**FAIR data management plan guidelines**

This document is aimed at supporting university staff members in writing Data Management Plans (DMPs) in order to make data produced along the research data lifecycle: Findable, Accessible, Interoperable and Reusable (FAIR).

This document has been realised with the support of *Marco Carraro PhD* (Department of Biomedical Sciences, University of Padua) in the context of the [ELIXIR-CONVERGE project](https://elixir-europe.org/about-us/how-funded/eu-projects/converge) (GA 871075).

*Useful contacts:*

Data protection officer - [Responsabile Protezione Dati](https://www.unipd.it/privacy)

Ethical committees

* [Comitato etico per la sperimentazione clinica CESC](http://www.aopd.veneto.it/sez,3450)
* [Comitato etico della ricerca psicologica](http://ethos.psy.unipd.it/)
* [Comitato etico Human Inspired Technology Research Center](http://hit.psy.unipd.it/comitato_etico_HIT) [Comitato etico Dipartimento di Scienze Biomediche](http://www.biomed.unipd.it/dipartimento/organi-collegiali-e-commissioni/)

GDPR advisor - [International Research Office](https://www.unipd.it/en/international-research-office)

Intellectual Property (IP) advisor - [Technology Transfer Office](https://www.unipd.it/trasferimento-tecnologico)

*Useful resources:*

Data policy - [Research Data Management Policy](https://www.unipd.it/sites/unipd.it/files/2018/policy%20dati%20ricerca.pdf)

Open access policy - [Regolamento per l’Accesso Aperto alla produzione scientifica](https://www.unipd.it/sites/unipd.it/files/2017/REG%20accesso%20aperto%20produz%20scientifica%2001082017.pdf)

Local data repository - [Research Data UniPd](http://researchdata.cab.unipd.it/)

All about publishing at UNIPD - [The University Library System](http://bibliotecadigitale.cab.unipd.it/en/)

Consultancy service for Data Management / DMP - [Biocomputing UP lab.](https://protein.bio.unipd.it/) - [Fees applied](https://drive.google.com/file/d/1UEkqFJdMHe4Zrye7013h0wk6GdHORi-i/view?usp=sharing)

**Table of content**

[What](#_ye4b7xy0u9t)

[Why](#_j30angqz6ijr)

[Before you start](#_lsyfvgxnxbn1)

[Online tools](#_gttl27k5xlc3)

[Recommended chapters](#_vjkryb9hahm5)

[1. Project overview](#_r59rptlvvp3b)

[2. Data summary](#_6ubgees460cn)

[3. FAIR data](#_bdck7bld09u9)

[4. Allocation of resources](#_cqgnds9tjm8l)

[5. Data security](#_9kwa50vmcr7i)

[6. Ethical aspects](#_wf5baqptxeo5)

# 

# What

A DMP is a document where researchers describe how they are going to manage data during and after the project: what data they will generate; how the data will be documented, described, secured and curated; and who will have access to those data after the research is completed.

# Why

DMP is a tool to encourage researchers to value the data they generated as much as the scientific article, in which that data is used. DMP should help preventing data loss and encourage data sharing according to the FAIR principle. Sharing data reduces data duplication and increases reproducibility and transparency of the scientific research. Moreover, accessible data can be reused by others to generate new knowledge.

Overall, DMP is a tool to improve the quality of your research and boost its impact.

In addition to this, projects funded under the Horizon Europe program, are requested to submit aDMP as deliverable by month 6 ([Horizon Europe GA](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf), article 17, annex 5, pag 108) and have it reviewed towards the end of the project. For guidance on open science practices and research data management, please refer to the relevant section of the [HE Programme Guide](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf).

# Before you start

Prior to write a DMP, it is important to think about a number of aspects of your research project. The paragraphs listed in this document will provide practical guidance on how to tackle many Data Management (DM) issues that need to be addressed in a DMP, compliant with [FAIR principles](https://en.wikipedia.org/wiki/FAIR_data).

FAIR is a faceted concept and it could be easily confused with similar principles like Open, since they share a similar philosophy when it comes to managing and sharing data. Due to this, it is crucial to understand what FAIR is not.

