Guidelines: EUR Data Management Plan

Version 4.6

This guideline document was created by the RDM specialists at UL/EDSC and the data stewards. It will be updated and expanded in the future.

Table of contents

The guideline document is meant to accompany the EUR DMP Template 4.6, which is available on the web page of Erasmus Research Services. You can also find these guidelines on DMPonline, where the guidance is already included next to each question.

GENERAL

For the contact information of the support staff than can help you with writing a data management plan (DMP), please refer to the following link:

- https://www.eur.nl/en/research/research-services/research-data-management/data-managementplan

Code of Conducts and guidelines:

You can find *The European Code of Conduct for Research Integrity* (ALLEA, 2023) here. To find out more, follow this link:

- https://allea.org/code-of-conduct/

You can find *The Netherlands Code of Conduct for Research Integrity* (VSNU, 2018) here.

ADMINISTRATION & PROJECT DESCRIPTION

DMP versioning

Researchers are advised to update their DMP once a year. List here the version number of your updated DMP (e.g. version 2, 3, 4).

Data Management Costs

For a checklist with examples of costs (financial and time) that might be needed to implement data management in your project, see this ERS webpage about Data Management Costs.

PREPARATION: LEGAL ARRANGEMENTS AND POLICY

Legal arrangements

If you (think you) need to make legal arrangements, contact your faculty's legal officer.

For more information about informed consent, see these EUR informed consent templates.

Examples:

- Research participants: Informed Consent, Confidentiality Agreement.
- Third party: Funder, Organisation, Company.
- Multiple research partners: Consortium Agreement.

Agreements to uphold

You should list agreements and other data management policies, this can be Funder, Institutional, or Discipline related policies relevant to Research Data Management.

Please make sure to include the terms of service (or use) for all the software tools that you will use for data collection, data analysis, collaborative writing, etc.

Please DOWNLOAD and save the agreements with your documents

Ethical approval

Ethical approval is required in case research participants are involved in the study to ensure what is asked of research participants does not infringe on participant health, safety, or wellbeing, in any way. Each faculty has its own ethical board and review process.

Be aware that some faculties have the requirement that **every research project** needs ethical approval, independent of whether it involves participants. You should therefore make sure that you know if this is the case at your faculty.

DURING RESEARCH: COLLECTING AND ANALYZING

File formats and file sizes

File format and size can differ significantly and may be reason to choose alternative tools. Video files, for example, can be several GB each in size, implying that standard storage capacity might not be enough. Here you can find a tool for calculating video file size and information on preferred file formats. For this question it is sufficient to estimate the total file size (e.g. <1GB; 1-5GB; 5-10GB; 10-15GB).

Data classification

The EUR data classification distinguishes four levels of data: Public, Internal, Confidential, Secret.

Personal data & sensitive personal data

If you have answered question 12 with 'yes', or 'I do not know', contact your faculty's Privacy Officer. You can find an overview of Privacy Officers at the EUR here (login with your ERNA).

(Sensitive) personal data:

- Data is personal when it can be used to identify a natural person, thus a specific bit of info may in itself not be personal, but in combination with other bits, it may become personal. Examples: name, address, e-mail, student number.
- Personal data is sensitive if it contains data on e.g. vulnerable groups or of sensitive content e.g. health or political views.

You can find more information on personal data by the European Commission here.

Pseudonymization & Anonymization

If you need to share (personal) data during the research project the data must at least be either pseudonymized or anonymized to protect the participants' privacy. If you need to transfer (sensitive) personal data, it is recommended to consult the faculty privacy officer for more information on how to do this and to determine whether additional measures are required such as sending the files encrypted. These measures depend on the degree of sensitivity of the data.

Pseudonymous means that the participant can only be indirectly identified, when, for example there is only a random participant ID in the datafile and a separate file (the key-file) contains the random ID and the personal data such as participant name.

(Fully) Anonymous means that no-one can directly or indirectly identify the research participant. This includes the research team. There is no key-file. In practice it is nearly impossible to truly anonymize a dataset.

