PROJECT: XXX

Deliverable:

DX.X

Title:

Data Management Plan

Date: _____

Version: _____

Website: https://www. _____

Status

☐ Final	
☐ In Progress. Please explain:	☐ Iterative Process – This year's results have been 100% achieved. ☐ Delay – This year's results were not fully achieved.

Tracking Changes

Initial Version	XX.XX.202X	For review, provided by XXX
Deliverable VX.X for upload	XX.XX.202X	Reviewed by XXX

Level of Dissemination

☐ Confidential
☐ Public

Author(s)

	Partner name	Name of the author
Main Author		

Executive summary

This data management plan establishes guidelines to manage all data collected, created, processed and published in the **[funding organization]** project **[project name]**. It also regulates the mechanisms to be used at the end of the project to share and preserve the project's data. To this end, the data management plan outlines all intended activities regarding data management and data governance. By its nature, the document is a living document, intended to be adapted and modified as required during the course of the project.

Action Number: XX
Action Acronym: XX
Action title: XX
Date: XX.XX.202X
DMP version: X
Contact: [name], [institution], [email]

Suggested citation

Last, first, etc. (202X). Data Management Plan of the Project [name], 202X (Deliverable No. DX.X, public). Retrieved from the [project] website: [link].

Table of contents

1 Introduction

1.1 Purpose of the DMP

The data management plan (DMP) organizes data handling in the [project name] project. The project is a data-driven, complex project requiring a common data language to organize

- experimental data,
- analysis data,
- business models, and
- communication / dissemination / exploitation of project results.

Data management ensures a **smooth and timely workflow** inside and outside of the project. The DMP guarantees **data quality** by setting up processes and standards (e.g., for documenting the provenance of data, methods of processing of data, etc.). The DMP also provides the **agreements among partners on the sharing and reusing of data** (internally within the consortium and externally).

A data management plan (DMP) is a living document, establishing an agreement between partners about data handling, data governance, and data practices. It is an integral part of project management. By its nature, it will be regularly updated.

1.2 Role of the DMP with respect to other deliverables

Data governance regulated in the project's DMP is coherent with data handling provisions in project deliverables and the Grant Agreement regulated in:
- ARTICLE XX — CONFIDENTIALITY AND SECURITY,
- ARTICLE XX — DATA PROTECTION,
- ARTICLE XX — INTELLECTUAL PROPERTY RIGHTS,
- ARTICLE XX — RECORD-KEEPING.
- etc.

1.3 Revision strategy for the DMP

The DMP will be updated as needed. Decisions concerning updating are made at the regular meetings of the project management board and annual partner meetings. The need for review is also synchronized with the milestones and timing of deliverables making sure that the data flow supports the tasks foreseen and proper agreement about data use between partners is documented in this document well ahead of time.

2 Overview of [Project Name] data

2.1 Administrative information

This document is the data management plan for the Horizon Europe project '
- Action Acronym: XX

- Action title: **XX**
- Project id (via CORDIS): **XX**, [https://cordis.europa.eu/project/id/XX]
- Funder id: **XX**, [link]
- Call for proposal: **XX**
- Funded under: **XX**

Parties responsible for ensuring the quality and fitness of data related to **[project name]** are all partners in **[project name]**. Specific responsibilities may be assigned in this document. All responsibilities described herein are set in accordance with the regulations outlined in the Grant agreement (GA) and the consortium agreement (CA). In case of disagreement, GA and CA regulations take precedence. This data management plan complements and details other deliverables such as D7.3 "GDPR Compliance Report", D8.10 "Ethics Requirements", and D8.1 "Collaborative Portal".

Information on data governance in support of implementing project activities in time and at high quality are given in the following sections. In particular, this document will provide related information and identify the tools needed to validate the results of the project.

[project name] has assigned **Dr. Name, institution**, as data steward (DS) and data protection officer to coordinate all activities regarding the fitness and quality of all data, both as content as well as metadata.

ORCIDs of involved researchers are the following[1]:

Partner	Name of staff	ORCID
Example: XXX	*XXX XXX*	*0000-000X-XXXX-XXXX*

Third parties[2]:

Partner	Name of staff	ORCID
Example: INDITEX		

2.2 Data sets in the project

The figure below gives an overview on data flows in the project. The color code follows the schema in the excels, collecting an overview on the answers to the DMP questionnaire.

