

Deliverable 1.2

Data Management Plan (DMP)

Authors	Tomasz Miksa (TU Wien)
Contributors	Ute Gunsenheimer (EOSC Association), Iiris Liinamaa (CSC), Roosa Ripatti (CSC), Gloria Soriano (EOSC Association)
Work Package	WP1 Project Coordination
Reviewed by	Ute Gunsenheimer (EOSC Association), Andreas Rauber (TU Wien), Chris De Loof (Belnet), Oskar Wolski (NCN), Franciska de Jong (CLARIN), Sara Garavelli (CSC)
Due Date of Deliverable	30/11/22, M6
Actual Submission Date	01/12/22, M7
Dissemination Level	PU
Approval Status	Submitted to EC
Version	V1

Deliverable Abstract

Deliverable 1.2 describes the Data Management Plan establishing the policy regulating collection, management, sharing, archiving, and preservation of research data in the EOSC Focus project throughout the project lifecycle and after the end of the project.



The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



DOCUMENT LOG

Issue	Date	Comment	Author	ORCID ID
V.1	01/12/2022	Submitted to REA EOSC Focus project officer	Tomasz Miksa	N/A

TERMINOLOGY

Terminology/Acronym	Definition
AAP	Additional Activities Plan
DMP	Data Management Plan
CSV	Comma Separated Values
EOSC	European Open Science Cloud
EOSC-A	EOSC Association
FAIR	Findable, accessible, interoperable, and reusable
MB	Megabyte
MF	Monitoring Framework
PDF	Portable Document Format
TU Wien	Vienna University of Technology
WP	Work Package

Table of contents

<i>Executive Summary</i>	1
1 <i>Introduction</i>	2
2 <i>Data Summary</i>	3
3 <i>FAIR data</i>	4
3.1 Making data findable, including provisions for metadata.....	4
3.2 Making data accessible	4
3.3 Making data interoperable.....	6
3.4 Increase data re-use	6
4 <i>Other research outputs</i>	7
5 <i>Allocation of resources</i>	8
6 <i>Data security</i>	9
7 <i>Ethics</i>	11
8 <i>Other issues</i>	12

List of figures

Figure 1 - ZENODO Repository	4
Figure 2 - Screenshot of the EOSC Collaboration Platform.....	9

List of tables

Table 1 - Produced datasets.....	3
Table 2 - Overview of datasets and repositories used for their publication, including a license.....	5
Table 3 - Access control rights for datasets during the project.	10
Table 4 - Datasets to be deleted at the end of the project.....	10

Executive Summary

The EOSC Focus Data Management Plan (DMP) describes the data management life cycle for the data to be collected, processed and/or generated by the project.

The DMP indicates the way in which the EOSC Focus consortium will manage the datasets that will emerge from the project, and how best practices in terms of metadata and archiving will be used to ensure that the data will be findable, accessible, interoperable, and reusable (FAIR).

This deliverable introduces the first version of the DMP and includes information on:

- 1 Handling of research data during & after the end of the project.
- 2 Data to be collected, processed and/or generated.
- 3 Methodology & standards to be applied.
- 4 Data to be shared/made open access.
- 5 Curation and preservation of the data during and after the end of the project.

This deliverable will be updated throughout the lifetime of the project.

1 Introduction

As stated in Article 8.1 of the Consortium Agreement '*A Data Management Plan (DMP) will be produced as a part of Task 1.1 in Work Package 1, including a plan for the management of the metadata of Background and Results that are software. The Data Management Plan will be approved by the General Assembly. Once approved, the Data Management Plan will be applied to data, including software, within the Project.*

A Data Management Plan (DMP) is a structured document that keeps record of what research data is created 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 Horizon Europe DMP template¹.

This is the initial version of the DMP. We consider it as a living document, that is, we will update it during the project to make sure that it is always up to date. This also implies that the contents of it are subject to changes.

Gloria Soriano Palomo will act as the Data Manager who will oversee the actions described by this DMP.

¹ Horizon Europe DPM template
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

2 Data Summary

Table 1 presents an overview of datasets that we will create within the project. It is based on the analysis of outputs of all work packages. We will collect the data in face-to-face and online meetings. A subset of data that will be made available at the end of the project may be relevant to everyone who is interested in EOSC and its activities. We do not reuse any existing data in this project.

