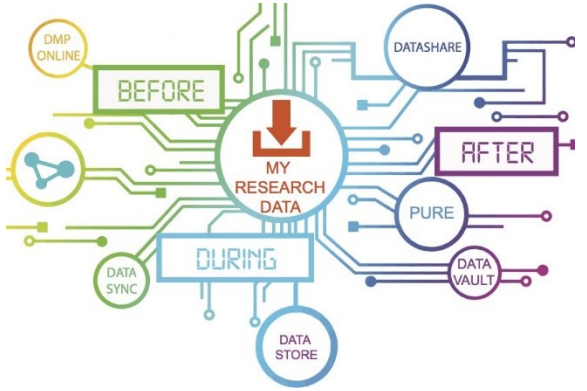
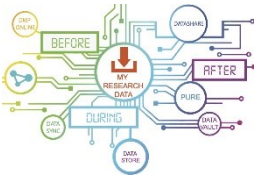


Writing a Data Management Plan

Simon Smith
Research Data Service

Overview



Data Management Planning

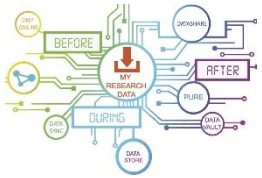
Writing a Data Management Plan (DMP)

Support for DMP & Research Data Management

Data Management Planning

Start as you mean to go on

What is data management planning?



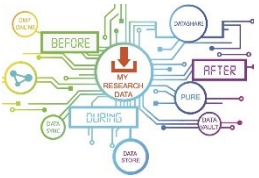
The act of planning how you will manage all aspects of your data before you begin collecting or creating it.

It may include:

- Where and how you will store and back-up your data;
- How and when it will be shared with collaborators;
- The steps required to anonymise sensitive data;
- When, how, and where your data will be preserved and shared.

Proper Preparation Prevents Poor Performance!

Why do I need a DMP?



To improve…

Efficiency & effectiveness;

Validation of results;

Use and/or re-use opportunities;

Visibility & impact;

The ability to share data.

To avoid…

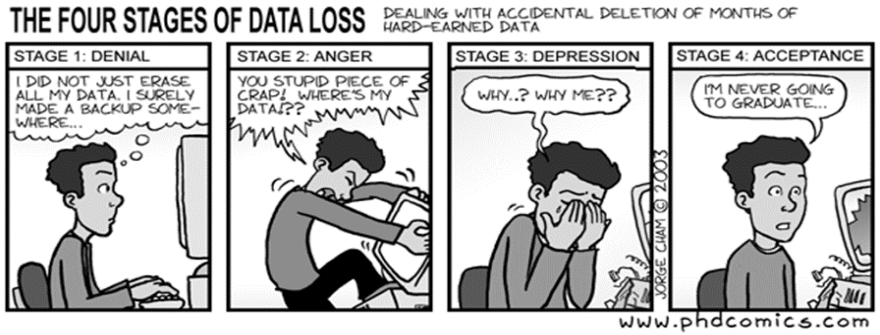
Hardware failure;

Software glitches;

Accidental deletion / corruption;

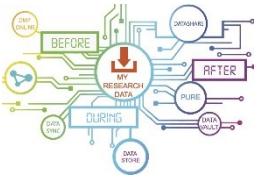
Media degradation;

Fire, flood, or natural disaster.



Jorge Cham ©
http://phdcomics.com/comics/archive.php?comicid=382

Essential Tips



- Keep your DMP short and specific;
- Seek advice from within your school / institute;
- Start early, don’t leave it till the last minute;
- Ensure that your ethics application, DPIA, & DMP are consistent;
- A DMP is a living document – keep it up-to-date;
- Ensure all team members have read, understood, and agree with the DMP.

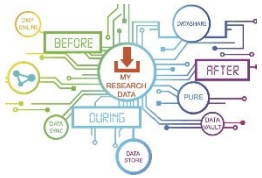
New Research Data Management policy

What does the policy mean for me?

- You must create a data management plan (DMP) for any new research proposal.
- You must include full costs for research data management in grant proposals.
- You should aim to link your datasets and research outputs such as published papers using persistent identifiers such as DOIs and an ORCID.
- Consider carefully who has rights to access the data; use a written agreement when necessary.
- You should not give exclusive rights to data to others, such as publishers.

https://blogs.ed.ac.uk/edinburgh_open_research/2022/01/13/new-research-data-management-policy/

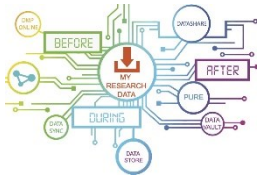
Funder requirements



Similar but different!

