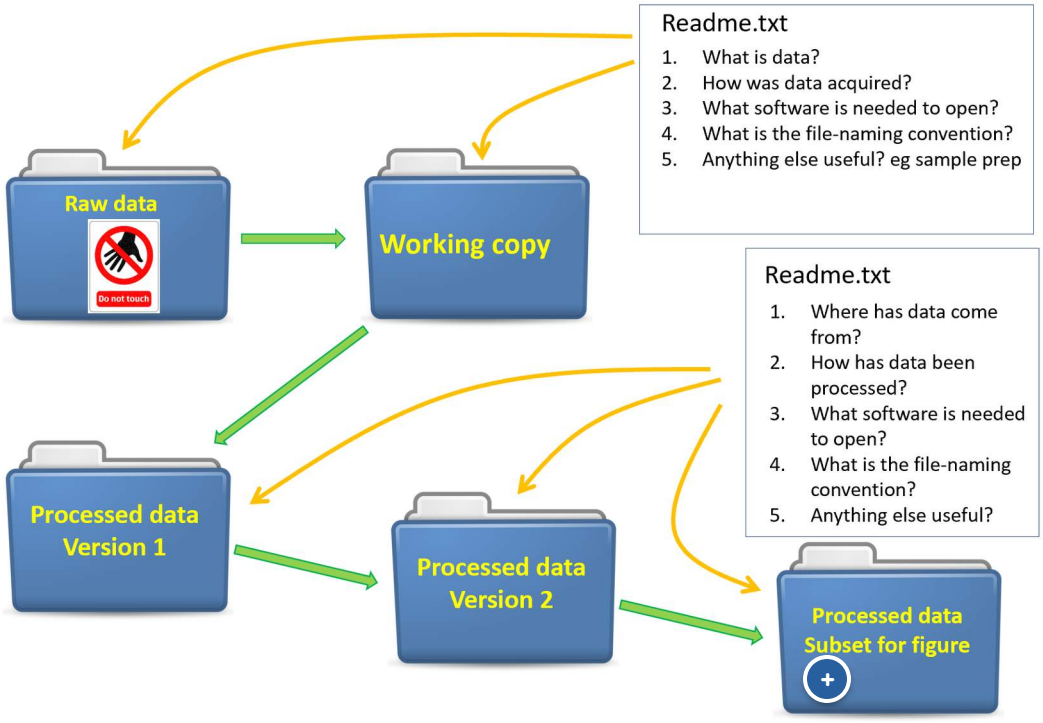
Processed data

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 change it consciously and still enjoy the 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 for figure

Anonymised extract from interview

Subset of results used to produce a visualisation

Illustration

Additional documentation

Workflow documentation produced by statistical tool

Semi-structured interview template

Record of types of identifier removed from personal data

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/

Further reading

University of Glasgow RDM: Guidance on file names and folder structure

The US Library of Congress hold the definitive list of recommended file formats: https://www.loc.gov/preservation/resources/rfs/

The UK Data Service has prepared a guide to recommended file formats: https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

CONTINUE

Storing data while you work on it

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 GDPR 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?

CC-BY

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 3 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.

	J:\	Nextcloud	OneDrive	
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	✓	✓	✓	✗

	J:\	Nextcloud	OneDrive	
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 GDPR.

	J:\	Nextcloud	OneDrive	
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	✓	✓	✓	✗

	J:\	Nextcloud	OneDrive	
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.

	J:\	Nextcloud	OneDrive	
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	✓	✓	✓	✗

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 OwnCloud are on campus.

> Free personal cloud storage must not be used for 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 GDPR compliant, 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. Confidential research data is personal data or data that is otherwise under a confidentiality agreement - for example if it comes from an industry partner.

- 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

Further reading

Guidance on data storage during your project:
https://www.gla.ac.uk/myglasgow/openresearch/researchdatamanagement/datastorageforresearchinprogress/

Take a look at the guidance on handling confidential data: https://www.gla.ac.uk/myglasgow/it/informationsecurity.

Choosing which data to retain after the project

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 the University's or a funder's policies, from a collaboration agreement between partners, from your ethical approval. You may also have statutory (legal) obligations - for example, certain medical data needs to be retained for a statutory period.

Click the slides below to read more about some considerations when deciding what data to retain.

Personal Data?

Do you need to keep personal data, or is your de-identified data enough to show how you reached your conclusions? Removing personal data is a good way to reduce the risk of accidentally disclosing

Raw and processed data?

Most researchers retain their raw data, the final processed version, and any intermediate processed versions that someone would need to see to understand how the dataset was produced.

Different versions for deposit?

If your dataset is going to be restricted (for example if it contains personal data), or if it is too large to easily download, you might want

to make a small (anonymised) sample

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 any 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/

Next we will look at depositing and sharing your research data.

CONTINUE

3. Depositing your data

This section will cover the process for depositing your data in an appropriate repository, choosing a licence for your data, and linking your datasets to publications and your thesis.