* FAIR is **not equivalent to open** (and open is not equivalent to ‘free’): there are many reasons why data may be non-open and only available under certain conditions to certain users. As long as accessibility conditions are properly described, non-open data can be entirely FAIR. Reciprocally, fully open and unrestricted data may score very low in FAIR metrics as they may for instance be non-actionable for machines.
* FAIR is **not a standard**, although the acronym is frequently used in that context, FAIR has to be considered more like a guiding principle.
* FAIR principles do not, in themselves, guarantee **intrinsic data quality or ethics**.

# Online tools

A range of tools and services are available online, designed to help researchers in writing DMPs and to provide support in managing research data.

Below is a list of recommended online tools that meet the needs of a wide range of users providing either advice on writing the very first DMP, or support in highlighting possible weaknesses of DM practices already in place. Below is a list of recommended online tools that meet the needs of a wide range of users providing either advice on writing the very first DMP, or support in highlighting possible weaknesses of DM practices already in place.

[**DMP Online**](https://dmponline.dcc.ac.uk/) **- Generate your DMP, step by step.**

*This resource lists the main topics that have to be described in a DMP.*

DMP Online is a web-based tool for creating DMPs, developed in order to help researchers to meet funder requirements on DM. The tool is centered on DMP templates created by the funders (or based on funders guidelines). Once selected a funder template, the user is guided through “sessions” and “questions” to be addressed by filling out text boxes. Users are able to include and exclude individual clauses according to their specific needs. At the end of the process, researchers will be allowed to export their plans in PDF format or other formats.

When all the requested questions will be properly addressed, the final document downloaded from DMP Online could reasonably be considered the draft document of your DMP.

As this tool is focused on posing the right questions, we encourage researchers to use DMP Online in association with another web based tool, the Data Stewardship Wizard, in order to be guided in finding useful suggestions on how to address those questions.

In addition to this, we encourage the use of DMP Online in order to have access to a plethora of DMP documents made available for free by other users.

[**Data Stewardship Wizard**](https://ds-wizard.org/) **- Build your knowledge model.**

*This resource raises the main questions to be addressed in a DMP and suggests useful answers.*

The Data Stewardship Wizard consists of an open source web questionnaire tool.

Data Stewardship Wizard has a different (complementary) approach if compared to other on-line resources, like DMP Online: most questions are closed questions with a limited set of possible answers. Questioning in the Data Stewardship Wizard is modelled after a conversation a researcher could have with a DM expert (they refer to this expert embodied system as: *Knowledge Model*). Questions are based on the previous answer that is selected by the user and follow-up questions are added to the questionnaire. Following the questionnaire users will be guided in dissecting their problem and advised on where to find possible solutions as some answers may be obtained from linked services, such as [FAIRsharing.org](https://fairsharing.org/) containing a curated database of standards, policies and databases, mainly (but not limited to) life science.

Based on their answers, at the end of the questionnaire users will end up with a customized *Knowledge Model*, rather than a complete DMP document, that could be of help in filling out empty text boxes provided by other tools, such as the DMP Online tool.

[**ELIXIR Research data management toolkit**](https://rdmkit.elixir-europe.org/) **- Find suggestions about your DM issues.**

*This resource provides tips, suggestions and use cases along the whole research data lifecycle.*

The ELIXIR Research data management toolkit is a resource designed to guide researchers in managing their data following the FAIR principles in the area of life sciences. It is based on the stages of the data lifecycle and content can be navigated based on your ‘role’, ‘domain’, ‘examples’ and ‘problems’. Contents are generated and maintained by data stewards and experts of the [ELIXIR](https://elixir-europe.org/) community working together with final users.

# Recommended chapters

This section is based on the [Horizon 2020 DMP template](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm) and it outlines the main topics to be addressed in a DMP document.

