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, 1 April 2022)

V4.3 18.09.2024

UNIVERSITÀ DEGLI STUDI DI PADOVA

HR EXCELLENCE IN RESEARCH

TABLE OF CONTENTS

INTRODUCTION

This document is intended to support University of Padua (Unipd) researchers in writing a Data Management Plan (DMP), ensuring that data produced throughout the research lifecycle are Findable, Accessible, Interoperable, and Reusable (FAIR).

Researchers are encouraged to:

- carefully read the introductory guidelines in Section 1 before approaching the DMP document
- fill out the annotated template in Section 2 as a draft of their DMP

SECTION 1 - DMP GUIDELINES

What is a Data Management Plan?

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

Why is it necessary and important?

DMP is a tool to support researchers to value the data they generate 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 principles. Sharing data reduces data duplication and increases reproducibility and transparency of 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.

Additionally, **some open science practices are mandatory per specific work programmes or call conditions, e.g., projects funded under the Horizon Europe programme, are requested to submit a DMP as a 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.

What FAIR stands for

FAIR is an acronym that stands for Findable, Accessible, Interoperable, Reusable.
In simple terms, FAIR principles are a rule of thumb on how research data can be made secure, findable and exchangeable.
Despite being in the air for more than a decade, it is commonly recognised that the publication in which FAIR principles were thoroughly outlined for the first time is the article "The FAIR Guiding Principles for scientific data management and stewardship" by Marc D. Wilkinson et al. 2016, where

the authors intended to provide guidelines on how to improve the Findability, Accessibility, Interoperability, and Reuse of data assets.

What FAIR is not

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, in order to be consistent 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 difficult to be retrieved or not compliant with any standard.
- 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 alone **do not ensure data intrinsic quality or compliance with ethical standards** (e.g., fabricated or manipulated can be totally in agreement with the FAIR principles).

Online tools

A range of on-line tools and services are available 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 based on DMP templates created by the funders (or based on funders' guidelines). Once a funder template is selected, 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 are properly addressed, the final document downloaded from DMP Online can 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 in 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 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.

Useful resources at Unipd

- **Data management**
  University of Padua, Data policy – Unipd Research Data Management Policy
  University of Padua, Data repositories - Research Data Unipd - Phaidra
  Consultancy service for Data Management / DMP - Biocomputing UP lab. - Fees applied
- **Publications**
  University of Padua, Open access policy - Regolamento per l'Accesso Aperto alla produzione scientifica
  All about publishing at Unipd - The University Library System

Contacts at Unipd

GDPR and Data Protection Officer (DPO) - Web ITA ENG - privacy@unipd.it
DMP advisor - international.research@unipd.it
Intellectual Property (IP) advisor - trasferimento.tecnologia@unipd.it
Ethical committees
Discipline-specific ethics committees at Unipd are established at department level (or for groups of departments).
Only researchers affiliated to such department or department group are entitled to request support:

- *Ethics Committee for psychological research* - Web ITA - comitato.etico.area17@unipd.it
- *Ethics Committee for human-technology interaction research* - Web ENG - ethical.hit@unipd.it
- *Ethics Committee of the Department of Biomedical Sciences* - comitatoetico.biomed@unipd.it
- *Ethics Committee of the Department of Political Science, Law and International Studies* - Web ITA (a submission form is accessible on the website)
- *Ethics Committee of the Department of Land, Environment, Agriculture and Forestry* - comitatoetico.tesaf@unipd.it

Researchers at the University of Padua involved in clinical studies are eligible to request an ethical feedback to:

- Ethics committees for clinical research at the Padua University Hospital - Web ITA - prc.unitaricercaclinica@aopd.veneto.it
- Ethics committees for clinical research at the Veneto Institute of Oncology IOV-IRCCS - Web ITA - comitato.etico@iov.veneto.it

SECTION 2 - THE DMP ANNOTATED TEMPLATE

This annotated DMP is based on the Horizon Europe DMP template (V1.1 – 01.04.2022) and summarises in 6 sections the aspects that must be addressed in a DMP document.
Only sections relevant/applicable to your project have to be filled out.

The template has to be filled out and uploaded as deliverable in the Portal Grant Management System, at the expected due deadline.

Using the template is recommended but not mandatory. If you choose not to use it, please ensure that you still meet the research data management requirements outlined in Article 17 of the Grant Agreement.

COLOURS LEGENDA

- *Instructions in green (like sentence below) are provided by the EU Commission.*

  Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus est libero, feugiat ac metus a, faucibus aliquet tortor. Nam et mauris quis turpis lobortis auctor. Fusce dictum ultrices diam, sit amet pharetra quam faucibus sit amet. Vivamus aliquam leo dapibus hendrerit consequat. Phasellus quis sem ante.

