

# ELIXIR Belgium: ELIXIR-BE DMP Template

## General Information

Name applicant.

Project number.

Project title.

Project acronym.

Affiliation.

## Datasets Information

[FWO, ERC, H2020] Will you generate/collect new data and/or make use of existing data? Specify how you will use existing data.

- Generate new data
- Reuse existing data

*Example Answer:*

1) I will re-use existing open access datasets on [add your topic] by combining them with my new data; no limits for sharing and re-use. 2) I will only generate new data; existing data on [this topic] are sparse and inadequate because [give reasons]. 3) I will re-use only open access metadata of [Dataset Name] with closed access data. 4) I will pay [X]€ to access [Dataset Name] from [Company Name]: data will not be shared, only metadata will be open access.

*Guidance:*

- **Existing data:** data previously generated by you or others.
- **New data:** data that will be generated/collected during this project.
- **Please, state the following in your answer:**
  - If there are existing data or not about your topic.
  - If there are restrictions or costs on the reuse and share of third-party data (specify the details in the next questions).
  - If re-use of any existing data has been considered but discarded and why (ex: incomplete data, data not reusable for legal reasons etc...).
  - How existing data will be used.

Where can I find existing data? [Read here](#).

[ERC] Name and list here all datasets that will be used and/or generated in this project. Add a reference for existing datasets.

*Example Answer:*

1) Existing datasets: [name and access number or reference or DOI]. New datasets: [your analysis type] by [your technique] on [your organism]. Software [Name] for [your application]. 2) New: RNA-seq on Arabidopsis. Phenotype analysis by imaging. Existing: Arabidopsis dataset E-MTAB-XXXX. 3) New: Methylation by LC/MS. 4) New: Biomodel of X protein complex.

*Guidance:*

The name of a dataset should be **short, informative and meaningful**. It can describe:

- the used technique (RNA-seq, Imaging, LC/MS etc...)
- and/or sample origin (organisms, literature, etc...)
- and/or collection methods (experiment, simulation, survey, questionnaire...)

**Reference for existing dataset:** any identifier or accession number for keeping track of data provenance.

How do I cite an existing dataset? [Read here](#).

[H2020] Per dataset, state its purpose, explain the relation to the objectives of the project, specify to whom it will be useful.

*Example Answer:*

1) Datasets [Name] and [Name] are needed to evaluate the role of [your factor] on [your variable], as described in the objective number [X] in this project. Other researchers and industries involved in [list topics] will be interested in these data.

*Guidance:*

A description of the purpose of the data to be collected/generated will help reviewers understand the impact of your research on academic community, industry and society.

[FWO, ERC, H2020] Per dataset, state its origin/source.

*Example Answer:*

1) Datasets [Name] and [Name] are [laboratory/field/preclinical experiments] on [organism]. 2) Datasets [Name] consists of measurements performed in the lab by a partner in a different (part of the) country. 3) Dataset [Name] is a combination of existing data and new experimental data of [X] and [Y] platform/consortium. 4) The source of [dataset name] is a collection of existing [studies/books/publications]. 5) The source of datasets [Name] and [Name] are quantitative [survey/interview/questionnaire] collected by a team of survey takers we hire. 6) Dataset [Name] is a [qualitative or quantitative observational] study on [population/topic/subject].

*Guidance:*

List of possible attributes for data source/origin or collection mode:

- Observational
- Experimental
- Quantitative
- Qualitative
- Surveys or questionnaires or interviews
- Simulation
- Derived/compiled from other sources
- Digital (born-digital or digitized) or non-digital nature (e.g. paper surveys, questionnaires...)
- Primary (generated by the researcher for a particular research purpose or project)
- Secondary (originally created by someone else for another purpose)
- Raw
- Processed

Specifying where the data come from, or when and by whom data will be generated/collected helps to identify implications for privacy (GDPR), IP and other legal or ethical aspects.

[FWO, ERC, H2020] Per dataset, state digital format(s) of raw and processed data files, distinguishing proprietary from open format(s).

