UoB Research Data Management Plan Template Guidance and examples

(v7.1) January 2025

This document contains Guidance and Examples for the **University of Birmingham DMP template**

# 1. Overview

1. Researcher’s Name (Principal investigator):
2. Title and ID of Research Project:
3. Length of Project
4. Start Date:
5. End Date:
6. A brief statement of the aim(s) of the project:
7. Funding body and programme:
8. Related policies:
9. Version of DMP:

# 2. Data collection and documentation

## a. Briefly describe the data that you will collect for the project

**Help, Guidance**

* Details on the type of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media. Details of existing datasets to be re-used. Methods by which data will be collected or created. In the event of interview recordings you must name any platform that will be used (e.g. MTeams, Zoom). Any specific data protection measures applied during initial data collection or processing should also be included.

**Example Answers**

* I will conduct structured online interviews using Microsoft Teams, following the GDPR principles to protect participants personal data. Once the automated transcript has been checked and downloaded, the MS Teams recording, alongside personal details recorded in the MS Teams calendar, will be deleted. Transcripts of each interview will be stored as individual MS Word files.
* I generate data using model code that I’ve written. The code outputs data as a csv file. Those files are then process in various ways to produce visualisations.
* I combine existing data from a number of sources [e.g.…] and reanalyse them to derive new conclusions.

## b. Frequency of new data (how often will you get new data and over what time period?)

**Example Answers**

* All of my data will come from a single 3-month field trip in my second year.
* I expect to run two or three experiments each week through my second year and much of my third year – about 100 in total.

## c. Quantity of data (Terabytes, other forms of storage)

**Guidance**

* Details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns)

**Example answers**

* Each experiment produces about 50MB of data, so over the course of my PhD I expect this to add up to about 5GB.
* 2 drawers of a standard filing cabinet

## d. What format is the data in?

**Guidance**

* Details on the data format: how the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).
* Explain why certain formats have been chosen and indicate if they are in open and standard format. Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such ‘preferred formats’). If a proprietary format is used, it explains why.
* Clearly state, if applicable, that no new data will be produced or generated by the project.

**Example Answers**

* The format used by a particular instrument paper notebook.
* Audio recordings from interviews will be stored in MP3 format.
* Survey data will be stored in SPSS format.

## e. Describe the system to name and structure any electronic files and records.

**Guidance**

* UK Data Service guidance on organising files: [Organising — UK Data Service](https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/organising/)

**Example answers**

* I will use the structure <thesis chapter>/<date>-<experiment number>.
* A folder for each project phase and a folder for each interview.
* Each filename starts with the date on which the data was collected
* Interview transcripts will be stored as separate files and all 20 files will be placed in a single folder.

## f. Could the data be considered high value and/or vulnerable? E.g. is your data likely to attract “hactivists”? How could this be mitigated? What measures will you take to comply with the security requirements and to mitigate the risks?

**Guidance**

* Describe how you can restore your data in the event of data loss and who is responsible. If applicable, please describe procedures to ensure personal data are handled confidentially and who is responsible.
* Precaution examples: Access restrictions (physical or digital), Encryption,/ reduce data sensitivity, Regular and timely back-ups, Master (locked) copy stored on university network storage, Master (locked) copy stored elsewhere, Data handling procedures and/or training for data handlers, name any other.
* Information security policies, standards and guidance resources are available at: https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx

**Example Answers**

* The working data will be stored on University of Birmingham (UoB) storage (BEAR Research Data Store), accessed via network shares mounted on the computers of the Principal Investigator (PI), Co-PI’s and other researchers. Backups are made overnight from the Research Data Store (RDS) and any files that are created or changed that day will be backed up. Backups are also copied to a second location for disaster recovery purposes .

## g. What standard will you use to describe your data?

**Guidance**

* Please refer to any metadata standards in your field if they exist. And Where will metadata be registered?

