Euro-BioImaging's Template for Research Data Management Plans

Version: V 1.0

Good documentation of data management practices is key for the effectiveness and sustainability of every research project. The purpose of a Data Management Plan (DMP) is to ensure digital traceability of project data which allows for a continuous and unambiguous flow of information during the research project - from acquisition to publication. Thereby a DMP supports the optimized use of resources through better planning. Further a DMP facilitates the practical implementation of open science policies and promotes the transparency, verifiability and reproducibility of a research project.

This DMP addresses the following topics:

- General project description and definition of roles and responsibilities (Section 1/ Section 2)
- What type(s) of data will be used and how it will be collected (Section 3)
- How data and metadata will be documented (Section 4)
- How data will be stored and backed up during research and preserved beyond projects end (Section 5)
- Perspectives of sharing or providing access to project outputs and materials (Section 6)
- How any ethical and legal issues will be dealt with (Section 7)

How to use this document:

- This template is a set of questions that you should answer at a level of detail appropriate to the project.
- Fill in this document and keep it with your data unless instructed otherwise.
- Update this document regularly as changes from the original plan occur or as plans become more concrete.
- Review the items carefully and discuss open questions with your supervisor, institution or data steward.
- Hover your mouse over the fillable fields marked with ⓘ for further instructions and example answers.

Note that if your Euro-BioImaging project is funded by the European Union's Horizon Europe research and innovation programme (such as ISIDORe, grant agreement number 101046133), it is mandatory to make your research outputs traceable and openly accessible on common data platforms wherever possible. These requirements to produce data in accordance with the FAIR principles and the legal provisions of the General Data Protection Regulation (GDPR) can be ensured by conscientiously completing this DMP and following its recommendations.

This document was created by Isabel Kemmer (FAIR Image Data Steward, Euro-BioImaging) with input from Aastha Mathur (Coordinator Image Data Services, Euro-BioImaging) and Arina Rybina (Scientific Project Manager, Euro-BioImaging).

The content is based on the data management checklist by Jean-Karim Heriche (EMBL Heidelberg), EMBL data management plan template, Horizon Europe data management plan template, the NC State University Libraby 'elements of a data management plan (https://www.lib.ncsu.edu/do/data-management/elements-of-a-dmp)' and the WUR DMP Template and guidance (10.5281/zenodo.7233369).

For questions or feedback please contact: fairdata@eurobioimaging.eu

1-General Project Description

Project Name	
Project ID	
Project Leader	
Project description	
Project Start Date	
Project End Date	
Funding	

History Information

Current Date	
Version #	
Version # description	

2- Roles and Responsibilities

Wherever possible name individuals which are responsible for the following aspects of the data lifecycle and indicate their roles and affiliations or name an institution. Caution: be mindful about storing personal data and when in doubt do not store the contact details here.

Data owner	
Data collection	
Data analysis	
Data organization/storage	
Long-term preservation	
Data stewardship	
Other roles	

3- Research Outputs

Outputs (i.e. "Data") encompasses general research data (i.e. microscopy images, cell traces, value tables). But also consider other types of digital research output such as produced analysis software or development of new procedures and workflows. Further, also include physical research outputs (i.e. antibodies, cell lines, plasmids, reagents etc.) and laboratory protocols.

3.1 – Research Data

Will you re-use existing data? ☐ yes

Describe the data that will be re-used and for which purpose. Specify the terms and conditions under which you are allowed to use them:

☐ no

Explain why data-use is not an option in your project:

Will you generate new data? ☐ yes

Consider both raw data as well as analysed data.

Describe **which** data will be generated and **how**:

☐ no

Estimate the total size of data you work with:

Describe the types of data that will be generated and used, as well as their formats/standards.

You are encouraged to use open formats instead of proprietary formats wherever possible to increase accessibility, interoperability and re-usability of the data. While many (proprietary) image formats can be read by the Bio-Formats library (plug-in for Fiji available), please check if all data, metadata and analysis layers are correctly read and stored.

Data type	File format

Will you (re)use existing software to generate, process and analyse data?

☐ no

☐ yes, please fill the table below

Software ⓘ			
Version			
Purpose ⓘ			
License ⓘ			
Obtainable from ⓘ			

3.2 – Computational workflows

Will you generate image acquisition or analysis workflows?

☐ no please fill Section 4.3 instead

☐ yes, but only for certain steps please fill the table below and Section 4.3 additionally

☐ yes for everything please fill the table below

Workflow ⓘ			
Purpose ⓘ			
Author(s)			
Generation tool ⓘ			

Please also cover Section 6.3 to describe your plans of making the workflow available.

3.3 – Research materials, equipment and procedures

Will you generate new materials, reagents or equipment?

☐ no

☐ yes, please describe

Will you generate new experimental procedures?

☐ no

☐ yes, please describe

4 – Data Documentation and Metadata

Metadata is data that describes other data and enables discovery, interpretation, and reuse of the original data. It includes information such as data type, origin, organization, and relationships to other data. The metadata can be embedded in the image file or stored in a separate file which details the experimental procedure, image acquisition settings, and data analysis performed to obtain the results.