- Use DMPonline to find the most up-to-date version of your funder's template;
- Read them carefully – answer the questions as asked;
- Contact them for clarification if you are unsure;
- RDS can help you complete your plan to meet specific funder requirements;
- Ensure that any additional resources (money, time, equipment, staff) to deliver your DMP are included in any grant application.

DPIAs and ethics applications



A good DMP will already have addressed many of the questions that are included in ethics applications and Data Protection Impact Assessments (DPIAs).

You can often reuse answers from you DMP in these forms, saving you time and effort.

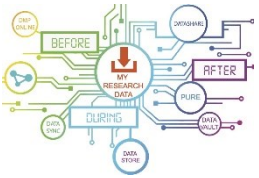
DPIA templates and guidance can be found at:

https://www.ed.ac.uk/data-protection/data-protection-impact-assessments

DMPonline

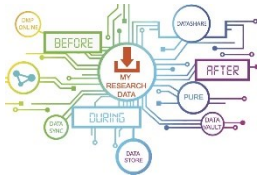
How to use

Live Demo!



https://dmponline.ed.ac.uk/

Your turn

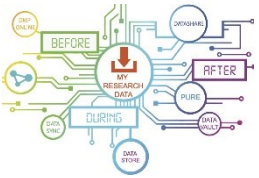


1. **Log-in to DMPonline**
2. **Select a University of Edinburgh template**
3. **Complete the administrative details on the first page**

Writing a Data Management Plan

To fail to plan is to plan to fail!

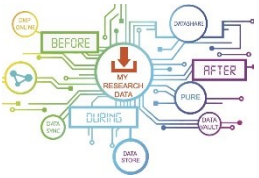
What should a DMP include?



A good DMP will answer all the following questions:

1. What data will be collected or created?
2. How will the data be documented and described?
3. How will ethics and intellectual property be handled?
4. How will the data be stored, backed-up & secured?
5. How, when, & where will data be preserved?
6. How will you achieve appropriate data sharing?
7. Who is responsible & what will it cost?

1. Data Description & Collection



Things to consider:

- Include code and “physical” data as well as digital;
- Is your digital data qualitative, quantitative, or a mixture;
- Consider the data formats you will use, wherever possible choose open formats;
- When using existing data outline any access requirements you’ll need to comply with;
- How many files will you create and what size will each one be? Can you make a reasonable estimate of the total amount of data you’ll collect?
- Don’t repeat anything already detailed in other parts of your documentation.

Standard file formats

Why?

Ensuring long-term usability of data requires consideration of the most appropriate software and file formats.

How?

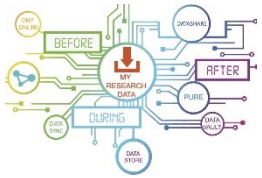
Using standard, interchangeable or open formats for long-term preservation and popular common formats to make data available to users.

Specific recommendations:

- Choose the best file formats | The University of Edinburgh https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats
- UK Data Service » Recommended formats https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages

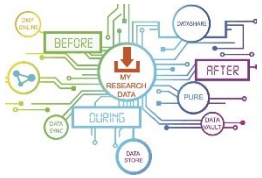
Writing your DMP – Section 1



Complete section 1 of the University of Edinburgh template

https://dmponline.ed.ac.uk

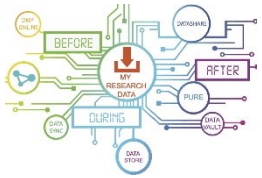
2. Documentation and Metadata



Things to consider:

- Where disciplinary or local standards already exist use them;
- Documenting your data well will also enable you to reuse it in the future;
- Do it as you go along – creating high quality documentation & metadata at the end of the project is very difficult and time consuming;
- Agree things like directory structures and file naming conventions with all members of a project team – be sure any new members are trained in these;

Types of documentation



Varies from project to project and may include:

- Data collection methodology
- Code annotations, details of libraries, hardware configurations, etc
- Data protection Impact Assessment (DPIA) / consent form
- Transcripts
- Questionnaires
- Laboratory notebooks
- Field notes
- Standard operating procedures
- Reports of decisions made that relate to conduct of the research