Storage during the project

Several tools are available to store data during the research project, see the Tooling page for an overview. Which one you need depends on the type of data you have, the classification of the data (Public, Internal, Confidential, or Secret), and personal preference. For example, SURFdrive and SURF Research Drive may be used to store and share Internal or Confidential data, but not Secret data.

It is recommended to use EUR-supported tools as almost all are cloud-based (thus accessible from anywhere), safe, and regular back-ups are automatically made. However, sometimes these tools do not meet all needs, resulting in private tooling to be used. If this is the case it is recommended to consult the faculty privacy officer to determine whether measures need to be taken to ensure safety and privacy such as using a tool with encryption or setting passwords.

Software catalog

More info on the EUR software catalog can be found here.

Data backups

When using EUR supported tools for data storage (as listed in Q16), only to a limited extent backups are made automatically. These backups are implemented to prevent data loss in case of disaster or hardware failure on the server side (e.g., see the SURF Research Drive Backup Policy). Note that these backups usually do not suffice as backups for erroneously deleted or modified files. We therefore advice you to always arrange regular backups of your data yourself (on EUR supported tools). It is also good practice to separate raw data (saved as read-only) from processed data in different folders, to prevent accidental data loss.

Folder and file naming and structure

We recommend:

Naming your files - use descriptive and logical file names

1. Machine readable
   - No spaces, accents, or odd punctuation
   - Use hyphens, underscores, capitalization for separation of elements (snake_case, kebabcase, camelCase)

2. Human readable
   - Informative elements
   - Make use of formats (do not use document in the name since the format will tell you what it is) (.doc/.png/.pdf) (avoid redundancy)

3. Logically sorted
   - E.g. Begin file names with a date YYYYMMDD to sort chronologically
     So an example file name could be: *20210805_transcript_subject1234_v1.docx*

File versioning

We recommend:

Formalize updates using semantic versioning - Versioning has usually 3 elements 1.0.2

1. [1.] Major (changed drastically, e.g. new chapter)
2. [.0.] Minor (new functionalities, e.g. revised chapter)
3. [.2] Patch (small repairs, typos)

RESEARCH PUBLICATION: DATA SHARING AND RE-USE

Research data repository

There are many different types of data repositories. There are university-, funder-, and journal specific repositories, but also domain- or data type (e.g. code or MRI-scans) specific repositories, all of which can either be free to use or require payment.

The EUR also has its own data repository which can be used to deposit data (in all formats) underlying publications. To help make this data more interoperable and reusable, a data curator checks all files and makes suggestions for e.g. file format and metadata description. If possible, all files underlying a publication should be placed in the EUR Data Repository; if not possible, then the metadata should be there.

Please note that for each file restrictions may be set such as making it confidential. These and other features allow the majority of datasets to be placed in one repository. If a domain- or data type specific repository is more suited for your files, make sure you enter the metadata and the digital object identifier (DOI; the persistent 'digital address' of your files) in the EUR Data Repository

Metadata standard

For some research types or research domains, there are so-called meta-data standards available that describe a uniform way to assign metadata, thereby contributing to interoperability.

Restrictions on data re-use

Embargo = a period between publishing and release. This period can range from several weeks to several months.

Restricted access = as long as the data are held in the repository, any interested party needs to contact the owner (the researcher) for permission to access the files, and user conditions can be specified by the researcher.

Licensing for data

When you share data, you want to make sure that others who re-use it give you the credit you deserve and only use it for its intended purpose (e.g. yes or no to commercial use, share alike). There are many different types of licenses, but most common are the Creative Commons (i.e. CC-license).

The standard license in the EUR Data Repository is CC-BY, which lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation. If you do not know which license to choose from, this website will help you.

AFTER RESEARCH: ARCHIVING

Archive of EUR library

Currently, the EUR is working on a pilot for a data archive that is available for use; similarly there may be faculty specific archives or archiving protocols available.