- Suggestions in grey *(like sentence below)* are provided by the International Research Office in order to guide 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**.

  Cras vel ligula vel metus egestas sagittis a at lorem. Phasellus euismod justo in arcu efficitur elementum. Quisque volutpat justo at purus dignissim porttitor. In sollicitudin dictum ipsum, at luctus erat finibus sit amet. Nunc dapibus vestibulum lectus, vitae elementum diam dignissim sit amet.

- Orange boxes *(like box below)* provide examples and in-depth information for the 3 ERC domains (PE - Physical Sciences & Engineering, SH - Social Sciences & Humanities, LS - Life Sciences)

Lorem ipsum dolor sit amet

Sed auctor nunc eu nisl pulvinar elementum
Donec laoreet condimentum sapien in feugiat. Nulla cursus tempor egestas. Quisque efficitur erat mollis mauris euismod condimentum. Sed et erat vitae turpis scelerisque aliquet. Nunc metus turpis, accumsan vel vulputate et, tincidunt et nisi. Nunc vel iaculis nisl, id pulvinar risus. Phasellus leo nisl, hendrerit eget diam et, blandit porta arcu. In in risus in arcu vestibulum egestas. Donec quis efficitur turpis. Sed gravida tincidunt nunc ac posuere.

Cover page

For all deliverables, we suggest creating a cover page including the following information.

PROJECT ACRONYM/NAME

Project emblem/logo

PROJECT DATA	
Grant agreement number:	*[grant agreement number]*
Project acronym:	*[acronym]*
Project name:	*[project title]*

DATA MANAGEMENT PLAN	
Deliverable ID:	*D1.2*
Due date:	*[dd/mm/yyyy]*
Submission date:	*[dd/mm/yyyy]*
Version:	*[DMP version]*
Dissemination level:	*(Public, Sensitive, EU restricted, EU confidential, EU secret)*
Authors:	*Dr. John Smith, Mr.Juan Pérez, Prof. Mario Rossi, Dr.Jane Doe*

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 999999.

The above sentence may be **different depending on the action funding the specific project**. When in doubt, ask the project officer for the official statement applicable to the specific action.

HISTORY OF CHANGES		
Version	**Submission date**	**Change**
1.0	DD.MM.YYYY	Initial version.
1.1	DD.MM.YYYY	Updated section on wind speed data provenance.
1.2	DD.MM.YYYY	Updated anonymization policies on human gender-related data.

1. Data summary

Introduce your research project in order to allow people who are not directly involved in the study (e.g. project officers, reviewers) to have an overview of the focus and purpose of your research action. While writing this paragraph, take some time to review the project proposal. Go through the steps of your proposal, identify the datasets/research assets that will be involved in your research project and dissect how they are linked to expected outcomes, milestones and deliverables. This process will allow you to generate an exhaustive list of the main datasets that will be needed, generated, collected, reused or analysed, to address a specific scientific question along your project. With this shortlist of research assets in your hands it should be much easier to address the topics listed in the DMP template.

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

List the research datasets generated in the previous step and map them to the corresponding Work Packages (WPs)/Deliverables/ Tasks

1.2 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 research assets will be reused and how (if any).

Examples of data summary table

The following table addresses questions 1.1, 1.2.

All DMP relevant research assets PROJECTacronim is expected to produce/re-use can be tracked down to work packages and deliverables as listed in Table 1.

Table 1.

WP No.	Deliverable No.	Lead Beneficiary	Deliverable Title	Research asset name	Research asset ID*	Generated/ Reused
WP1	D1.1	Unipd	Report on transcriptomic and proteomic data in healthy and dystrophic muscles.	Datasets of transcriptomic data in healthy and dystrophic muscles.	D1	Generated
				Datasets of proteomic data in healthy and dystrophic muscles.	D2	Reused
WP2	D2.3	KU Leuven	Report on youth attitude toward Kant's critical philosophy.	Dataset of interviews administered to teenagers and young adults.	D3	Generated

WP3	D3.4	Unipd	SatBuild automatic pipeline.	Software tool for the automatic recognition of buildings from satellite images.	S1	Generated
WP4	D4.2	UGraz	Report on field campaign on Graz palaeolithic settlements.	Virtual reality models dataset of Palaeolithic archaeological sites in Graz (spring campaign 2025).	D4	Generated

*D: dataset/database, S: software/app, Do: documents/datasets for internal use, P: publication
The line above offers just a brief, non-comprehensive list of examples for datasets/research assets categories.