***Cover page***

We suggest creating a cover page including the following information

|  |  |
| --- | --- |
| ***PROJECT NAME*** | |
|  |  |
| EU flag | Project logo |

*Grant Agreement number:*

*Action Acronym:*

*Action title:*

*Deliverable number and title:*

*Work package number and title:*

*Dissemination level: (Public, Sensitive, EU restricted, EU confidential, EU secret)*

*Authors:*

*Delivery Date:*

*DMP version:*

## 1. Project overview

Introduce your research project allowing people who are not directly involved in the study (E.G. project officers, reviewers) to have an idea of the purpose of your research. While writing this paragraph, take some time to review the grant request proposal. Go through the various steps of your research and highlight milestones, deliverables and expected outcomes. This process will help you in focusing on the data that will be involved in your research project. It will support you in the identification of which datasets will be needed, generated, collected, reused or analysed to address a specific scientific question, allowing your research project to move forward. While going through your proposal, track down all these datasets and tag them with the reason why that data is needed for the proceeding of your research. With this shortlist in your hands it should be much easier to address the next topic listed in the DMP template.

## 2. Data summary

Based on the list you generated in the previous step, try to address the following questions for each dataset:

● What types of data your dataset is about?

*e.g. interview, survey, clinical measurements, medical records, electronic health records, administrative records, archive, database, images, video recordings, etc..*

● What is the purpose of the data collection/generation and its relation to the objectives of the project?

● Specify the file format of your dataset and the software used for its generation

[List of open formats from Wikipedia](https://en.wikipedia.org/wiki/List_of_open_formats)

[Recommended formats from the UK Data Service](https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats)

[Recommended formats from the Data Archiving and Network Services](https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats)

● Specify if existing data is being reused and how (if any)

● State the expected size of the dataset

*e.g., MB, GB, TB, PB*

● Outline the data utility: to whom will it be useful

## 3. FAIR data

For each dataset, describe how to make it FAIR.

**Making a dataset findable, including provisions for metadata**

*Tip: is it all about unique persistent identifiers?*

Are the dataset produced and/or used in the project discoverable with metadata, identifiable and locatable?

*Dataset should be identifiable and locatable by means of a standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers)*

What naming conventions do you follow?

*Define file and folder naming convention used. Naming records (according to agreed conventions) is a logical and predictable way to distinguish similar records in order to easily store and retrieve them.*

[*Best practices for file naming at Stanford University*](https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming)

Will search keywords be provided that optimize possibilities for re-use?

*Indicate how potential new users can find out about your data and identify whether they could be suitable for their research purposes. For example, you may provide basic discovery metadata online (i.e. the title, author, subjects, keywords and publisher)*

Do you provide clear version numbers?

[*Document versioning guidelines at National Institute of Health NIH*](https://files.nccih.nih.gov/s3fs-public/CR-Toolbox/Version_Control_Guidelines_ver2_07-17-2015.pdf)

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

*Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a ‘readme’ text file, in file headers etc. Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place.*

[*Metadata standards catalog from Digital Curation Center (DCC)*](http://www.dcc.ac.uk/resources/metadata-standards)

[*Metadata standards catalog from University of North Carolina at Chapel Hill (UNC)*](https://guides.lib.unc.edu/metadata/standards)

**Making a dataset openly accessible**

*Tip: is it all about controlled access?*

Which data produced and/or used in the project be made openly available as the default? If it cannot be shared (or need to be shared under restrictions), explain why. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement.

*If some data will be closed provide the rationale for doing so. Specify any restriction (e.g. the study team’s exclusive use of the data, patentable data, Intellectual property rights (IPR) protection), clearly separate legal and contractual reasons from voluntary restriction. Restrictions on the release of data may be allowed, to protect confidentiality and for other ethical and legal considerations. Access to and use of Sensitive and confidential data can be restricted and regulated using end user licenses, data sharing agreements or by using a data enclaves.*

Where will the data and associated metadata, documentation and code be deposited?

*Preference should be given to certified repositories that support open access and long term preservation and curation where possible (e.g. by deposition in a repository).*

[*Find a discipline-specific repository at re3data.org*](http://www.re3data.org)

*UniPD Local data repository -* [*Research Data UniPd*](http://researchdata.cab.unipd.it/)

What methods or software tools are needed to access the data?