**Example Answers**

* I will use a discipline-specific metadata standard, namely DDI (Social Sciences)/Darwin Core (Natural history)
* I will make use of the device’s automatically generated metadata

## h. What data quality control measures will be used?

**Guidance**

* Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.

**Example answers**

* I will describe in a text document what is in a file, where it comes from, how it could be retrieved if needed, any existing problems, etc.I will document and record the data collection workflows

## i. What additional information is necessary to understand the data? E.g. abbreviations and supplementary notes.

**Example answers**

* Abbreviations used for column headings are kept in a separate text document.
* The content of digital photographs is recorded in the file name.
* Database schema
* Laboratory notebooks
* Software syntax and output files
* Any jargon language elements will be recorded in a separate file

# 3. Data Storage and Backup

## a. What different versions of the data do you create? E.g. versions of data files

**Help, Guidance**

* Version your files, e.g. by using a 'revision' numbering system. Any significant changes to a file can be indicated by numbers; for example, v01 would be the first version, v02 the second version
* Consider using version control software such as Git, which can be used on text files and code. There is a Library Carpentry session on Introduction to Git: https://librarycarpentry.org/lc-git/
* More details on Version control and authenticity from the UK Data Service.

**Example Answers**

* As I survey new cohorts, data is appended to the dataset and saved as a new file.
* There is only one version of each data file — new experiments create new data, which is stored in a new set of files.
* Each time I run a new version of my model, intermediate files are written over, but the final results are saved as a new file

## b. Where will the data be stored?

**Help, Guidance**

* Describe where the data will be backed up during research activities and how often the backup will be performed. How will data and metadata be stored and backed up during the research? Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution (i.e. UoB). Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.
* It is recommended that data be stored in at least two separate locations.
* Electronic data should be stored in multiple geographically distinct locations and any sensitive data on local removable drives should be encrypted. Note that the University Research Data Store is backed up to two physically separate data centres.
* We recommend reading the article‘Where should I store my research data’ for examples: <https://kb.bham.ac.uk/KB15356>

**Example answers**

* Working data is backed up on the UoB Research Data Store. I copy the latest versions of my working there there each day. I regularly scan my paper-notebook and store digital copies on the X: drive.
* Information security policies, standards and guidance resources are available at: https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx

## c. Describe the procedure to be used to ensure files can be restored from the backups.

**Help, Guidance**

* Backing up data is an essential practice to insure against the loss of valuable information. Check the information at the UoB [Backup & Retention Policy](https://www.birmingham.ac.uk/research/arc/policies/rds-backup-retention-policy)

**Example answers**

* Weekly checks of the files on the X: drive are still usable. The data will be stored on University of Birmingham (UoB) storage (BEAR Research Data Store), accessed via network shares mounted on the University-managed computers of the Principal Investigator (PI), Co-PI’s and other researchers. Backups are automatically made overnight from the Research Data Store (RDS) and any files that are created or changed that day will be backed up, with backup tapes copied to a second location for disaster recovery purposes. Backups can be recovered either via accessing snapshots (within 4 days of deletion) or by contacting IT Services (within 90 days of deletion).
* Information security policies, standards and guidance resources are available at: https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx

## d. Will the project generate any non-digital data or outputs? Where will these outputs be stored?

**Help, Guidance**

* Do you have a protocol for storage and deletion of non-digital data? Please specify briefly and describe who is responsible for the storage of these outputs.

**Example answers**

* Signed consent forms will be stored in a locked cabinet in the office and will be destroyed after the completion of the data analysis phase
* Physical lab notebooks will be securely stored in a locked cabinet in the research office.