Types of metadata

•Categories of metadata

•Descriptive

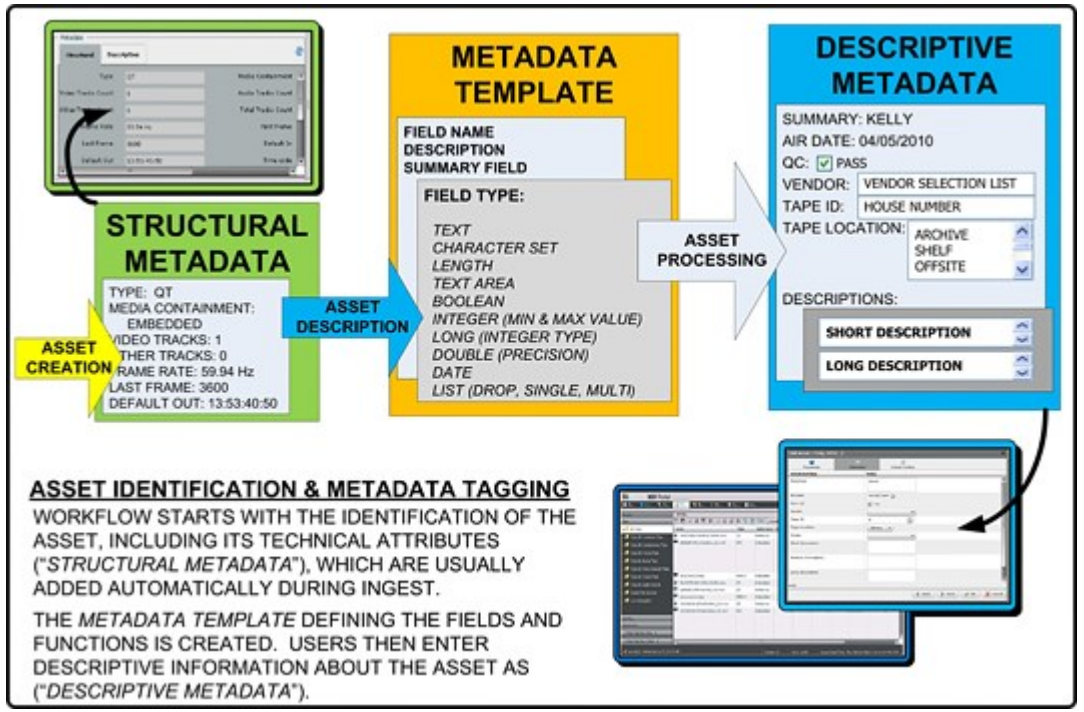
- Title
- Author
- abstract,
- location,
- keywords for discoverability

•Administrative

- terms of access
- rights management
- preservation

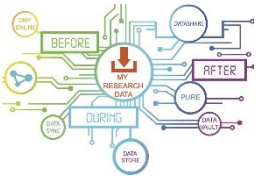
•Structural

- components of the dataset
- their relationship to each other



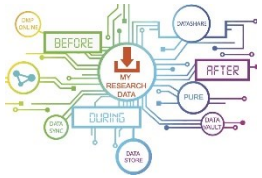
Acknowledgement: www.tvtechnology.com

Writing your DMP – Section 2



Continue in DMPonline and answer the question(s) in section 2

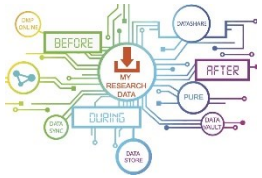
3. Ethics and Legal Compliance



Things to consider:

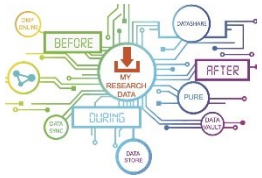
- Make explicit mention of consent, confidentiality, anonymisation and other ethical considerations, where appropriate.
- Demonstrate that you have sought advice on and addressed all copyright and rights management issues that apply to the data.
- Are any restrictions on data sharing required, for example to safeguard research participants or to gain appropriate intellectual property protection?
- Get advice on IPR from Edinburgh Innovations at an early stage.
- Also seek advice on ethics from School Ethics Committee & data protection from Data Protection.

What counts as sensitive data?