*Example Answer:*

1) Raw and/or processed [numeric/video/audio/text...] data files of datasets [Name] and [Name] are in this [proprietary/open] [formats]. 2) Raw images of the dataset "Phenotype analysis by imaging" are in open format JPEG; spreadsheet of the processed quantification data are in open .csv format.

*Guidance:*

- Raw and processed data file formats can be instrument/software-dependent, so check the formats generated by the instrument/software you will use.
- Try to convert proprietary format into open format to ensure that your data will be usable in the future.
- The required digital formats of raw and processed data could vary depending on the data repository you will use to share the data, so check what formats are accepted by the chosen repository on its website.
- Is my format open or proprietary? [Read here](#).

[ERC, H2020] What methods or software tools are needed to access data files in proprietary format? Is documentation about the software needed to open the data file provided in the metadata? Is it possible to provide the relevant software (e.g. in open source code)?

*Example Answer:*

1) Data files in proprietary formats [x,y,z] can be accessed by the software [X and Y], which are open; software info will be described in the documentation associated with the data files. 2) Format [x] can only be opened with the proprietary software [Y]; no open format nor open software exists for this data type; software info will be described in the documentation associated with the data files. 2) Software to access data files will be provided as open source.

[FWO, ERC, H2020] Per dataset, state its expected volume at the end of the project.

*Example Answer:*

1) Phenotype analysis: X images in Y format is about XXX GB; RNA-seq on Arabidopsis: 200 files are about 200GB.

*Guidance:*

Data volume doesn't have to be precise; a realistic range of the data volume is sufficient.

#### **How to estimate dataset volume:**

1. Consider at least all raw data files. Check if processed data are also required by repositories or journals.
2. Estimate file size per sample or experiment, based on files previously generated using similar setting.
3. Multiply the estimated file size by the number of samples or experiments you are going to generate during the project.

#### **To estimate the volume of**

- Sequencing data files: [see here](#).

## **Datasets Sharing and Reuse**

[FWO, ERC, H2020] Which datasets will be shared and made available (for the society) after the end of the project? Distinguish open access datasets from restricted or closed access datasets.

*Example Answer:*

1) Open access: Datasets [Name] and [Name]; Restricted access: Datasets [Name] and [Name]; Closed access: Datasets [Name]. 2) All datasets will be open access.

*Guidance:*

What does open, closed and restricted access mean?

- **Open access:** there are no restrictions on access to the data; anyone can view and download a copy.
- **Closed access:** a description of your dataset is published (metadata); however, the dataset can only be accessed by its subject, owner or holder.
- **Restricted access:** a description of your dataset is published (metadata) and provides information on how to apply for access to the data (registered access, access upon request etc...).

Embargoed: the dataset can't be accessed until the end of a specified time period, after which it becomes Open, Closed or Restricted access.

What is the best type of access for my dataset? [See here](#).

[FWO, ERC, H2020] Are there any factors restricting or preventing the sharing of (some of) the data? Per dataset with restricted or closed access, provide a rationale for doing so.

- No
- Yes, Intellectual Property (IP): the researcher involved and the IP team shall make the necessary arrangements in order to maintain the embargo on the data access, at least until essential steps in securing intellectual property have been taken.
- Yes, 3rd party agreement or classified data: the researcher involved and the Legal Team of the TechTransfer Office will specify the restrictions below, in this DMP.
- Yes, ethical and/or legal issues (personal/sensitive data): the researcher involved and the Data Protection Officer will specify how compliance to GDPR impacts data sharing below, in this DMP.
- Other. Specify below.

*Example Answer:*

1) The legal agreement with a 3rd party states the following: [explain briefly]. 2) Dataset [Name] contains personal data that require restricted access, only metadata will be open.

*Guidance:*

Explain the legal and/or ethical restrictions that impact data sharing.

If you are processing personal data, refer to the GDPR section.

[FWO, ERC, H2020] Per dataset, list the repositories in which data, metadata, code, software and related documentation will be shared. If research materials and protocols will be shared in centralised repositories, specify it below.

