**[projectname]**

**[acronym]**

**Data Management Plan (DMP)**

|  |  |
| --- | --- |
| **Lead partner:** |  |
| **Version:** | 1.0 |
| **Status:** |  |
| **Dissemination level:** | PU: Public |
| **Document link:** |  |

|  |
| --- |
| **Deliverable abstract** |
| This deliverable is the initial Data Management Plan (DMP) for the [projectnameText] project, delivered in [M4]. It will be kept updated as a living document. The DMP addresses the relevant aspects of the management of data and other outputs produced by the project according to the principles outlined in the [section name] section. |

**COPYRIGHT NOTICE**

Ein Bild, das Text, ClipArt enthält.

Automatisch generierte BeschreibungThis work by parties of the [projectnameText] project is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The [projectnameText] project is funded by the European Union Horizon Europe programme under Grant Agreement No [grantid].

**DELIVERY SLIP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Name** | **Partner/Activity** | **Date** |
| From: |  |  |  |
| Moderated by: |  |  |  |
| Reviewed by: |  |  |  |
| Approved by: |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| **Issue** | **Date** | **Comment** | **Author** |
| 1.0 | [dd.mm.yyyy] |  |  |
|  |  |  |  |
|  |  |  |  |

**TERMINOLOGY**

<https://eosc-portal.eu/glossary>

|  |  |
| --- | --- |
| **Acronym** | **Definition** |
| DMP | Data Management Plan |
| CSV | Comma Separated Values |
| EOSC | European Open Science Cloud |
| FAIR | FAIR-Principles: Findable, Accessible, Interoperable, and Reusable |
| MB | Megabyte |
| PDF | Portable Document Format |
| WP | Work Package |

**CONTRIBUTORS**

|  |  |
| --- | --- |
| **Role** | **Persons** |
| Project Coordinator Principal Investigator | [coordinator] |
| Project Coordinator Principal Investigator | [contact] |
| Contributors | [contributors] |

Content

[1. Executive Summary 5](#_Toc156287338)

[1.1. Introduction 5](#_Toc156287339)

[1.2. Data Summary 5](#_Toc156287340)

[1.3. Methods and software used for data generation and reuse 5](#_Toc156287341)

[1.4. Foreseeable research uses and /or users: 6](#_Toc156287342)

[2. FAIR data 6](#_Toc156287343)

[2.1. Making data findable, including provisions for metadata 6](#_Toc156287344)

[2.2. Making data accessible 6](#_Toc156287345)

[2.3. Making data interoperable 7](#_Toc156287346)

[2.4. Increase data re-use 8](#_Toc156287347)

[3. Other research outputs 8](#_Toc156287348)

[4. Allocation of resources 9](#_Toc156287349)

[5. Data security 9](#_Toc156287350)

[5.1. Storage and backup facilities 10](#_Toc156287351)

[5.2. Data security and protection of sensitive data 10](#_Toc156287352)

[5.3. Long-term preservation and deletion of data 10](#_Toc156287353)

[6. Ethical and legal issues 11](#_Toc156287354)

[6.1. Personal data 11](#_Toc156287355)

[6.2. Intellectual property rights and ownership 11](#_Toc156287356)

[6.3. Ethical issues 11](#_Toc156287357)

[7. Other issues 11](#_Toc156287358)

# Executive Summary

## Introduction

A data management plan (DMP) is a structured document that keeps record of what research data is created or reused and what happens to that data during and after a project. It helps with planning the research process and defining responsibilities in a research project involving several researchers or institutions. For writing this DMP, we followed the [Horizon Europe DMP template](https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.tuwien.at%2Ffileadmin%2FAssets%2Fforschung%2FZentrum_Forschungsdatenmanagement%2Fdata-management-plan-template_HE_2021.docx&wdOrigin=BROWSELINK).

[datamanager] [datamanagerInfo]

## 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.
* What types and formats of data will the project generate or re-use?
* What is the purpose of the data generation or re-use and its relation to the objectives of the project?
* What is the expected size of the data that you intend to generate or re-use?
* What is the origin/provenance of the data, either generated or re-used?
* To whom might your data be useful ('data utility'), outside your project?