1. Data concerning human participants - This kind of sensitive data is often referred to as ‘personal data. Personal data identifies individuals, either directly (e.g. a name) or indirectly (when combined with other information to identify specific individuals).
2. Data relating to species of plants or animals - Data which includes information on rare or endangered species, or other conservation activities, is often classed as sensitive data.
3. Commercially sensitive data - Data where disclosure could cause economic harm, or prejudice the interests of any person, is deemed to be sensitive. This includes information such as references to ongoing negotiations, trade secrets, or data generated as part of a commercial funding agreement.
4. Data that poses a threat to others - Information which, if made available, would pose a threat to national security or would have a negative public impact.

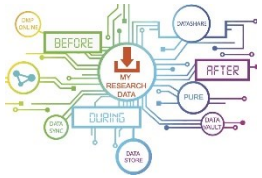
Working with sensitive data



Requires particular care to avoid:

- Causing direct or indirect harm to research participants
- Breaching confidentiality agreements or contractual arrangements
- Damaging the researchers reputation and that of the UoE
- Breaking the Law!

Before you start



Complete the mandatory UoE training:

“Data Protection Training” (Learn)

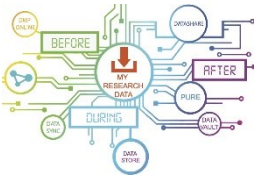
“Data Protection Training for Research” (Learn)

“Introduction to the Information Security Essentials” (Learn)

Other relevant training:

“Working with Personal and Sensitive Data” (workshop)

Guidance on data security



Informatics provides extensive guidance on:

- Encryption of devices and files
- Creating strong passwords
- Using VPNs
- Securing mobile devices
- Protecting sensitive data in different situations

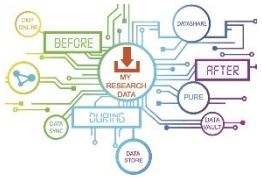
All available on the Informatics Intranet - https://computing.help.inf.ed.ac.uk/data-security

Information Security provides guidance on:

- University policy
- Password managers, inc. LastPass
- Secure deletion

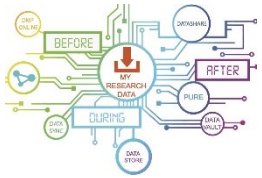
All this and more - https://www.ed.ac.uk/infosec

Writing your DMP – Section 3



Continue in DMPonline and answer the question(s) in section 3

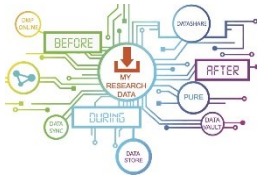
4. Storage and Backup During the Project



Things to consider:

- Can you use University networked storage for all your data collection and storage?
- Do you have enough storage space available for all of the data you will create or generate?
- If not, how will you ensure that your data is regularly synchronised between devices and properly backed-up?
- Will you need to collaborate with non-UoE partners, if so how will this be done in a secure and organised manner?
- Are any of your data sensitive? If so specialist storage space may need to be procured.

Informatics managed storage

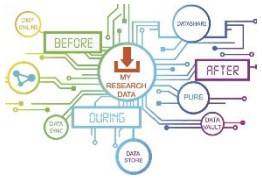


Informatics provides AFS for most file storage. It's authenticated with load balancing, local caching, and global access.

Full details of storage options available within Informatics are here: https://computing.help.inf.ed.ac.uk/storage

Choose the solution that will best meet your needs while providing adequate security, resilience, and back-up.

DataStore



DataStore provides active storage for all research staff and postgrad students.

RDM DataStore provides a free ‘at point of’ use allocation (currently 0.5TB).

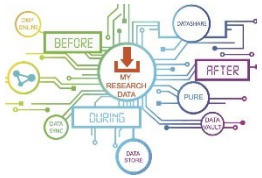
Additional capacity can be purchased for £175 per TB per annum.

Support for very large data (>1PB) hosting available.

This facility also provides a data services cloud for hosting specific data access mechanisms, or for integrating additional computational infrastructure.