- In an Open Access repository
- In a restricted access repository
- Upon request by email
- Other (specify below)

*Example Answer:*

1) Datasets [Name] and [Name]: [Repository Name]. 2) RNA-seq on Arabidopsis: ArrayExpress. Code on Github. 3) Biomodel of X protein complex: BioModels. Code and software on Github. 4) Plasmids in Addgene.

*Guidance:*

**Important:** some repositories give open access to the data as default; while other repositories allow restricted or closed access. Check the access conditions to the data before choosing a repository.

How to choose the appropriate repository: [see here](#).

[FWO, ERC, H2020] Per dataset with restricted access, explain how access will be provided (under what conditions or procedure).

*Example Answer:*

1) Access to [restricted dataset] can be granted only [explain under what condition]. 2) Access to restricted datasets will be granted by a Data Access Committee (DAC). 3) Access to restricted datasets will be granted by the Institutional authentication procedure.

*Guidance:*

Check what conditions or procedures are required by repositories to access restricted datasets.

To know more about types of restricted access [see here](#).

[FWO, ERC, H2020] Per dataset, state when it will be made available. Is embargo foreseen?

- Immediately after the end of the project
- Upon publication of the research results
- After an embargo period. Specify the length of the embargo and why it is necessary below

*Example Answer:*

All datasets: Upon publication of research results.

*Guidance:*

Embargo: it is a defined period of time, imposed by the author, which prohibits access to data. After the embargo period, data becomes Open, Closed or Restricted access.

Check if the repositories of your choice allow embargo period

[ERC, H2020] Per dataset, state how it will be licensed to permit the widest reuse possible. If the re-use of some data is restricted, explain why and specify the length of time for which the data will remain re-usable.

- Each dataset will follow the license policy of the repository in which it will be shared.
- All datasets: CC-BY-4.0. No time limit for reusability.
- Other. Specify below.

*Example Answer:*

1) All datasets: CC-BY-4.0. No time limit for reusability. 2) Software [Name]: AGPLv3.

*Guidance:*

Check if:

- Your funder requires application of specific license on research data.
- The repository applies one default license to all datasets.
- You can attach your own license to your datasets in the repository.

What type of license should I use for my dataset? [See here](#).

## Documentation and Metadata

[FWO, ERC, H2020] Will a metadata schema be used? Per dataset, state the metadata standards or the customized metadata that will be used.

- Yes. Each dataset will follow the metadata schema of the repository in which it will be shared.
- Yes. The chosen metadata standards will be described below.
- Yes. The customised metadata schema will be described below.
- Other. Specify below.

*Example Answer:*

1) All datasets will be described according to the discipline-specific metadata standard [Name of the standard]. 2) Since standard metadata schemas are not suitable, we will use the following metadata schema to describe all datasets: i)Title, ii)Description, iii)Date, iv)Creator, v)Rights/license, vi)Format, vii)Volume, viii)Experimental factors, ix) Species, x)Observational Unit, xi)Replicates, xii)Response variable, xiii) Technique, xiv)Experimental design [other].

*Guidance:*

A metadata schema is a set of information (also called metadata fields or attributes) that you need to give about your dataset. Metadata should provide all the necessary information to enable finding, understanding and reuse of datasets by anyone.

Metadata gives information both at the overall study/project level (such as authors, aims, date etc) and at the individual data point or observation level (such as variables names and relation between files).

Check if/what metadata schema is required by repositories.

For information about metadata schema [see here](#).

[ERC, H2020] Do you make use of persistent identifiers or unique identifiers such as Digital Object Identifiers (DOI)?

- Each dataset will have the identifiers or accession numbers provided by the repository in which it will be shared.
- We will generate a persistent identifier or a DOI for each dataset. Details will be specified below.
- Other. Specify below.

[ERC, H2020] Will you use a controlled vocabulary or standard ontology to describe your data? Per dataset, state the standard ontologies you will use or the mapping of your vocabulary to a standard ontologies.