## e. How will data security and protection of sensitive data be taken care of during the research?

**Help, Guidance**

* Which institutional and/or national data protection policies are in place - provide a link to where they can be accessed, who will have access to the data during the research, and how access to data is controlled, especially in collaborative partnerships.
* A Data Protection Impact Assessment (DPIA) must be carried out by the relevant School or Professional Service and approved by the Data Protection Officer (or nominee). This must be retained and regularly reviewed and revised as necessary for the duration of the processing. For more information follow the link here: <https://www.birmingham.ac.uk/documents/university/legal/data-prot-policy.pdf>
* Clearly describe the additional security measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe, when sensitive data are involved (for example, personal data, politically sensitive information, or trade secrets). Information security policies, standards and guidance resources are available at: <https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx>
* BEAR Services [The Data Matrix](https://itservicedesk.bham.ac.uk/sys_attachment.do?sys_id=a0b5acf91b3eb5147bc4a609b04bcb12&view=true) provides guidance on how to handle sensitive data

## f. What are the main risks to data security, and what would happen if the data got lost or unusable?

**Help, Guidance**

* Examples of risks: Accidental deletion or file corruption, Theft of, or damage to, equipment, Overwriting or version loss, Data leak, unauthorised access, or unauthorised use, or explain any other event that might occur

**Example Answers**

* Original versions of files will always be kept on the server. If copies of files are held on a laptop and edits made, their file name will be changed.
* Information security policies, standards and guidance resources are available at: <https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx>

# 4. Ethics and legal compliance

## Who owns the data? (Copyright and intellectual property rights)

**Help, Guidance**

* Explain who will be the owner of the data, meaning who will have the rights to control access
* Make sure to cover rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement. to understand data ownership, consult the University’s Code of Practice for Research (subject to change January 2025), and/or discuss with your Principal Investigator. Also, check your agreements or contracts with any sponsors or collaborators.

**Example Answers**

* I own the copyright of the newly created research data.
* The media information is in the copyright of the respective publishers and will be attributed accordingly

## b. Are there restrictions on who can use the data, and if so, what are they?

**Help, Guidance**

* Explain what access conditions will apply to the data. Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses. More information: Licences.
* Clearly explain, if applicable, why data sharing is limited or not possible and who can access the data under which conditions (for example, only members of certain communities or via a sharing agreement).
* Describe the procedure to manage access to only authorised users.

**Example Answers**

* “A confidentiality agreement covers all my data and, for that, cannot be shared.”
* “Some of my data identifies individual patients and must be anonymised before sharing.”
* "Personally identifying information (PII) will be removed and stored separately from the data files. Access to these separately store PII files will have the added protection of encryption, such as via a password."

## c. If personal sensitive data are processed, how will compliance with legislation on personal data and security be ensured?

**Help, Guidance**

* Ensure that when dealing with personal data, data protection laws (for example, GDPR) are complied with:
* Gain informed consent for preservation and/or sharing of personal data.
* Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data). Check the following tool: https://amnesia.openaire.eu/
* And the guidance on storing sensitive data from UoB IT Services:

**Example Answer**

* ”To ensure anonymity, all personal identifiers will be removed from the dataset before sharing”.

# 5. Archiving (preserving the data for future use)

## What data should be kept beyond the end of the project and under what requirements (contractual, legal or regulatory)?

**Help, Guidance**

* Explain the foreseeable research uses (and/ or users) for the data.

**Example Answers**

* All data, both raw and processed.
* Only simulation code and input parameters.
* Anonymised transcripts of all interviews, but not recordings.
* Data underlying publications will be shared openly via a trusted repository

## b. How long should it be kept?

**Help, Guidance**

* This should align with funder, institutional, national policies and/or legislation, or community standards.
* Data will be preserved in the long term and clearly indicate for how long.

**Example answers**

* Until the end of my PhD
* For 10 years after the end of the project

## c. Where will the data be archived?

**Help, Guidance**

* Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.

**Example answers**

* In the UK Data Service ReShare repository
* On the Ubira eData repository￼

## d. Who will create and maintain the archival of data?

**Help, Guidance**

* Who (for example, role, position, and institution) will be responsible for data management (i.e. the data steward)?

