

UNIVERSA
UNIVERSIS
PATAVINA
LIBERTAS

FAIR DATA MANAGEMENT PLAN GUIDELINES AND ANNOTATED TEMPLATE

*Released by the International Research Office of the University of Padua
Based on the Horizon Europe Data management Plan Template (version V1.1 of
1 April 2022)*



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



This document, released by the International Research Office at the University of Padua (Unipd) is aimed at supporting researchers in writing their Data Management Plan (DMP), as to make the data produced along the research data lifecycle “FAIR”, that is Findable, Accessible, Interoperable and Reusable.

Read the introductory guidelines in Section 1 carefully, before you start filling in your own DMP, using the annotated template in Section 2 as a draft.

TABLE OF CONTENTS

SECTION 1 - DMP GENERAL GUIDELINES	3
What is a Data Management Plan?	4
Why is it necessary and important	4
Before you start.....	4
Online tools.....	5
SECTION 2 - THE DMP ANNOTATED TEMPLATE.....	7
1. Data summary.....	8
2. FAIR data	8
2.1 Making data findable, including provisions for metadata	8
2.2 Making data accessible	9
2.3 Making data interoperable	10
2.4 Increase data re-use	11
3. Other research outputs	11
4. Allocation of resources	11
5. Data security	12
6. Ethics	12
7. Other issues	12

SECTION 1 - DMP GENERAL GUIDELINES

When tackling the DMP, be aware of the following contacts and resources that are relevant to the University of Padua (Unipd).

Contacts

Data Protection Officer (DPO) - see the dedicated web page (in Italian) [Responsabile Protezione Dati](#) | (in English) <https://www.unipd.it/en/privacy> - privacy@unipd.it

GDPR and DMP advisor - international.research@unipd.it

Intellectual Property (IP) advisor - trasferimento.tecnologia@unipd.it

Local ethical committees - At the University of Padua, there are several discipline-specific ethics committees. These committees were established by individual departments or clusters of departments. Only researchers affiliated with these departments are eligible to request an ethical opinion

- webpage (in Italian): [Comitato etico della ricerca psicologica](#) (*Ethics Committee for psychological research*) - contact: comitato.etico.area17@unipd.it
- webpage (in English): [Comitato etico Human Inspired Technology Research Center](#) (*Ethics Committee for human-technology interaction research*) - contact: ethical.hit@unipd.it
- Comitato etico Dipartimento di Scienze Biomediche (webpage not available) (*Ethics Committee of the Department of Biomedical Sciences*) - contact: comitatoetico.biomed@unipd.it
- webpage (in Italian): [Comitato etico Dipartimento di Scienze Politiche, Giuridiche e Studi Internazionali](#) (*Ethics Committee of the Department of Political Science, Law and International Studies*) – On line submission procedures, email contact not available
- Comitato etico Dipartimento di Territorio e Sistemi Agro-Forestali (webpage not available) (*Ethics Committee of the Department of Land, Environment, Agriculture and Forestry*) - contact: comitatoetico.tesaf@unipd.it

There are two ethics committees for clinical research established at the [Padua University Hospital](#) (contact: prc.unitaricercaclinica@aopd.veneto.it) and at the [Veneto Institute of Oncology IOV-IRCCS](#) (contact: comitato.etico@iov.veneto.it). Only researchers of the University of Padua involved in clinical studies are eligible to request an ethical opinion from those committees.

Useful resources

Data policy – [Unipd Research Data Management Policy](#)

Open access policy - [Regolamento per l'Accesso Aperto alla produzione scientifica](#) (Regulations for Open Access to the scientific production of the University of Padova)

Local data repository - [Research Data Unipd](#) - [Phaidra](#)

All about publishing at Unipd - [The University Library System](#)

Consultancy service for Data Management / DMP - [Biocomputing UP lab.](#) - [Fees applied](#)

What is a Data Management Plan?

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 is it necessary and important

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 programme, are requested to submit a DMP as deliverable by month 6** ([Horizon Europe GA](#), article 17, annex 5, page 111) 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](#).

Before you start

Prior to writing 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](#).

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](#) - 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](#) - 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://www.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](#) - 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](#) community working together with final users.