4.1 – Research data organization and structure

How do you organize your data during data generation and further analysis?

☐ file-system based

Explain the directory structure and the file/folder naming scheme:

If needed consult the RDM best practice – file and folder naming.

☐ other

Please elaborate:

How will changes in the data organization be documented?

4.2 - Data interpretation, metadata and reproducibility

Explain which types of metadata will be associated with the data and which standards they follow:

Please also consult the RDM best practices – metadata section.

Where is the metadata stored and how will the data and metadata be linked? How will the link between data and metadata be preserved as the data organization changes?

How are (biological) entities (e.g. strains, materials, equipment, methods ...) referenced using controlled vocabularies/ontologies?

Describe the system used for file versioning:

4.3 – Management of computational processes

Will a computational workflow management system be used?

☐ yes please fill Section 3.3

☐ no

Explain how you manage and document your computational workflows instead. How will software configuration parameters and manual data processing interventions be documented?

4.4 – Management of research procedures and materials

Will an experimental workflow management system be used?

☐ yes

Which management system/electronic lab notebook is used:

☐ no

Explain how you document your experimental workflows instead:

Which metadata do you store for research material? How is the research material managed?

5- Data storage and long-term preservation

5.1 – Data storage during the project's lifetime

Ideally before collecting data, check with the facility about their data storage, access and transfer policies. Consider the expected size of each dataset and the total data generated to decide which data storage and transfer protocols are suitable for your data (in terms of local disk space, cloud storage requirements and data transfer speed).

Initial data storage location

Data storage lifecycle

Describe where the data will be stored during the project's lifetime and how it will be moved.

Data access pattern

Who is going to access the data during the project's lifetime and for which purpose?
Explain the requirements this puts on your data storage (bandwidth, permissions…) and if they are met.

Data back-up strategy

Describe which data is backed-up and detail the back-up process.

5.2 – Long-term storage

Check if your institution has a policy on long-term storage (usual time 10 years). Also, many grant funders require a minimum data retention period so be certain to comply with these.

Data selection	How will you select data for preservation? Explain which data must be retained or destroyed for contractual, legal or regulatory purposes.
Minimal retention period	
Place of long-term storage	
Data preservation strategy	Briefly describe the procedures your long-term storage facility has in place for preservation and back-up, if known.

6 - Sharing of research outputs

Following the FAIR (Findable, Accessible, Interoperable, Reusable) principles, you are strongly encouraged to make your data (& workflows, code) publicly available wherever possible especially for data underlying a publication. Also, more and more funders and journals now require data to be openly available wherever possible, usually at the time of the article's publication.

6.1 – Journal publication

Are you integrating the data into a journal publication?

☐ no why not? (skip the rest of section 6.1)

☐ yes

Which data will be included in the publication and which not? Describe your selection process:

Will a pre-print be available? ☐ yes

Please specify where:

☐ no

Describe the journal and its requirements for data publication:

Journal

Publishing license

☐ open-access Further specify the license:

ⓘ

☐ subscription-based model (closed-access)

Explain why you chose and specify the license:

ⓘ

☐ other please specify:

Specific requirements ⓘ

6.2 – Data sharing

Which datasets will be shared?

☐ all data produced

☐ data used in publication

☐ certain datasets (describe your selection criteria)

☐ only the metadata (give reasons)

☐ none (give reasons)

When will your data be shared?

- ☐ as soon as it is generated
- ☐ together with paper publication
- ☐ upon project completion
- ☐ after an embargo
- ☐ other (please elaborate)

Where will the data be shared?

- ☐ domain-specific, open repository
- ☐ domain-specific, restricted-access repository
- ☐ institutional repository
- ☐ general purpose repository
- ☐ upon request/publication only
- ☐ other, please specify

Describe the standards that the repository requires:

Under which **licence** will the data be shared?

- ☐ Fully open (CC0)
- ☐ Attribution (e.g. CC BY)
- ☐ Non-derivative (e.g. CC BY-ND)
- ☐ Non-commercial (e.g. CC BY-NC)
- ☐ other (please specify)

How can the data be accessed?

How long will the data remain findable in this repository?

6.3 – Distribution of computational workflows

Are the computational workflows and software made available?

☐ yes
Describe which ones, when and in which way:

Specify the licence:

Describe which documentation will be made available with the software:

☐ no Describe why not:

6.4 – Distribution of research procedures and materials

Will you share developed experimental protocols?

☐ yes Describe in which way:

☐ no Describe why not:

Will you share physical research outputs?

☐ yes Describe in which way:

ⓘ

☐ no Describe why not:

7- Sensitive data, intellectual property and ethics

Will you store or process sensitive or personal data?

For example patient-related data, health records etc.

ⓘ

☐ yes

Describe how the data is handled according to legal requirements:

☐ no

Are there any issues that will restrict access to the data or impact data sharing?

ⓘ

☐ yes, please describe

☐ no

Are there any other ethical considerations?

ⓘ

☐ yes, please explain

☐ no