The University requires that all your 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

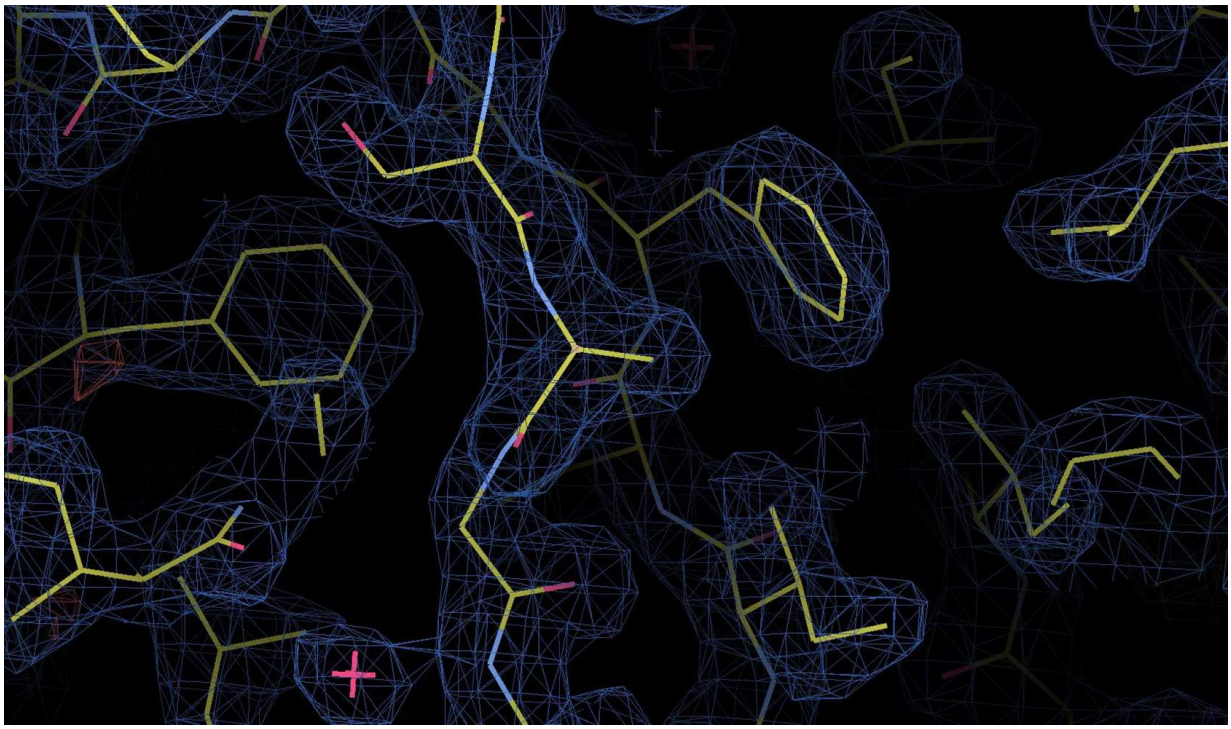
We do this in order to ensure the integrity of our research by making sure a definitive copy of the data is deposited at the end of the project, and to comply with funder and University requirements.

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 in some disciplines. For example, lots of crystallographic researchers conduct their research on open platforms like GitHub and deposit the data in the CCDC.

Further reading

Research Integrity at the University of Glasgow: https://www.gla.ac.uk/myglasgow/ris/researchpolicies/researchintegrity/

Open Research at the University of Glasgow: https://www.gla.ac.uk/myglasgow/openresearch/

Next we will look at depositing data in a repository.

CONTINUE

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 archive staff 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, but 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.

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

A Chrome 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 submit your work. 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]."

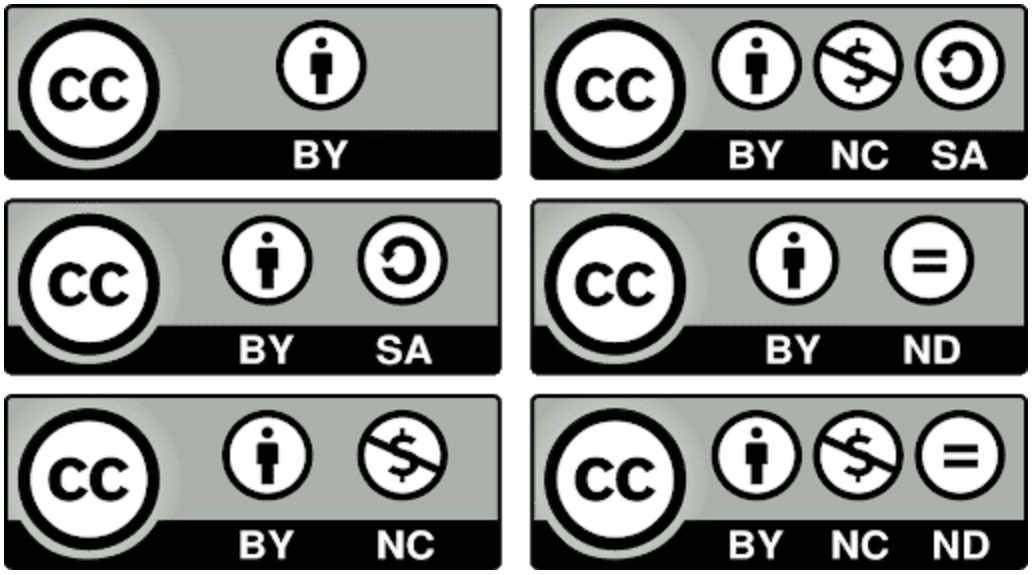
You can include this text in your **data accessibility statement**, which should appear as a note at the start of your thesis or article, or in a footnote or endnote in an article if the journal does not have a dedicated place for one.