- Each dataset will use the ontology supported by the repository in which it will be shared.
- A customised vocabulary will be used and the mapping to standard ontologies will be explained in documentation and summarised below.
- Datasets require the use of uncommon or project specific ontologies or vocabularies that cannot be mapped to commonly used ontologies. Definition of the used terms will be provided in documentation.
- Other. Specify below.

*Example Answer:*

1) Ontologies required by the chosen data repositories will be used. 2) All datasets will be described by using the following ontologies: Experimental Factor Ontology (EFO), Statistics

Ontology (STATO) and Plant Ontology (PO). 3) For datasets [Name], standard ontologies are not suitable, but we will create a controlled vocabulary and map it to the standard ontologies [Names].

[FWO, H2020] What documentation will be provided to enable understanding and reuse of data collected/generated in this project?

- Metadata will be provided as an attached README.txt file in each repository. Details will be specified below.
- Metadata will be provided in the repositories' dedicated metadata fields. Details will be specified below.
- Metadata will be provided as structured machine readable files, according to the repositories's requirements. Details will be specified below.
- Other. Specify below.

*Example Answer:*

1) In ArrayExpress "Protocols" section, we will explain the file naming convention, column names for tabular data, performed calculations, materials and methods. Identifier of the associated open access article describing materials and methods will be provided. 2) In Github, each well-annotated script will have a README.txt file containing all metadata and the following additional information: version, how to use the script, [etc...]. 3) In Zenodo, metadata and additional information will be provided in the "Description" section. Metadata about each data entry will be provided as structured machine readable .csv file in attachment.

*Guidance:*

Describe what documentation will be provided for the public when sharing your datasets. Documentation, **containing metadata and additional information**, should allow anyone to fully understand the content of the datasets. As metadata, documentation can contain general information about the overall study/project level (such as authors, aims, date etc) and/or structured description at the individual data point or observation level (such as variables names and relation between files).

Check if the repositories of your choice allow upload of documents or README file as an attachment to datasets.

If you want to know more about documentation and README file [see here](#).

## Costs for Datasets Sharing and Reuse

[H2020] What are the estimated costs for making data more FAIR (Findable, Accessible, Interoperable, Reusable)?

*Example Answer:*

1) Making data more FAIR will not have any cost, because we will use open source ELN, open source resources and software for data management and analysis during the project (DMPonline, OSF, FAIRDOMHub, UseGalaxy.be, R, GitHub). 2) Making data more FAIR will not have any cost, because we will use recommended open source software and resources for data management. A common system for organisation of folders/files and curation of documentation has been implemented by the lab manager. 3) Costs of data compliance consultancy will be approximately X€.

*Guidance:*

Possible costs of making data more FAIR could be due to payment for tools or experts in order to:



- make this DMP
- acquire or transcribe or clean existing data
- harmonise file naming and organisation of folders in your project
- adopt open and standard file formats
- make your dataset machine readable
- implement metadata schemas and ontologies in your project
- draft consent form for data sharing and/or apply license
- curate documentation about datasets
- create DOIs/PIs
- guarantee data security and/or transfer data across sites
- comply with legal requirements

**(for costs about data storage, backup, archiving and data sharing, see next questions)**

[H2020] How will these costs (for making data FAIR) be covered?

*Guidance:*

Check what costs are eligible as part of the grant.

[FWO, ERC, H2020] What are the expected costs for data sharing and reuse?

- Data will be shared and made available in public repositories which are free of charge.
- Other. Specify below.

*Guidance:*

Possible costs for data sharing and reuse:

- fee to publish data in repositories
- fee to transfer data
- paying an expert for support about licensing

[FWO, ERC, H2020] How will these costs (for data sharing and reuse) be covered?

*Example Answer:*

1) No costs expected. 2) These costs will be covered by [X] funding.

*Guidance:*

Check what costs are eligible as part of the grant.

## **(Meta)Data Storage During the Project**

[FWO, ERC, H2020] Where will the data be stored during the project?