Accessing DataStore: http://www.ed.ac.uk/information-services/computing/desktop-personal/network-shares/

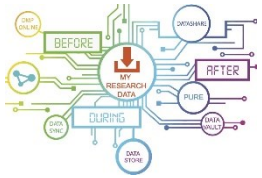
DataStore standard answer



"DataStore provides enterprise-class storage with guaranteed backup and resilience. Data is retained on DataStore until deletion by the data owner. The backups provide resilience in the case of accidental deletion and against incidents affecting the main DataStore storage. The data are automatically replicated to an off-site disaster recovery facility, with 10 days of file history visible online. Off-site tape backups keep 60 days of history of the filesystem. The 60 day rolling snapshots allow important data to be recovered to a prior state, by request if beyond the visible period.

Sensitive data stored on DataStore will be further protected by the use of 256 bit encryption as required by University policy"

OneDrive for Business



OneDrive For Business is cloud-based file storage for all staff and students as part of the Office 365 suite, allowing you to store and access personal and work files from anywhere.

Upload or sync any document from your local computer to OneDrive. It will then be available to you from any computer, tablet or phone.

Users receive 1 Terabyte (TB) of storage on OneDrive to store all personal and University documents.

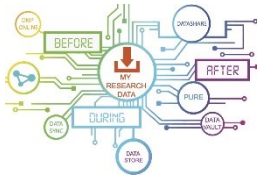
Any files created in Office Online are automatically backed up in your OneDrive space.

Share and work documents with friends and colleagues both inside and outside the University (latter may require Microsoft login).

Easily recover deleted documents or roll back to previous versions.

https://www.ed.ac.uk/information-services/computing/comms-and-collab/office365/onedrive-for-business

GitLab



Fully supported GitLab Community Edition (CE) environment for researchers.

GitLab is a version control tool which allow users to store and track changes to code and other documents.

Gitlab documentation is available at:

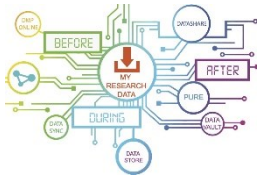
https://www.wiki.ed.ac.uk/display/ResearchServices/Version+Control+Service+-+GitLab

The GitLab service can be accessed at: https://git.ecdf.ed.ac.uk/

Informatics also provides a combined Git/Gerrit service

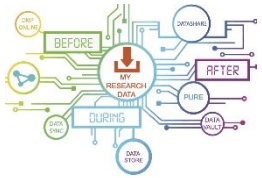
https://computing.help.inf.ed.ac.uk/git-and-gerrit-school-informatics

Writing your DMP –Section 4



Continue in DMPonline and answer the question(s) in section 4

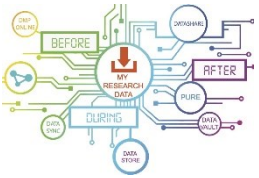
5. Data Selection and Preservation



Things to consider:

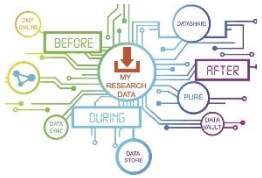
- Does all the data need to be preserved?
- Which data are necessary to validate existing research outputs?
- Where will you deposit your data for preservation? Use existing infrastructure wherever possible.
- Will any of the data need to be destroyed, e.g. to protect the privacy of participants or comply with confidentiality requirements of industry partners?
- What is the minimum period that the data should be preserved for?

How to preserve data



- Document data clearly and comprehensively
- Apply consistent quality assurance processes from the outset
- Choose file formats that are open, and accessible
- Select an appropriate repository or archive
- Deposit data in the chosen repository

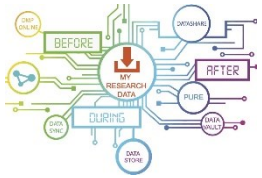
Comparison of archiving options



	DataStore	DataShare	DataVault
Editable?	Yes	No	No
DOI	No	Yes	Yes
Cost	£175 / TB / year* 500 GB free per research user	FREE	£50 / TB / year* Free for Projects under 100 GB
Size	Depends	x < 100 GB (per dataset)	A deposit may be up to 10 TB. Vault – no limit in theory. But > 100 TB would have to be scheduled for deposit to manage queues.
Access	University login aka EASE Shared area	Open	Roles: User roles relate to a specific vault or all vaults belonging to a School.
Embargo	N/A	Temporary	N/A

*** PRICES CORRECT AS AT 17 January 2022.**

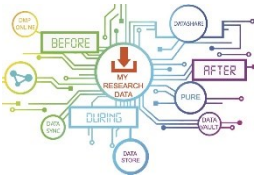
6. Data Sharing



Things to consider:

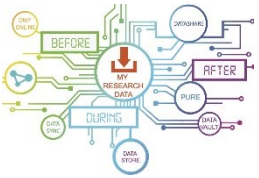
- Can all your data be made openly available, or will some of it require access controls?
- What will you do to promote the findability of your data?
- How will you obtain a persistent identifier for your data?
- Will your chosen repository record data downloads or citations?
- Do you want an embargo on your data to delay its release?