 Fig.: Overview on data flows in the project (example)

Data flow diagrams help to get an overview of how data flows between work packages and partners. It is useful to construct such an overview. See for example https://en.wikipedia.org/wiki/Data-flow_diagram and for templates check: https://commons.wikimedia.org/wiki/Category:Data_flow_diagrams.

---

[1] Note that this table may be moved to Appendix if partners request the data being confidential and only available for internal purposes.

[2] Note that this table may be moved to Appendix if partners request the data being confidential and only available for internal purposes.

Data in [project name] are categorized as follows:

Id	Name & origin	Type of data	Format of data	Purpose (WP, Task)	Partner responsible
001					
002					
xxx					

All data sets exchanged between partners and work-packages should come with metadata detailing all relevant information. This metadata can be listed in a separate document as text. In the course of the project, metadata will be integrated to the datasets in a machine-readable format, e.g. JSON-LD or CSV-LD. The convention for the naming of datasets is the following: Id_TaskId_Topic_PartnerId_Date_Nr.ext, where the extension is indicating the format of the dataset. For most of the datasets before FAIRification, this will be xls or xlsx. For metadata information to be FAIRified, the dataset should have the same filename, but use the extension .txt for plain text.

2.3 Data for re-use and sharing

Data for sharing and reuse outside of [project name] are the following:

Id	Name & origin	Format of data	Deliverable	Repository & licensing	Partner responsible
001					
002					
xxx					

All datasets will be assigned a unique identifier and name. The format, the relation to deliverables, the storage at a repository, connected licensing information, as well as the responsible partner will be recorded. The preferred repository is zenodo.org. Zenodo is free of charge, supports datasets up to 50 GB, and has a rather low workload for uploading datasets. It supports the reporting guidelines of the European Commission. All datasets which will be disseminated publicly in the course of the project will follow FAIR principles. For details, see the information below. All datasets will be machine-actionable. Preferred formats for datasets are e.g. JSON-LD, CSV-LD, or RDF. Partner xxx will support the conversion process of the original datasets into these preferred formats. Afterwards it is the responsibility of the partner who owns the data to upload it to a repository.

3. Curation of data

[project name] commits to the implementation of FAIR principles (Wilkinson et al. 2016) as much as feasible and as outlined in the following sections. These agreements within the

consortium are a result of a questionnaire and discussions with individual partners. They are in alignment with the Grant Agreement and decisions taken in the consortium. The questionnaire is detailed in the Appendix.

3.1 FAIRification effort & FAIR workflow

FAIR principles are implemented in [project name], focussing on rich and standardized metadata agreements, standardized vocabularies, and agreements about the handling of licensing and IPR):

- Addressing FAIR - On metadata & vocabularies
- Addressing FAIR - On licensing
- Addressing FAIR - On IPR

Describe the process on how to approach FAIRification and add a picture for the workflow to be followed. Below general principles of the FAIRification workflow will be followed. The figure is subject to regular updates.

Fig. Overview on the FAIRification workflow (to be updated with project specifics). Source of picture: EERAdata platform.
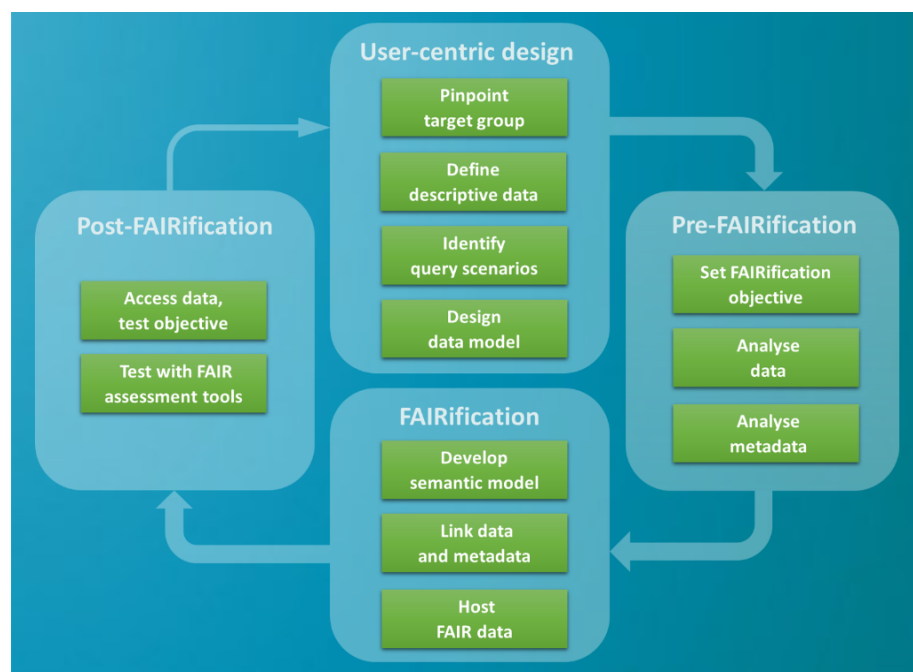
3.2 Addressing FAIR - On metadata & vocabularies