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

Licences

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.** 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, which are identified by their grey logos. The default licence for open access data in the University's repository is CC-BY, which means that users must attribute the original author of the material but are otherwise free to re-use 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.

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.

A 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.

Further reading

Guide to Open Access at the University of Glasgow:
https://www.gla.ac.uk/myglasgow/openresearch/openaccess/

DOIs and other identifiers at the University of Glasgow: https://edshare.gla.ac.uk/1419/

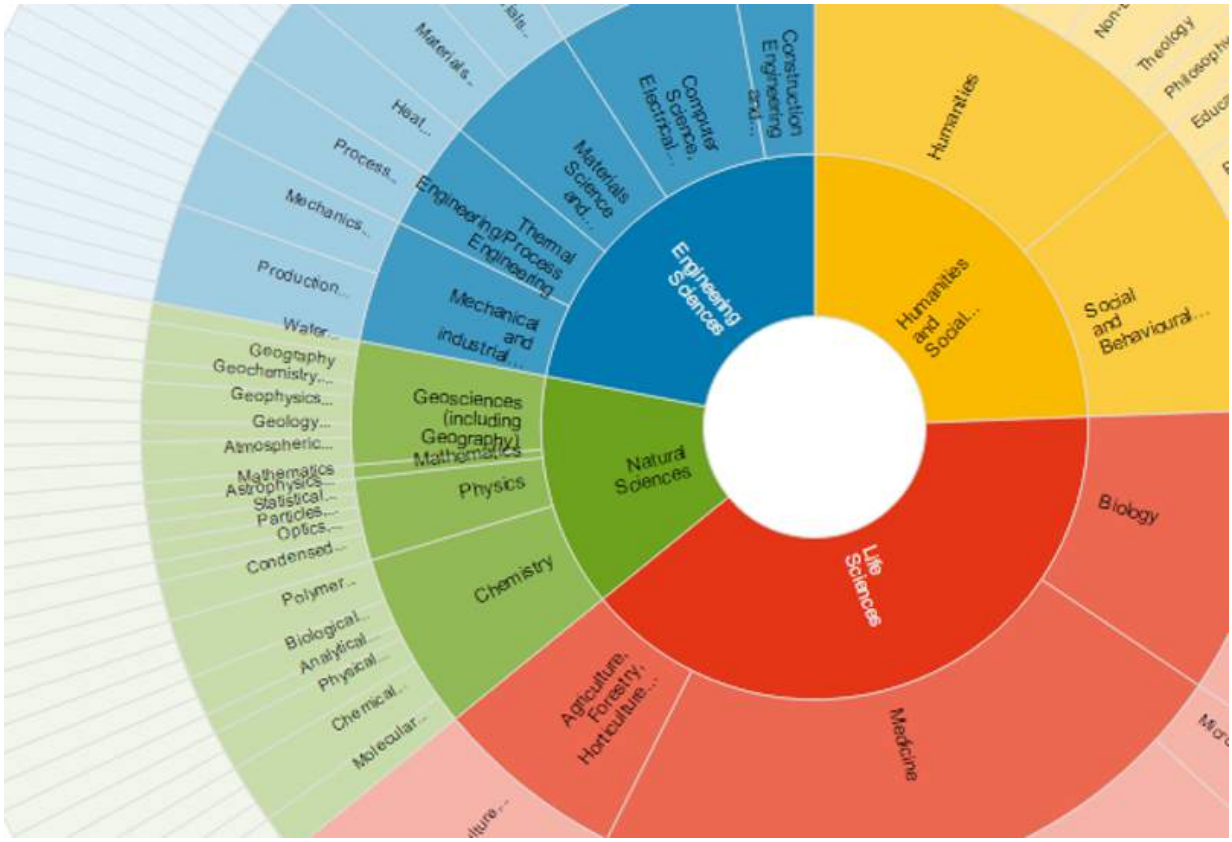
Creative commons licences: https://creativecommons.org/licenses/

Dataset Licensing: https://datasetlicencing.wordpress.com/outputs/

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

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 or restricted data. Some deposit charges may apply; contact us if you think that your dataset will be larger than 5GB.

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

Preparing for deposit

Prepare your data for deposit by dividing your data into open access data and restricted data. Make sure that you have permission from participants or rights holders to deposit all your data.

1

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 5GB, get in touch as soon as possible, as some charges may apply.

2

Send us the data

Send us your open access and restricted data in separate zipped folders. 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 it is restricted, we will pass access requests on to the ethics committee that approved the original application.

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:

Matt Mahon, Mary Donaldson, Mick Eadie, Valerie McCutcheon and Jacqui Brannan. 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.