How to make data shareable



- Plan for sharing (via a data management plan) at an early stage.
- Don’t collect personal information that is not needed.
- Principle of informed consent: get consent to share data.
- Attribute, anonymise, or aggregate individual’s data.
- Document all data processing (inside & outside analysis package).
- Consider the timescale for release (embargo).
- Check the infrastructure for sharing.
- Be aware of rights management & licensing

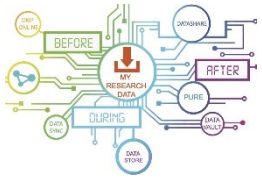
Choosing a repository



When choosing a repository consider:

- Is the repository sustainable?
- What will happen if the repository closes down?
- How much will it cost? Are costs upfront or annual?
- Will outputs be easily accessible to the researcher and to third parties?
- How does the repository promote discoverability?
- Does the repository record when outputs are accessed, downloaded, or cited so the researcher will get recognition for their work?

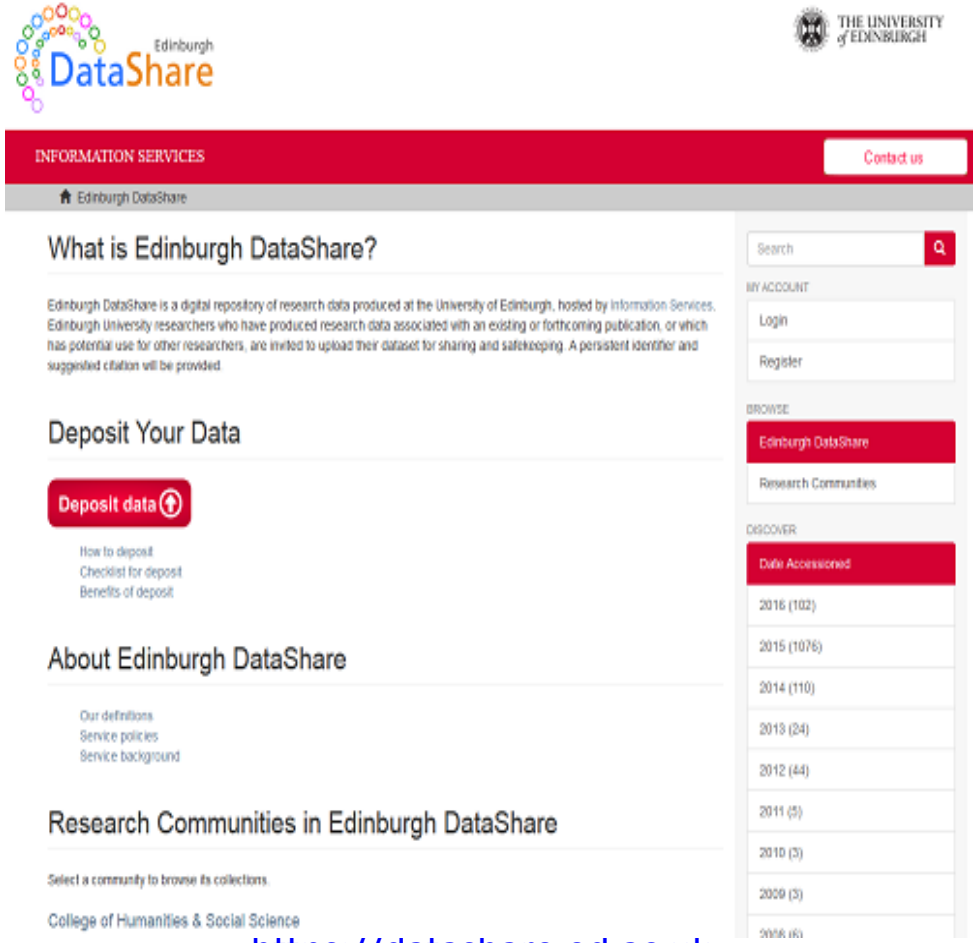
Edinburgh DataShare



Edinburgh DataShare is the University's OA multi-disciplinary data repository hosted by the Data Library : https://datashare.ed.ac.uk