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

- What types of data is your dataset 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

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

Data provenance is the documentation of where data come from and the methods through which they 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.

Examples of data formats and provenance table

The following table addresses questions 1.3 , 1.4.
All DMP relevant research assets PROJECTacronim is expected to produce/re-use will be compliant with formats and provenance tracking policies summarised in Table 2.

Table 2.

Research asset ID	Research asset Format	Data provenance
D1	.gtf	Standard Operating Procedures will be outlined via provision of a SOP.txt file in agreement with the Saint Mary's University recommendations. Details about the specific experimental conditions will be

		outlined via provision of a README.txt, file in agreement with MIAME guidelines
D2	.xlxs	Dataset reused from Morelli et al. 2022. DOI 10.5281/zenodo.7463257
D3	.xlxs	Details and relevant notes on the survey administration procedures will be outlined by the provision of a README.txt file
S1	.exe, .apk	Provided via GITHub documentation in agreement with Lee BD. Ten simple rules for documenting scientific software. PLoS Comput Biol. 2018 Dec 20;14(12):e1006561. DOI: 10.1371/journal.pcbi.1006561.
D4	.vrml	Virtual Reality Models generations procedures and relevant notes will be outlined by the provision of README.txt file

1.5 What is the expected size of the data that you intend to generate or re-use?
e.g., MB, GB, TB, PB

Examples of data size table

The following table addresses question 1.5.
All DMP relevant research assets PROJECTacronim is expected to produce/re-use will have their size outlined in Table 3.

Table 3.

	Research asset ID	Research asset Size
	D1	< 500 MB
	D2	< 500 MB
	D3	< 100 MB
	S1	< 50 MB
	D4	40 GB (< 10 GB for modelling each archaeological site)

1.6 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?

Datasets should be identifiable and locatable by means of a stable identification mechanism such as persistent and unique identifiers like the Digital Object Identifiers (DOI). Persistent identifiers, such as DOIs, provide a stable connection between the resource and its location, even if the resource is moved to a different hosting service/provider.

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.

Metadata are additional information that allow a better characterization of the research asset. Metadata could have different focuses. Metadata supporting findability are often described as "descriptive metadata" and they are usually defined as the descriptive information about a resource, used for discovery and identification of the specific asset among large datasets.

Descriptive metadata usually include information as title, abstract, author, and keywords.

The following documents could be useful in clarifying metadata definition and usage.

Ten simple rules for improving research data discovery
Metadata: A Key to Dataset Discovery
A dataset describing data discovery and reuse practices in research

Will keyword search be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Some disciplines have specific metadata standards, while in other cases, general metadata standards may be used to describe the research asset.

For more information on comprehensive/discipline-specific metadata standards see:

Dublin Core Metadata Initiative
Digital Curation Center (DCC)
Research Data Alliance (RDA) Metadata Standards Catalog
FAIRsharing Metadata standard

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

The ability to harvest and index metadata depends significantly on the policies and functionalities of the data repository. Please review the repository fact sheets and specifications when addressing this topic in your DMP.

Examples of findability section

Findability applied to general purpose repository/general purpose metadata schema

All PROJECTacronym DMP-relevant research assets (datasets, publications, main and supplementary data, software) will be made available in the Zenodo repository. To ensure unique identification, a Digital Object Identifier (DOI) will be automatically assigned by the repository at the time of data record creation.

A minimum set of metadata, including the DOI, is mandatory for dataset registration. The Zenodo metadata scheme complies with DataCite's Metadata Schema mandatory and recommended terms, with further enrichment fields available to meet specific field needs.

In addition to Zenodo's minimal mandatory metadata, PROJECTacronym will require the inclusion of the following mandatory additional information:

- the terms “European Union (EU)” and “Horizon Europe”
- the name of the action, acronym, and grant agreement number
- the research asset type (publication, dataset, software, etc.)
- the publication date (ISO 8601 format, YYYY-MM-DD)
- the creators/authors of the deposition

Findability applied to discipline specific repository/metadata schema - Example A
The dataset of human chemical compound biomonitoring experiments will be stored in the discipline-specific HBM4EU repository. The IPCheM metadata schema for human biomonitoring will be used to describe the data collected during the implementation of PROJECTacronym. A metadata file will be attached to both raw and processed data stored in the HBM4EU repository.