Dataset ID	Title	Type	Format	Estimated volume	Contains sensitive data
P1	Deliverables and reports	Standard office documents	PDF	100 MB	No
P2	Event video recordings	Audio-visual data	MP4	1 GB	(No)
P3	Stakeholder Interview transcripts	Standard office documents	PDF	100 MB	Yes
P4	AAP Survey data	Structured text	PDF, CSV	100 MB	Yes
P5	MF Survey data	Standard office documents, Structured text	PDF, CSV	100 MB	Yes
P6	Interview recordings of EOSC-A member organisations	Audio-visual data	MP3	100 MB	No

Table 1 - Produced datasets.

For dataset P2 we will get consent from people to be in the videos as identifiable persons.

3 FAIR data

3.1 Making data findable, including provisions for metadata

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.

As there are no domain specific metadata standards applicable, we will provide a README file with an explanation of all values and terms used next to each file with data.

Additionally, we will provide common metadata such as title, description, or keywords when publishing data in open access repositories. In such a case, we will follow the default template provided by the repository, such as Data Cite Metadata.

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

3.2 Making data accessible

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.

We will use Zenodo to make data accessible:

“Zenodo builds and operates a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. Zenodo enables researchers, scientists, EU projects and institutions to:

- *Easily share the long tail of small research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science.*
- *Display their research results and get credited by making the research results citable; and,*
- *Integrate them into existing reporting lines to funding agencies like the European Commission. easily access and reuse shared research results”².*

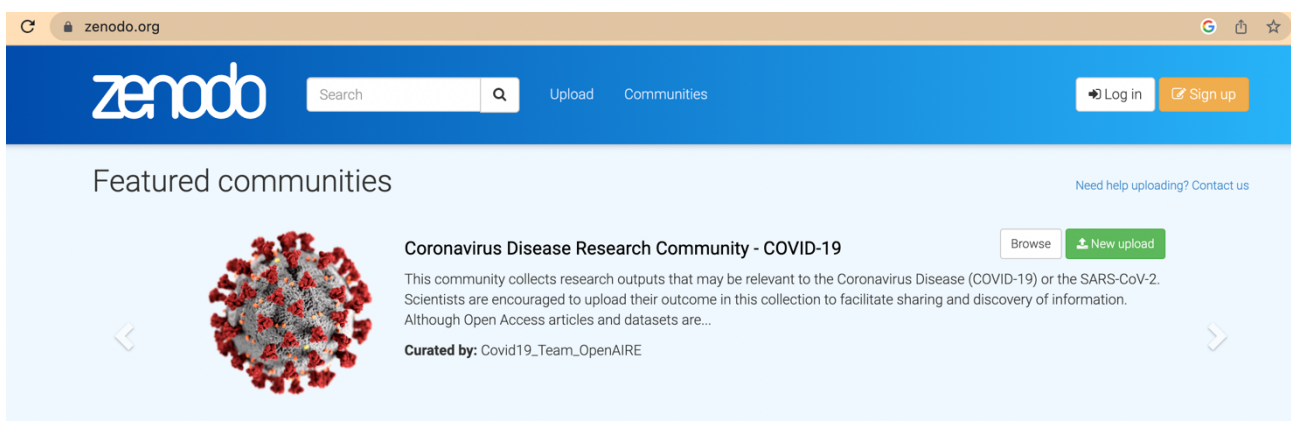


Figure 1 - ZENODO Repository

² Zenodo description. <https://www.re3data.org/repository/r3d100010468>

The repository will preserve the data for a minimum of 10 years, but longer retention period is expected.

Table 2 provides an overview of datasets, their location and expected license used for publication. It only includes datasets that can be made openly accessible via a repository. For this reason, datasets P4 and P5 are not listed.

Dataset ID	Access conditions	Restrictions / Embargo reasons	Estimated publication date	Location For Publication (Repository)	License
P1	Open	None	2025-05-30	Zenodo	CC BY 4.0
P2	Open	None	2025-05-30	Zenodo	CC BY 4.0
P3	Open	Must be anonymised before publication	2025-05-30	Zenodo	CC BY 4.0
P6	Open	None	2025-05-30	Zenodo	CC BY 4.0

Table 2 - Overview of datasets and repositories used for their publication, including a license.

3.3 Making data interoperable

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.

3.4 Increase data re-use

We will make our data reusable by adding metadata and comprehensive Readme files to all published datasets. The descriptions 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.

4 Other research outputs

There are no other research outputs, e.g., no software produced in this project.

5 Allocation of resources

The data manager will direct the data management process overall, with the work package leaders 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.

There will be no costs related to data management.

6 Data security

Currently the project is using Google Drive as its main internal collaboration and working tool to co-create and to share documents with each other. For the duration of the project, the data manager will ensure storage and backup of data in the shared drive.