Produced datasets:

| **dataset ID** | **title** | **type** | **format** | **estimated volume** | **contains sensitive data** | **description** |
| --- | --- | --- | --- | --- | --- | --- |
| P1 | [datasetTable] | … | … | … | … | … |
| ID | [dataset1name] | [dataset1type] | [dataset1format] | [dataset1vol] | [dataset1sensitive] | [dataset1description] |

Reused datasets:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **dataset ID** | **title** | **source** | **rights (e.g. license)** | **contains sensitive data** | **description** |
| R1 | [reusedDatasetTable] | … | … | yes/no | … |
| R2 | … | … | … | … | … |

## Methods and software used for data generation and reuse

[datageneration]

## Foreseeable research uses and /or users

[targetaudience]

# FAIR data

## **Making data findable, including provisions for metadata**

* Will data be identified by a persistent identifier?
* 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.
* Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?
* Will metadata be offered in such a way that it can be harvested and indexed?

We will make our data findable by uploading it to a data repository that provides a persistent identifier, and adding relevant metadata. The repository will provide means for harvesting the metadata, including its machine-actionable representation.

[dataorganisation]

[metadata]

Additionally, we will provide common metadata such as title, description, or keywords when publishing data. In this case, we will follow the default template provided by the repository, such as Data Cite Metadata or Dublin Core.

As far as possible, we will use controlled vocabularies for our data to allow inter-disciplinary interoperability and machine-actionability.

## Making data accessible

**Repository:**

* Will the data be deposited in a trusted repository?
* 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 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?
* 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)?

We will make our data accessible by providing open access to data, wherever possible. In cases where open access is not possible, we will provide meaningful metadata plus contact information for access requests. Some datasets cannot be published and even need to be deleted at the end of the project. See section 5 for more details.

| **dataset ID** | **access conditions** | **restrictions / embargo reasons** | **estimated publication date** | **location for publication (repository)** | **PID** | **license** |
| --- | --- | --- | --- | --- | --- | --- |
| P1 | [datasetPublicationTable] |  |  |  |  |  |
| ID | [dataset1access] | [dataset1restriction] | [dataset1pubdate] | [dataset1repo] | [dataset1pid] | [dataset1license] |

Repository description:

[repoinformation]

[tools]

[restrictedAccessInfo]

## 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?
* 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 to other data (e.g. other data from your project, or datasets from previous research)?

We will make our data interoperable by providing and describing data in a way that is common within our domain. As far as possible, we will use controlled vocabularies for our data to allow inter-disciplinary interoperability and machine-actionability. We will provide good documentation for all our datasets.

## 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?
* 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.
* 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.

We will make our data reusable by adding metadata and comprehensive Readme files to all published datasets. The descriptions will include details on the methodology used, analytical, and procedural information. In case of publication, licenses for code and data will always be assigned and clearly marked. The digital research data obtained will be published Open Access under a Creative Commons CC BY license, provided that there are no data protection concerns.

[documentation]

[dataqualitycontrol]

# 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.

…

# 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.)?
* 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?
* 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)?

[costs]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **cost name** | **cost type** | **description** | **unit** | **value** |
| [costTable] | [costTable] |  | € | 250 |
| [cost1title] | [cost1type] | [cost1desc] | [cost1currency] | [cost1value] |
| **Estimated total costs** |  |  |  | **[costtotal]** |

Coverage of costs:

…

The data manager will direct the data management process overall, with [workPackageLeaders] being responsible for ensuring metadata production, day-to-day cross-checks, back-up and other quality control activities are maintained. The researchers will be responsible for routine supervision of the dataset development.

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

## Storage and backup facilities

[storageintro]

[storage]

## Data security and protection of sensitive data

We pay strict attention to compliance with the relevant institutional and national data protection policies.

[sensitivedata]

Access to data during research:

| **dataset ID** | **selected project members** | **all other project members** | **the public** |
| --- | --- | --- | --- |
| P1 | [datasetAccessTable] | [dataset1allaccess] | [dataset1publicaccess] |
| P2 | [dataset1selectedaccess] | writing | reading only |

All incidents will be handled individually by an incident response team that is maintaining the affected service.

## Long-term preservation and deletion of data

| **dataset ID** | **location for long-term storage** | **minimum retention period (≥ 10 years)** |
| --- | --- | --- |
| P1 | [datasetRepositoryTable] | 100 years |
| P2 | [dataset1repo] | [dataset1period] |

Overview of (unpublished) data that will be deleted:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dataset ID** | **kind/name of data** | **date of deletion** | **reason for deletion** | **responsible person** |
| P1 | [datasetDeleteTable] | [datasetDeleteTable] |  |  |
| ID | [dataset1delete] | [delete1date] | [delete1reason] | [delete1person] |

# **Ethical and legal issues**

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

## Personal data

[personaldata]

## Intellectual property rights and ownership

[legalrestriction]

## Ethical issues

[ethicalissues]

# 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)?