This metadata reference is structured into four sheets:

- a standard set of general information about the dataset
- a standard set of information about the conditions of data access and use, as well as the contact point of the responsible organization
- a standard set of information on sampling and analytical details for each substance recorded in the dataset
- a standard set of information about the study population

Findability applied to discipline specific repository/metadata schema - Example B:
The dataset of virtual reality models of Palaeolithic archaeological sites in Graz generated during the spring champain 2025, will be stored in the discipline-specific Archeology Data Service (ADS) repository. The ADS_3D metadata template for virtual reality models will be used to describe the archaeological sites identified during the field campaigns along the PROJECTacronym action. A metadata file will be attached to data made available in the ADS repository.

This metadata reference is structured into 8 sheets:

- a ReadMe sheet describing the metadata reference
- a standard set of general information about the dataset
- a standard set of information about the laser scanner capture instrument and operations
- a standard set of information the camera calibration details and operations
- a standard set of information on control operations
- a standard set of information about registration/alignment operations
- a standard set of information about the model objects to document the final 3D objects

- a standard set of information about the orthophoto objects

2.2 Making data accessible

Repository:

Will the data be deposited in a trusted repository?

The choice of the deposition repository impacts multiple layers of the FAIR principles implementation. The technological features of the repository will be key in determining also how research assets can be accessed.

The characteristics of the repository will indeed influence, among other things, how user identification is managed, how different privileges are granted or denied to specific users or groups, and how data can be accessed programmatically, for example through a RESTful API interface.

Repositories could be classified in:

Discipline-specific repositories
Find a discipline-specific repository at re3data.org

General-purpose repositories
Zenodo (not-for-profit, hosted by CERN, all domains): Zenodo website
EUDAT Collaborative Data Infrastructure (large number of general and discipline-specific repositories): EUDAT website
Unipd Local data repositories - 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 enclave.

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 a dataset.

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

The requirement for data retention in Horizon 2020 is not explicitly detailed, but it is considered a good practice to apply a data retention policy that requires research assets to be kept for at least 5 years after the end of the project. In practice, data is often preserved for 5 to 10 years but some datasets may be kept indefinitely if they are deemed of long-term value.

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)?

Examples of accessibility section

For illustrative purposes, in this section, we distinguish accessibility of raw data and processed data.

Case scenario of raw data hosted on project's resources and processed data hosted on on-line repository

All internal PROJECTacronym raw data will be hosted in the PROJECTacronym data repository.
The PROJECTacronym data repository is a virtual cloud space hosted at the IT premises of the Department of Mathematics at the University of Padua, and so located within the EU territory.
Access to the PROJECTacronym raw data repository will be closed and limited to individual researchers involved in the PROJECTacronym action.
Access to the repository will be granted to PROJECTacronym partners via SSO authentication scheme.
Data access will be granted by the project coordinator to researchers involved in the project.
Data access will have to be requested via an authorization form.

Processed data will be made public in the Zenodo repository, after publication or delivery to the Research Executive agency.

Case scenario of raw and processed data hosted in an on-line repository

All PROJECTacronym raw and processed data will be hosted in the Zenodo data repository.
Access to the PROJECTacronym raw data will be closed and limited to individual researchers involved in the PROJECTacronym action.
Zenodo allows users to login via local username/password, via ORCID or via GitHub credentials.
In all cases, Zenodo has a local account and links external identities to this local account.
This allows a user to log into the same account from both GitHub, ORCID or with a local password.
Access to the processed data from PROJECTacronym will be open.
The metadata of the processed data will link each specific research asset to the corresponding raw data used for its generation. No authorization will be required to access the processed data.

Summary table
All DMP relevant research assets PROJECTacronim is expected to produce will be made available as outlined in Table 4.

Table 4.

Research asset ID	Availability
D1	Restricted to PROJECTacronym consortium members.
D2	Open.
D3	Open after sensitive data anonymization/pseudo-anonymization.
S1	Embargoed for patenting. Expected embargo duration: 1 year.
D4	Open.

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

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 references[1] to other data (e.g. other data from your project, or datasets from previous research)?

Examples of interoperability section

Case scenario where standards formats are available - Example A
Dataset 1, storing information about traffic flow, will be in agreement with Datex II, the main standard format for this kind of data.
Datex II is a standard for data exchange that facilitates the sharing of traffic information among traffic management centers, service providers, operators, and media partners. It covers data on traffic incidents, ongoing road works, and other events related to traffic. The information is formatted in XML and modelled with UML. The standard is developed by the technical committee

[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/)*

for Intelligent Transport Systems (CEN/TC 278) of the European Committee for Standardization.