Metadata is information added to data to increase functionality, supporting findability, accessibility, interoperability, and reusability of [project name]`s project data. Metadata are used to capture the context of the data, split into descriptive, structural, administrative, reference, statistical, and legal. For a given document, metadata can apply to the document as a whole, to parts of the document, e.g. describing the content in columns or to single entries, e.g. describing the source of a particular value. Rich and comprehensive metadata

play an essential role in making data FAIR. The application of standardized vocabulary and formats is important.

As a minimum, [project name] provides core information according to the DUBLIN core standard for all digital objects as specified in Chapter 1 (see project internal list of data sets and data ids). For datasets to be published open for reuse, standards by the external repository owners where the data is stored will be followed (e.g., Zenodo, dataverse). Dublin Core Elements are:

DC Element (term name)	URI of DC Element
contributor	http://purl.org/dc/elements/1.1/contributor, see also Section 1 for researcher ORCIDS
coverage	http://purl.org/dc/elements/1.1/coverage
creator	http://purl.org/dc/elements/1.1/creator see also Section 1 for researcher ORCIDS
date	http://purl.org/dc/elements/1.1/date
description	http://purl.org/dc/elements/1.1/description
format	http://purl.org/dc/elements/1.1/format
identifier	http://purl.org/dc/elements/1.1/identifier
language	http://purl.org/dc/elements/1.1/language
publisher	http://purl.org/dc/elements/1.1/publisher
relation	http://purl.org/dc/elements/1.1/relation
rights	http://purl.org/dc/elements/1.1/rights
source	http://purl.org/dc/elements/1.1/source
subject	http://purl.org/dc/elements/1.1/subject
title	http://purl.org/dc/elements/1.1/title
type	http://purl.org/dc/elements/1.1/type

If applicable, the following metadata frameworks and controlled vocabularies/ontologies are applied in [project name].

- General descriptive metadata: Dublin core metadata elements (contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type) supplemented by keyword. See table above.
- specify as suitable for the field

For an overview of ontologies in field XXX, see the list compiled in LOV - Linked Open Vocabulary.

3.3 Addressing FAIR - On licensing

Agreements table on reuse rights <u>outside</u> of the project. See datasets & products identified in Section "Data for reuse and sharing".

Id	Name & origin	Repository link	License statement
001			
002			

As a default, all datasets connected to deliverables which have dissemination level PU (public) will be assigned a <u>Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)</u>. For datasets connected to deliverables with dissemination level SEN, an assessment for an appropriate license is down on an individual level. The default means that data reusers are free to share and adapt the data. However, the license requires to give appropriate credit to the creator (attribution). ShareAlike implies that any contributions made must be distributed under the same license as the original, e.g., when remixing, transforming, or building upon the material.

3.4 Addressing FAIR - On IPR

The figure below gives a visual overview on the level of confidentiality of [project name] data as a result from the questionnaire to the partners.
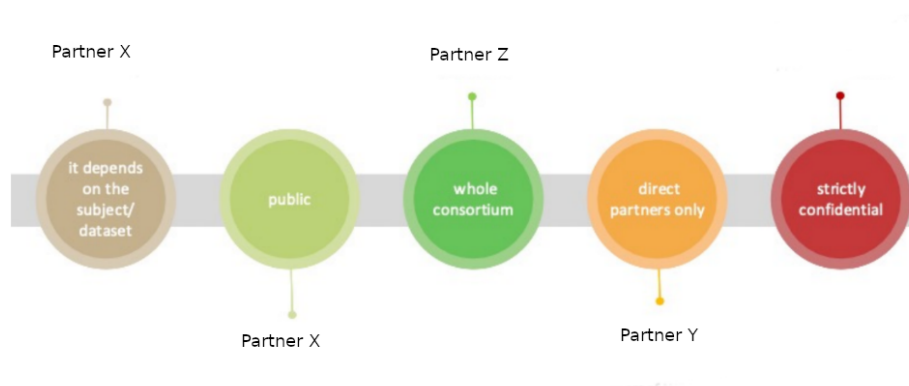
Internal agreements table on access rights within the project. Linked to table in Section "Datasets in the project". See also Grant Agreement and IPR regulations.