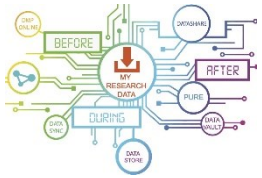
Assists researchers who want to share their data, get credit for data publication, and preserve their data for the long-term (DOI, licence, citation)

It can help researchers comply with funder requirements to preserve and share your data and complies with Edinburgh's RDM Policy



https://datashare.ed.ac.uk

DataVault

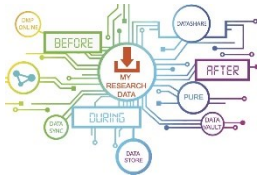


Safe, private, store of data that is only accessible to the data creator or her/his representative.

- provides data creators with a safe place to store sensitive research data that are no longer actively being developed and cannot be published;
- data will receive Digital Object Identifier(s) (DOI), which can be used in publications and other outputs;
- complies with funder and University requirements to preserve research data for the long-term;
- be confident that the data will be available for re-use in the future; and
- links research data to metadata records in PURE.
- costs £500 for 1TB over 10 years. Storage for data up to 100GB is free.

http://www.ed.ac.uk/is/research-support/datavault

Other Repositories Are Available



zenodo

Search

Upload Communities

Log in Sign up

Featured communities

Need help uploading? Contact us

Transform to Open Science

Browse New upload

Transform to OPen Science (TOPS) is a $40 million, 5-year mission, led by NASA's Science Mission Directorate's Open-Source Science initiative. Within the TOPS mission, NASA is designating 2023 as the Year Of Open Science, a community initiative to spark change and inspire open science...

Curated by: nasatransformtoopen

Recent uploads

April 19, 2023 (v1.18.4) Software Open Access

Flowminder/FlowKit: 1.18.4

Jonathan Gray; maxalbert; James Harrison; Thingus; dependabot-support; Bhavin Panchal; Dan Williams; OwlHute;

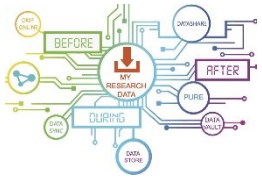
View

Need help?

Contact us

Zenodo prioritizes all requested related to the COVID-19 outbreak.

External repositories

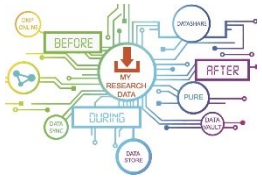


An external repository may be more suitable if:

- The funder requires or suggests using a particular repository;
- If there is an established national or international repository for the discipline;
- You need to preserve and share code;
- If it is necessary to implement access controls to protect the output(s).

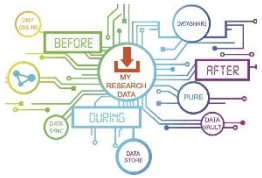
An international register of research data repositories is maintained at www.re3data.org as part of Datacite.

Writing your DMP – Sections 5 & 6



Continue in DMPonline and answer the question(s) in sections 5 & 6

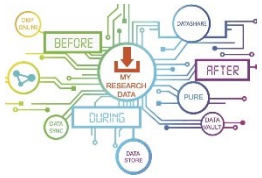
7. Responsibilities & Resources



Examples from funders:

- Responsibilities & Resources: persons involved in collection/management of data and implementing plan; training required; hardware/software requirements; charges for data repositories; costs and resources for preparation for preservation/data sharing. (UoE Template)
- You should consider what resources you may need to deliver your plan and outline where dedicated resources are required. (Wellcome Trust)
- Who will be responsible for data management? What resources will you require to deliver your plan? (EPSRC)
- Apart from the PI, who is responsible at your organisation/within your consortia for: study-wide data management, metadata creation, data security, quality assurance of data. (UKRI)

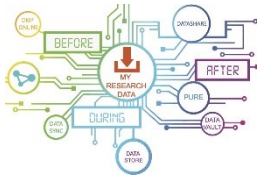
7. Who is responsible & what will it cost?