Case scenario where standards formats are available - Example B
Patients phenotype data, stored in Dataset 2, will be described using the Human Phenotype Ontology (HPO).
The HPO is the community standard to provide a structured and controlled vocabulary for the phenotypic features encountered in human diseases.
The HPO is a flagship product of the Monarch Initiative, an NIH-supported international consortium dedicated to semantic integration of biomedical and model organism data with the ultimate goal of improving biomedical research. The HPO, as a part of the Monarch Initiative, is a central component of the Global Alliance for Genomics and Health (GA4GH) strategic roadmap.

Case scenario where standards formats are available - Example C
The dataset of findings from the Paleolithic archaeological sites in Graz, collected during the spring campaign of 2025, will be described using Getty Vocabularies. These vocabularies serve as a community standard for architecture, decorative arts, archival materials, visual surrogates, art conservation, and other cultural works. Aligned with international standards for structured and controlled vocabularies, Getty Vocabularies aims at being multilingual, multicultural, and inclusive, emphasizing diversity, equity, unbiased and antiracist language, and accessibility. They are available as Linked Open Data, XML, relational tables, and through APIs. Additionally, they are freely accessible under the Open Data Commons Attribution License (ODC-By) 1.0.

Case scenario where standards formats are NOT available
As today, no community shared policies or standards/vocabularies/formats are available to describe mitochondrial respiration experiments data, stored in Dataset 3.
For this kind of experiments, data will be stored in simple text-based tabular format as CSV/TSV files supported by any open-source software.
Template tabular file storing mitochondrial respiration experiments data is attached to the present version of the DMP and it will be shared among all the partners of the PROJECTacronym consortia.

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 fits 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 discovering 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.

Examples of reusability section

Data management policies concerning data accessibility and reusability are closely related. Therefore, it is advisable to combine summary Tables 4 and 5 to enhance the document's readability.

All DMP relevant research assets PROJECTacronim is expected to generate will be licenced as outlined in Table 5.
Information about licensing will be mandatory integrated into the specific dataset metadata.

Table 5.

Research asset ID	Reusability licence
D1	Not applicable - raw sensitive data will not be made available for reuse
D2	CC 0
D3	CC 0
S1	CC BY
D4	CC BY

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 for monitoring activity to avoid risk of disclosure of information and countermeasures in case of information 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?

Examples of ethics section if human sensitive data are involved

Task 4.5 will involve the collection of human blood samples in order to compare gender concentration of vitamin K among the population of interest.
Participants will be thoroughly informed about the purpose and scope of the analysis, including any potential risks and benefits, as well as the intended use of their samples for research purposes. The consent forms will clearly detail this information, highlighting the voluntary nature of participation and respecting participant autonomy. Before giving their consent, participants will have the chance to ask questions and seek clarification. All procedures will comply with applicable data protection regulations, ensuring confidentiality and the secure handling of sensitive information at every stage of the research.

Documentation provided to study subjects.

In agreement with the Council for International Organizations of Medical Sciences recommendations, documentation provided to subjects involved in the study will mandatorily include:

- **Information sheet**
- **Informed consent form**

Further details towards documentation provided to study subjects will be overseen in coordination with the Unipd Department of Biomedical Science - Ethical Committee and the ethical officer.

Additional assessment of PROJECTacronym ethical issues is addressed via Ethics Summary Report, in coordination with project ethical officer.

Sensitive data anonymization/pseudo-anonymization.

Sensitive data will undergo a mandatory anonymization/pseudo-anonymization process.

- Anonymization
  - only data needed to address the scientific purpose will be collected. IE: if possible it will be avoided to request full names and other identifiable information in the case they will not be used.
  - subjects' direct identifiers will be removed via eliminating or obscuring the specific part in the survey's hard/digital copies, e.g., names, Social Security numbers, phone numbers, and email addresses.
  - anonymization will be performed during transcription or initial write-up e.g.: do not transfer Name/Surname/Address while transcribing data from survey's hard copies to electronic spreadsheets.
- Pseudo-anonymization
  - identifying information will be replaced with pseudonyms or unique identifiers.
  - sensitive data will be replaced with tokens or unique codes (Tokenization).
  - pseudo-anonymized data and documents matching pseudonyms with original IDs will be stored in a trusted repository and no access will be authorised to personnel involved in the research activities except project coordinator.

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)?

In this section, the following Unipd policy documents can be cited:

- Policy sulla gestione dei dati di ricerca, a light policy document on data management.
- Codice di integrità della ricerca dell'Università degli Studi di Padova, a comprehensive document about research integrity. Article 6 focuses on general principles for research data management.