Id	Name & origin	Data provider	Data user	Access specification
001				
002				

4. Other research outputs (if applicable)

Placeholder for providing additional information on the following matter: 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.

5. Allocation of resources

In line with the Grant Agreement, all project data is stored at least for up to five years after the project's completion. For the datasets made available to the public, no publication fees arise, as Zenodo.org is free of charge. Publications in scientific journals target gold open access. The Grant Agreement regulates the procedure for handling publication fees.

The following staff is responsible for data governance at each partner[3]:

Partner	Name of staff	Email
Institution 1		
Institution 2		
Institution 3		
xxx		

6. Data security

Each partner is responsible to ensure the security of data under control by the partner. Dynergie as project coordinator is responsible for data security of the project data infrastructure. For details about the collaborative platform, see deliverable D8.1. Refer also to the list of contacts for responsible staff at each partner in the section before.

7. Ethics

[project name] commits to ensuring the ethical handling of data, specified and inline with Deliverable D8.10 "[project name]'s ethics requirements - initial review" and Grant Agreement (see Appendix on IPR).

---

[3] Note that in agreement with IPR, the table may be a confidential appendix to the DMP, depending on the decision of the partners about the publication of names and email addresses. The DMP as such is required to be open (dissemination level - PU).

8. Other issues (if applicable)

Placeholder: Whatever else needs to be handled.

9. References

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

qudt.org. Quantities, Units, Dimensions and Data Types Ontologies.Available at https://qudt.org/.

European Commission - Joint Research Centre - Institute for Environment and Sustainability: International Reference Life Cycle Data System (ILCD) Handbook - Nomenclature and other conventions. First edition 2010. EUR 24384 EN. Luxembourg. Publications Office of the European Union; 2010

Appendix: Questionnaire to the partners

This questionnaire is used to inquire data requirements from partners. It was distributed after the kickoff workshop and will be followed up in discussions with individual partners. The answers form an integral part of the data management plan, which is the primary tool for understanding what data inputs, methods, and outputs are going to be involved in the [project name] project.

Please either fill it out or pass it along to the person in your organization best suited to completing it. Please complete this questionnaire by the end of September so that we can move on to the next steps in the creation of the data management plan.

1. Please specify the [project name] partner institution you are answering these questions for:

2. Who is responsible for data governance in your institution (a person who will engage in the [project name] project)? Please provide name and email.

3. Do you have a deputy? Please provide name and email.

4. Please give the name of the dataset

5. Context of data

   Please describe the type and formats of data input (data output) that is needed to carry out the task together with a short explanation or title describing the context and characteristics of the corresponding dataset. What is the origin of the data? Is it raw data? Is it processed data? (If so, what process was applied or what software has been used?)

   Example:

Table of numeric data in excel with the following headers: Name of material - chemical characterization - amount in kg etc.

The dataset comprises cleaned data that are purchased from the material supplier Company XYZ.

6. Standardization of data

Is the data connected to any standard, e.g., industry standards such as ISO-Codes or standardized vocabulary etc. Please provide the name of the standard and a link to its documentation. E.g., the data aligns with ISO-Code XYZ. The dataset uses the standardized vocabulary from Provide an example, e.g.IUPAC Standards Online - IUPAC | International Union of Pure and Applied Chemistry. If the standardization information is contained in an internal document, please send this document to xxx@institute.

Example:

The dataset follows an internal vocabulary that has been sent to the WP6 coordinators.

7. Use of the data in [project name]

Is the dataset critical for a specific procedure or task in [project name]? When is the dataset needed (or when will it be available)? Will the data be updated?

Example:

The dataset is critical input to task Tx.x and will be needed in M24.

8. Confidentiality of data

Please use this field to briefly flag if any of the data are restricted for sharing with the public / within the consortium / among selected partners only / strictly confidential. Also, do you have a license specification for the dataset? E.g. CC-BY,..

Example: The dataset can be shared within the consortium. A license has not been assigned.

9. Do you have another dataset that you would like to submit information for?

If no, the questionnaire will end

If yes, you will fill out this section again for the next dataset