Things to consider:

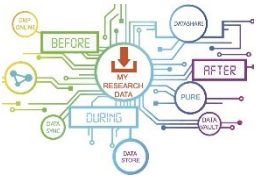
- Outline and justify costs: what resources will you require to deliver your plan.
- Include cost estimates for both active data storage and long-term preservation & storage.
- Be realistic about the human time and effort required, e.g. for data cleaning or anonymisation.
- Show that funds will be used efficiently and effectively.
- Some funders require named individuals to take responsibility for RDM tasks – think about the best people on a team to do specific tasks
- UKDA provides costing guidance: https://www.ukdataservice.ac.uk/manage-data/plan/costing

Writing your DMP – Section 7



Continue in DMPonline and answer the question(s) in section 7

What Happens Now?



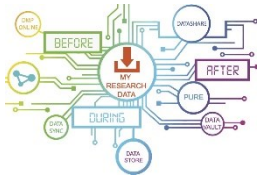
Get Your DMP Reviewed:

1. In Pairs/Small Groups: Share and Review your DMPs
2. TGIDRGS: Than God It's DMP Review Group Session
3. Make Your Changes
4. Surprise Your Supervisor: send them your DMP

Reviewing a DMP: What am I looking for?

- Is the plan complete?
- Could I be Data Manager for this project?
- Use the guidance down the right-hand side!

So What's The Plan?



Look out for:

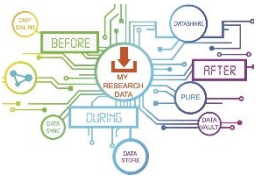
Section: 1. Description of the data

- Code/software
- Personal/Sensitive data
- Format/Size

Section 2. Data collection / generation

- Quality Assurance: how will you know that the code is doing what it's supposed to do?

So What's The Plan?



Keep looking out for:

Section 3: Data management, documentation and curation

- Storage: **NOT** flash drives, laptop or pc hard drive, other portable media
- Preservation Strategy: **NOT** DataStore, flash drives, laptops, etc.

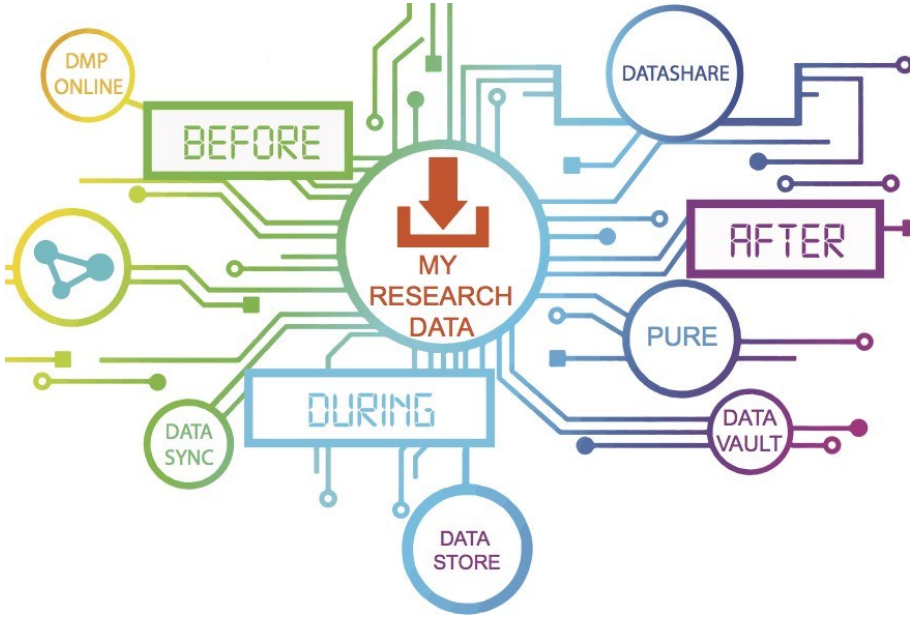
Section 4. Data security and confidentiality of potentially disclosive information

- Personal/sensitive data
- Data Security Risks

Section 5: Data / Code sharing and access

- Discovery: DOI, licence, metadata, DAS

Questions?



General RDM queries & requests for help writing a DMP should be sent to data-support@ed.ac.uk

RDM website: http://www.ed.ac.uk/is/research-data-service

Training Courses: www.ed.ac.uk/is/data-training

RDM blog: http://datablog.is.ed.ac.uk