*Example Answer:*

1) On [techn type], hosted [by]. 2) Small digital file containing confidential data will be stored on secure network drive [X], hosted and maintained by university ICT department. 3) Big digital file [>XXX GB], on external cloud platform [X], contracted by our research department. 4) On Sharepoint of our collaborator at [X] institute.

*Guidance:*

Check the storage services available in your Institute or department with the IT Team or on the website.

List data storage system for every type of data: big or small digital file; files generated with external or internal collaborators; files containing personal or not-personal data; paper or digital files.



[FWO, ERC, H2020] What is the backup strategy?

*Example Answer:*

- 1) On the shares, 'snapshots' of the data are made every [X] hours. Snapshots allow you to retrieve (older versions of) files.

*Guidance:*

Check the website or ask the IT Team of your department or Institute for information about the data backup system.

[FWO] Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes. See details below.
- No. See details below.

*Example Answer:*

- 1) Storage is allocated in a project based manner (according to estimated volumes of data generation and analysis); sufficient storage capacity will be ensured for the project at [Institute Name].

*Guidance:*

Check the website or ask the IT Team of your department or Institute for information about storage and backup capacity available for your project.

Briefly explain how storage and backup for the data will be guaranteed.

[FWO, ERC, H2020] What are the expected costs for data storage and backup during the project?

*Example Answer:*

- 1) 4 years storage and backup for this project will cost about 2000€.

*Guidance:*

Specify the costs charged to your research team.

Multiply the cost in €/TB/Year, by the number of years and the volume of data which you are going to generate (as specified in the section above).

[FWO, ERC, H2020] How will these costs (for data storage and backup during the project) be covered?

*Example Answer:*

- 1) Costs for data storage and backup during the project will be covered by [X] funding.

*Guidance:*

Check possible funding for covering data storage costs.

[FWO, ERC, H2020] Data security: how will you ensure that the data are securely stored (not accessed or modified by unauthorised persons), especially for sensitive data.

*Example Answer:*

- 1) My laptop will be password-protected. The internal external hard drive (for backups) will be encrypted using [Encryption system name] (as suggested by IT). Files holding personal data will also be encrypted with password-protection. Read-only permissions will be implemented for raw data files on the Network Shared Drive so that they cannot be edited. Files containing personal data will be stored in access-controlled folders on the university network drive to protect privacy.

2) Data will be handled in accordance with university information security guidelines, involving measures proportionate to their nature and the risks involved. Lab computers and external drives will be password-protected, and the rooms in which they are kept will be locked when no lab members are present. Security upgrades to operating systems will be regularly performed. Read-only permissions will be implemented for raw data files so that they cannot be edited.

*Guidance:*

Check the website or ask the IT Team in your department or Institution for information about the data security system.

For personal/sensitive data, specify how the system complies to legal requirements and refer to the GDPR section.

## **(Meta)Data Long Term Preservation**

[FWO, ERC, H2020] List which datasets will be preserved or not, for at least 10 years after the end of the project and explain why (legal, contractual restrictions etc..).

*Example Answer:*

1) All the generated datasets will be stored for long term. 2) All relevant data for reuse and experiment replicability will be preserved for 10 years after the end of the project. 3) Dataset [Name] will be deleted after [X] years because of the following [contractual, legal, ethical...] reasons: [explain].

[FWO, ERC, H2020] Where will datasets be stored for long term (archived) after the end of the project?

- The same repositories used for data sharing will ensure long term storage.
- Other. Specify below.

*Example Answer:*

1) All datasets will also be stored for long term and archived in the [X center] of [University Name]. 2) Datasets to which data sharing restrictions apply will be stored in-house according to the following secure measures: [specify].

*Guidance:*

Check the website or ask the IT Team in your department or Institute for information about long term storage and data archive.

[FWO, ERC, H2020] Estimate the costs for long term preservation (or archiving).

- The chosen data repositories provide long term storage free of charge.
- Other. Specify below.

*Guidance:*

Specify the costs charged to your research team.

Multiply the cost in €/TB/Year, by the number of years and the volume of data which you are going to generate (as specified in the section above).