In addition to this, we will use the EOSC Collaboration Forum to keep the data during the project. EOSC Collaboration Forum is an engagement tool to exchange information within the EOSC Association membership and its Task Forces. The platform is provided by Plek and customised for the EOSC Association’s needs. The platform will be also used as an internal project management tool for the EOSC Focus project. It’s a GDPR compliant tool. More information on it is available in Deliverable 1.1 – Project Quality and Risk Management Plan.

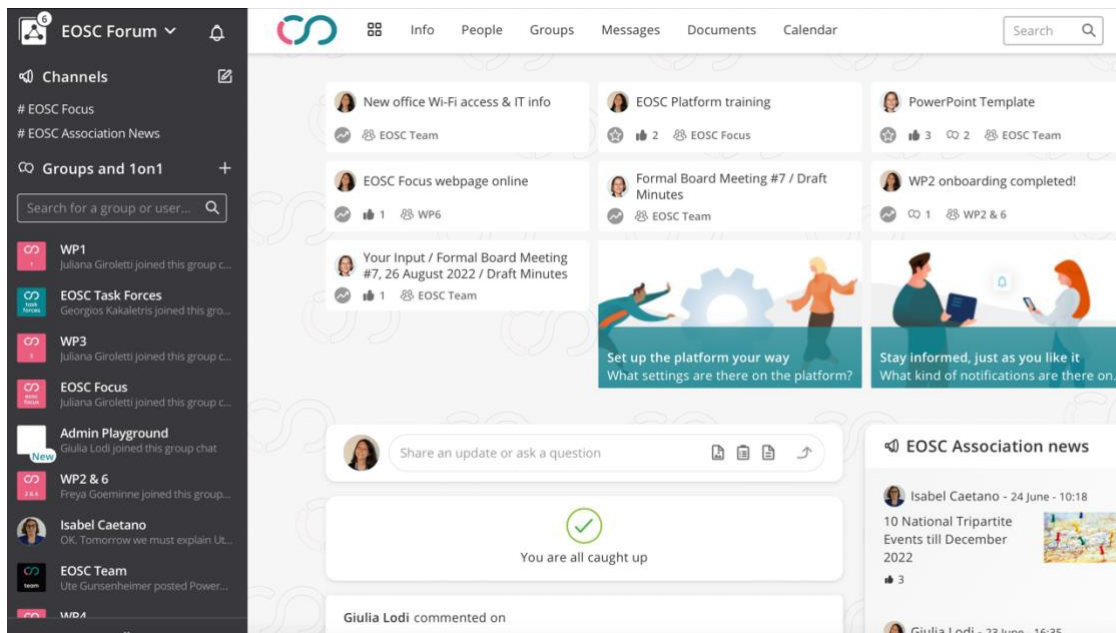


Figure 2 - Screenshot of the EOSC Collaboration Platform

We pay strict attention to compliance with the relevant institutional and national data protection policies. In this project there will be sensitive data on dataset P3 (Stakeholder Interview transcripts), P4 (AAP Survey data), P5 (MF Survey data). To ensure that the storage and transfer of sensitive data is safe, additional security measures such as individual log-in and password is taken. Selected work package members will be authorised to access sensitive data.

In this project, we will process personal data (see section 1). P3 (Stakeholder Interview transcripts), P4 (AAP Survey data), P5 (MF Survey data) will contain personal data. To ensure compliance with data protection laws, gaining informed consent for processing personal data and pseudonymisation of personal data will be used, for example, using Amnesia from OpenAIRE³.

Dataset ID	Selected members	project	All other project members	Public
P1	Writing		Writing	No Access
P2	Writing		Writing	No Access
P3	Writing		No Access	No Access
P4	Writing		No Access	No Access

³ <https://amnesia.openaire.eu>

P5	Writing	No Access	No Access
P6	Writing	Writing	No Access

Table 3 - Access control rights for datasets during the project.

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

Not all the datasets can be published at the end of the project. Table 4 lists datasets that must be deleted.

Dataset ID	Date deletion	Reason for deletion	Responsible person
P3	2025-05-30	Includes sensitive or personal data that cannot be made public.	Data Manager
P4	2025-05-30	Includes sensitive or personal data that cannot be made public.	Data Manager
P5	2025-05-30	Includes sensitive or personal data that cannot be made public.	Data Manager

Table 4 - Datasets to be deleted at the end of the project.

7 Ethics

No specific ethics issues are foreseen with the data to be used or produced by the project. This section will be updated if issues arise.

8 Other issues

None.