SECTION 2 - THE DMP ANNOTATED TEMPLATE

Instructions in green are provided by EU Commission.

Suggestions in grey are provided by the International Research Office to help you in writing a comprehensive DMP, fulfilling the requirements of the European Commission: they **must not appear in the final version of the DMP document**.

*This annotated DMP is based on the [Horizon Europe DMP template](#) (V1.1 – 01.04.2022) and summarises in 6 sections which main aspects must be addressed in a DMP document. **You are required to provide detailed answers only to the questions that are relevant for your project.***

Cover page

We suggest creating a cover page including the following information

PROJECT NAME



EU flag




Project logo

PROJECT	
Project number:	[project number]
Project acronym:	[acronym]
Project name:	[project title]

DATA MANAGEMENT PLAN	
Date:	[dd/mm/yyyy]
Version:	[DMP version]

(To be filled in and uploaded as deliverable in the Portal Grant Management System, at the due date foreseen in the system and regularly updated).

 *The template is recommended but not mandatory. If you do not use it, please make however sure that you comply with the research data management requirements under Article 17 of the Grant Agreement.*

1. Data summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

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

What types and formats of data will the project generate or re-use?

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

[List of open formats from Wikipedia](#)

[Recommended formats from the UK Data Service](#)

[Recommended formats from the Data Archiving and Network Services](#)

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

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.

Go through the various steps of your research and highlight expected outcomes and research outputs. This process will help project officers/reviewers in focusing on the data that will be involved in your research project, and in particular which datasets will be needed, generated, collected, reused or analysed to achieve the objectives of your project.

What is the expected size of the data that you intend to generate or re-use?

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

What is the origin/provenance of the data, either generated or re-used?

Data provenance is the documentation of where data comes from and the methods by which have been produced. Data provenance provides essential information for determining data quality and facilitating reproducibility and reliability of data. Provenance can be recorded in a single README text file that describes the data collection and processing methods.

To whom might your data be useful ('data utility'), outside your project?

The term “data utility” is used to describe the value of data. Describe potential target groups and/or relevant data users that will benefit from your datasets.

2. FAIR data

2.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Dataset should be identifiable and locatable by means of a standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers). A persistent identifier is a stable link where data can be directly accessed.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Store raw data it is not enough. Metadata should be used to properly describe the data making possible to identify, categorize and search large data sources efficiently. Metadata can describe physical items as well as digital items. 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, in standardised, structured, machine-readable content .etc. Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. For more information on disciplinary metadata standard catalogue see:

[Dublin core](#)

[Digital Curation Center \(DCC\)](#)

[Research Data Alliance \(RDA\) Metadata Directory](#)

[FAIRsharing Metadata standard](#)

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential 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). Following documents could be useful on how to identify relevant keywords.

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009768>

<https://library.hkust.edu.hk/sc/metadata-dataset-discovery/>

<https://www.nature.com/articles/s41597-020-0569-5>

Will metadata be offered in such a way that it can be harvested and indexed?

2.2 Making data accessible

Repository:

Will the data be deposited in a trusted repository?

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](#)

Zenodo (not-for-profit, hosted by CERN, all domains): <https://zenodo.org>

EUDAT Collaborative Data Infrastructure (large number of general and discipline-specific repositories): <https://eudat.eu/eudat-cdi>

Unipd Local data repository - [Research Data Unipd, PHAIDRA](#)

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant 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). 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.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

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

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Describe which metadata will be openly accessible even in case of embargo or restricted access to dataset.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

2.3 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Do the datasets produced adhere 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?

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

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? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include qualified references¹ to other data (e.g. other data from your project, or datasets from previous research)?

¹ A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

2.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement? Choose the licence that better fit your needs using these tools:

[Creative Commons license chooser \(beta\)](#)

[UFAL license selector](#)

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

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](#)

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?

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](#)

[Data management costing tool and checklist from the UK Data Archive](#)

[Calculate costs of data management from Data Stewardship Wizard](#)

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

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.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

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 or will be in place for data security (including data recovery as well as secure storage/archiving 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).

Will the data be safely stored in trusted repositories for long term preservation and curation?

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

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.

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

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?