*Example Answer:*

1) Long term preservation of [X]TB for 10 years at [University Name] will cost about 2000€.

[FWO, ERC, H2020] How will these costs (for long term preservation) be covered?

- No costs expected.
- Other. Specify below.

*Example Answer:*

- 1) These costs will be covered by [Name] funding.

*Guidance:*

Check possible funding for covering costs of data storage.

[ERC, H2020] Data security: are the datasets safely stored in repositories for long term preservation and curation?

- Yes, the chosen data repositories will ensure data security.
- Other. Specify below.

*Example Answer:*

- 1) The archiving system of [University Name] will guarantee data security.

## **Responsibilities**

[ERC, H2020] Describe data quality assurance processes and who will be responsible for it.

*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 ([Science Europe](#)). People responsible for quality assurance will ensure that appropriate quality documents, such as Standard Operating Procedures (SOPs), working instructions, conventions, guidelines, forms, templates, logs, tabs, and labels, are determined, developed and implemented.

[FWO, H2020] Who will be responsible for the data documentation and metadata?

*Example Answer:*

- 1) The experimentalist [PhD student, PostDoc...] will be responsible for day to day operations, and PI for the overall responsibility. 2) The PI (and/or dedicated people from his/her research team) bears the overall responsibility for documentation and metadata.

[FWO, H2020] Who will be responsible for data storage and backup during the project?

*Example Answer:*

- 1) The research and the IT team will be responsible for data storage and backup.

[FWO, H2020] Who will be responsible for ensuring data preservation and sharing?

*Example Answer:*

- 1) The the repositories [Names] holding the datasets will be responsible for data preservation and sharing, as outlined in the terms and conditions by which they offer their services. The Principal Investigator is responsible for local data preservation and sharing, with the support of the IT Team. 3) Local data preservation will be the responsibility of the IT Team of [Institute Name]; the PI bears the overall responsibility for data sharing.

[FWO, H2020] Who bears the end responsibility for updating and implementing this DMP?

*Example Answer:*

- 1) The Principal Investigator bears overall responsibility for all data management during and after data collection, including updating and implementing the DMP.

## Ethical and Legal Issues

[FWO, H2020] Will you use personal/sensitive data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.

- No.
- Yes. Specify below.

*Example Answer:*

- Privacy Registry Reference: - Short description of the kind of personal data that will be used:

*Guidance:*

Before starting a project, you should contact your Data Protection Officers (DPO) to be informed of GDPR compliance requirements for your institution.

You should inform your local legal office of your project's setup and identify the necessary agreements to be signed.

The proposal for processing of personal/sensitive data has to be submitted and approved by the appropriate Privacy Registry Reference. Write here the reference to your file.

[FWO, H2020] Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s).

- No.
- Yes. See below the list all the relevant ethical approval numbers and the title of the related project.

*Example Answer:*

1) Ethical approval XXXXXX: [Dataset name or project title]. 2) Ethical approval S56289: Transcriptomic in human primary cells.

*Guidance:*

The proposal has to be submitted and approved by the appropriate Ethics Committee. Write here the reference to your file.

[FWO] Will Intellectual property restrictions be claimed for your data? If so, for what data and which restrictions will be asserted?

- No.
- Yes. Specify below.

*Example Answer:*

1) If there is substantial potential, the invention will be IP protected (mostly patent protection); data will be made public after the IP application becomes public (X months after IP application).

2) If there is substantial potential, the invention will be IP protected. Data will be under embargo until IP application becomes public (X months after IP application). The Legal Team of the TechTransfer Office will be consulted.

*Guidance:*

Consult with the Legal Team of the TechTransfer Office.

[FWO] Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No.
- Yes, specify below.

*Example Answer:*

1) Because of 3rd party agreements, the following restrictions are in place for dataset [Name]:  
[explain].

[H2020] Indicate whether other national/funder/sectorial/departmental procedures for data management are used.

- No.
- Yes. Specify below.

## **Other**

[H2020] Indicate whether other national/funder/sectorial/departmental procedures for data management are used.

- No.
- Yes. Specify below.