**Example answers**

* I am responsible for initial archiving of data in the University’s edata repository. The repository is managed long-term by the University Library and Learning Resources department.

**Are there restrictions on who can access the archived data?**

**Example answers**

* Data will be embargoed for 12 months to enable patient protection.
* Full data will never be public as it contains sensitive personal information, but anonymised data will be available upon request.

## e. If the data can be made openly available, at what point can this happen?

**Help, Guidance**

* How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
* Explain how the data will be discoverable and shared (for example, by depositing it in a trustworthy data repository, indexed in a catalogue, using a secure data service, direct handling of data requests, or using another mechanism). Include any licence under which the data will be shared (e.g. CC-BY for maximum capacity for re-use)
* Outline the plan for data preservation and give information on how long the data will be retained.
* Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents.

**Example answers**

* When publishing research papers from the project, relevant data sets will be made openly accessible under a CC BY licence via the University’s edata repository and linked from publications via a data access statement
* At the completion of the project, data will be archived in a trusted repository under a one year embargo to enable final publications to be made.

## f. How will you enable findability and availability for the long term and maximise data’s discoverability?

**Help, Guidance**

* Identify where the data will be deposited, how the metadata will be made available and whether it will be given a persistent identifier such as a DOI.
* For the sharing of data with the general public, experimental data that underpins published work will be deposited in UoB Ubira eData repository. UoB uses PURE as its research information and management system; a metadata record will be created in PURE for the dataset. PURE exposes metadata to web services via standard metadata harvesting protocols.

**Example Answers**

* My data will be deposited in the UBIRA edata repository (based on Eprints) and be given a DOI to aid long term discoverability
* I will deposit data in a trusted data repository (e.g. DANS Easy, 4TU.ResearchData) as indicated below:
* According to the data protocol of my institute, I will archive data in the data repository indicated below (e.g.Ubira eData):
* I will deposit data in a discipline-specific data repository as indicated below:
* I will use an archive specifically for my collaboration, namely:
* I will not use a data repository and will explain below how I will make my data findable and accessible for the long term.
* I will not make my data findable and accessible and I will explain why

## g. What are the likely (estimated) costs of preserving the data?

**Help, Guidance**

* What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?
* Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in.
* Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges.
* Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.
* Consider the likely amount of storage you will need, and how long this will need to be archived for. The University of Birmingham provides the BEAR Archive for storing valuable data associated with your research project which may be needed in the future but is not currently being actively used, and is not needed to support a project. Up to 20TB of data storage is provided free of charge for up to 10 years. For more information see <https://www.birmingham.ac.uk/bear-archive>.

**Example Answer**

* Open and working data will be stored in UoB storage and sharing facilities that are offered free of charge for its researchers.

# 6. Implementing this Plan

## a. Name the person responsible for implementing this plan.

**Example Answer**

* I will take responsibility for carrying out the actions required by this plan and report them to my supervisor as appropriate.

## b. Frequency of review and/or updates of this plan.

**Example Answer**

* My supervisor and I will review it every six months and update its contents if necessary

## c. Actions required to implement this plan.

**Example Answers**

* Ask my supervisor to request research storage space for my project
* Set up a backup system.
* A test dataset II can be restored from my backup.
* Learn how to anonymise data for archival.

## d. List any further information needed to carry out the actions above.

**Help, Guidance**

* Where can you find this information?
* Who might you be able to ask?
* e.g. DCC, DMPonline research-data@contacts@bham.ac.uk
* https://www.birmingham.ac.uk/research/environment-culture/open-research-statement

### Document created: January 2025

**Useful resources:**

* <https://www.birmingham.ac.uk/research/environment-culture/open-research-statement>
* <https://intranet.birmingham.ac.uk/student/libraries/research/rdm/fair-data.aspx>
* [FAIR Data Self-Assessment Tool | ARDC](https://ardc.edu.au/resource/fair-data-self-assessment-tool/)
* <https://reshare.ukdataservice.ac.uk/>