*Are there well described conditions for access (i.e. machine readable licence)?*

*Is documentation about the software needed to access the data included?*

*Is it possible to include the relevant software (e.g. in open source code)?*

If there are restrictions on use, how will access be provided?

*Is there a need for a data access committee?*

*How will the identity of the person accessing the data be ascertained?*

**Making a dataset interoperable**

*Tip: is it all about standards/vocabularies/ontologies?*

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc?

*i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins.*

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

[*Find topic-related standards, vocabularies and ontologies at FAIRsharing.org*](https://fairsharing.org/)

Will you be using standard vocabularies for all data types present in your dataset, to allow interdisciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

**Making a dataset reusable**

*Tip: is it all about adding a license?*

How will the data be licensed to permit the widest re-use possible?

*Choose the licence that better fit your needs using these tools:*

[*Creative Commons license chooser (beta)*](https://chooser-beta.creativecommons.org/)

[*UFAL license selector*](http://ufal.github.io/public-license-selector)

When will the data be made available for re-use and how long will the data remain reusable?

*If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*

Are the data produced and/or used in the project usable by third parties, in particular after the end of the project?

*If the re-use of some data is restricted, explain why.*

Are data quality assurance processes described?

*Data quality assurance is a process aiming at discover data inconsistency, missing information, application of wrong data analysis methods, in order to guarantee data precision and accuracy.*

[This tool could be helpful to support you in making simple data quality checks: OpenRefine](https://openrefine.org/)

## 4. Allocation of resources

Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified.

What are the costs for making data FAIR in your project? How will these be covered?

*Note that costs related to open access to research data are eligible as part of the Horizon 2020 and Horizon Europe grant (if compliant with the Grant Agreement conditions).*

*Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services. Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data*

*A widely accepted rule of thumb envisages that 5% of total research expenditure would be required to properly finance data management, storage and long-term preservation.*

[*Costs of data management from the Utrecht University*](https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management)

[*Data management costing tool and checklist from the UK Data Archive*](https://ukdataservice.ac.uk/media/622368/costingtool.pdf)

[Calculate costs of data management from Data Stewardship Wizard](https://ds-wizard.org/)

Who will be responsible for data management in your project?

*List who is in charge for data generation/collection/management (e.g. researcher, research assistant, commercial partner, project coordinator); data security, data archiving (e.g. qualified data manager certified in disclosure risk management, data manager with expertise in collection and processing of research data, etc). Individuals should be named where possible.*

Are the resources for long term preservation discussed?

*Consider who decides what data will be kept and for how long. Define the computing environment for data preservation: ensure readability and re-usability of data long after they have been generated: in other terms employ data formats that are likely to be readable even after 10-20 years.*

## 5. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

*Briefly describe the technical measures that will be implemented in the short to medium term to ensure data integrity (data should remain intact and unaltered during updates, e.g., data access with password, input validation to preclude the entering of invalid data, error detection/data validation to identify errors in data transmission), recoverability (prevention of data loss, e.g., cloud backup service - avoid only local storage), security (to prevent unauthorized access, e.g. hard disk encryption or cloud solution with encryption, transmission control as SSL certificate for websites, virus/malicious intruder protection, protection of wireless network as WPA, WEP or PSK).*

## 6. Ethical aspects

Ethical issues may affect how you store data, who can see/use it and how long it is kept. You should show that you’re aware of this and have planned accordingly.

Are there any ethical or legal issues that can have an impact on data sharing?

*If your data is sensitive (e.g. detailed personal data, politically sensitive information or trade secrets) you should discuss any appropriate security measures that you will be taking. Identify risk level and risk probability (low, medium, high) and provide a plan of monitoring activity to avoid risk of disclosure of information and countermeasures in case of information’s disclosure. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements.*

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

**Acknowledgments**

The authors would like to acknowledge the following initiatives, inspiring the content of the present document.

*The* [*RDM Guide*](https://rdm.elixir-belgium.org/about_DMP.html) *- ELIXIR Belgium*

*The* [*GO FAIR initiative*](https://www.go-fair.org/go-fair-initiative/)

*The* [*RDMkit*](https://rdmkit.elixir-europe.org/data_management_plan.html) *- ELIXIR CONVERGE*