Data Management Plan

Lead Partners:	**GARR, TUWien**
Version:	**V1.0**
Status:	<mark>**Sumbitted draft not yet approved by the EC**</mark>
Dissemination Level:	**PU: Public**
Document Link:	**https://doi.org/10.5281/zenodo.7476530**
DOI:	**10.5281/zenodo.7476530**

Deliverable Abstract
This deliverable is the initial Data Management Plan (DMP) for the Skills4EOSC 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 Embedding Open Science in the Skills4EOSC Methodology section.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	Name	Partner/Activity	Date
From:	Tomasz Miksa	TU WIEN	13.12.2022
Moderated by:	Samah Jaber	TU WIEN	13.12.2022
Reviewed by:	Kevin Ashley	UEDIN	19.12.2022
Reviewed by:	Marta Teperek	TU-Delft	19.12.2022
Reviewed by:	Sara Di Giorgio, Emma Lazzeri	GARR	23.12.2022
Approved by:	Emma Lazzeri	GARR	23.12.2022

DOCUMENT LOG

Issue	Date	Comment	Author
v.0.0	*13.12.2022*	*First draft*	*Tomasz Miksa, Samah Jaber*
v.0.1	*19.12.2022*	*Comments by internal reviewers*	*Kevin Ashley, Marta Teperek*
v.0.2	*20.12.2022*	*Second draft*	*Tomasz Miksa, Samah Jaber, Emma Lazzeri*
v.1	***23.12.2022***	***Minor changes, version approved***	***Emma Lazzeri, Sara Di Giorgio***

TERMINOLOGY

https://eosc-portal.eu/glossary

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

Contents

1 Executive summary

1.1 Introduction

A data management plan (DMP) is a structured document that keeps record of what research data is created and used, 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.

Emma Lazzeri (GARR) will act as the data manager who will be in charge of the actions described by this DMP.

1.2 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 survey data using online tools. Specific tools will be determined later and the DMP will be updated to reflect it. 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 data in this project.

Table 1 – Produced datasets.

dataset ID	title	type	format	estimated volume	contains sensitive data
P1	Deliverables and reports	Standard office documents	PDF	100 MB	no
P2	General survey among the Skills4EOSC partners	Standard office document, Structured text	PDF, CSV	100 MB	yes
P3	Survey among the	Standard office	PDF,	100 MB	yes

		PhD students from the Skills4EOSC partners	documents, Structured text	CSV		
P4	Administrative data for the project	Structured text	CSV	100 MB	yes
P5	*Registry of RDM training resources*	Structured text	*CSV*	*100 MB*	*no*
P6	List of reference sources for minimum viable skillset properties	Structured text	CSV	100 MB	no
P7	OSCC Incubator Programme video recordings	Audio-visual data	MP4	*5 GB*	*no*
P8	Events video recordings	Audio-visual data	MP4	5 GB	no
P9	Reviews from training participants (questions, tests, answers...)	Structured text	CSV	*100 MB*	*no*
P10	Participants information	Standard office documents	CSV	100 MB	yes
P11	Participating institutions information	Standard office documents	CSV	*100 MB*	*yes*

2 FAIR data

2.1 Making data findable, including provisions for metadata

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

We will use the Data Documentation Initiative (DDI)[1] metadata standards whenever possible to document survey data. In all other cases, 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. In this case, we will follow the default template provided by the Zenodo repository, which is DataCite metadata[2].

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

2.2 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. All the data that cannot be made open or needs to be deleted is described in Section 5.

We will use Zenodo to publish data. Zenodo is 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

---

[1] https://ddialliance.org/
[2] https://schema.datacite.org

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.

All the project outputs can be found in the project community page[3] that is maintained by data manager.

The repository will host the data for minimum of 20 years, but longer retention period is expected. This meets our requirements.

Table 2 provides an overview of datasets, their location and expected license used for publication.

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

dataset ID	access conditions	restrictions / embargo reasons	estimated publication date	location for publication (repository)	PID	license
P1	Open		2025-08-30	Zenodo		CC BY 4.0
P2	Open	Must be anonymised before publication	2025-08-30	Zenodo		CC BY 4.0
P3	Open	Must be anonymised before publication	2025-08-30	Zenodo		CC BY 4.0
P5	Open		2025-08-30	Zenodo		CC BY

---

[3] https://zenodo.org/communities/skills4eosc/

						4.0
P6	Open		2025-08-30	Zenodo		CC BY 4.0
P8	Open		2025-08-30	Zenodo		CC BY 4.0
P9	Open		2025-08-30	Zenodo		CC BY 4.0

2.3 Making data interoperable

Whenever possible we will follow standards such as DDI or DataCite to foster interoperability of data. A 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.

2.4 Increase data re-use

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.

3 Other research outputs

One of the key outputs of the project will be learning materials, that will be produced by WP3-6 targeting different stakeholder. The design of Skills4EOSC materials will be based on the results of task 2.3, which started in M1 and will deliver a methodology for FAIR-by-design learning materials. The methodology is currently being defined based on previous works, and it will go through a co-design process engaging the Skills4EOSC participants and the larger EOSC community via specific consultation mechanism.

The methodology will include specification for metadata schema and repository to be used, process to ensure the material is in line with the FAIR principles, and thus fully reusable by both trainers and trainees, as well as a mechanism to assess the quality of the material itself.

The methodology will be the core of deliverable D2.2 due to M12.

4 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.

The costs of data management are all staff costs and accounted for.

5 Data security

For the duration of the project, the data manager will ensure safe storage and backup of data. Data during the project will be stored in **workplace.skills4eosc.eu** with limited access to people involved in the given WP. The service is hosted in Italy by GARR. It is GDPR compliant and the data is backed-up.

We pay strict attention to compliance with the relevant institutional and national data protection policies. In this project there will be sensitive data in dataset P2 (General survey among the Skills4EOSC partners), P3 (Survey among the PhD students from the Skills4EOSC partners), P4 (Administrative data for the project), P10 (Participants information) and P11 (Participating institutions information). 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 1a). P2 (General survey among the Skills4EOSC partners), P3 (Survey among the PhD students from the Skills4EOSC partners), P4 (Administrative data for the project), P10 (Participants information) and P11 (Participating institutions information) 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.

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

dataset ID	selected project members	all other project members	the public
P1	writing	writing	no access
P2	writing	reading only	no access
P3	writing	no access	no access
P4	writing	no access	no access

P5	writing	reading only	no access
P6	writing	reading only	no access
P7	writing	no access	no access
P8	writing	no access	no access
P9	writing	reading only	no access
P10	writing	reading only	no access
P11	writing	reading only	no access

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

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

Table 4 – Datasets to be deleted.

kind/name of data	date of deletion	reason for deletion	responsible person
P3	2025-08-30	Includes sensitive or personal data that cannot be made public.	Data Manager
P7	2025-08-30	Includes sensitive or personal data that cannot be made public.	Data Manager
P10	2025-08-30	Includes sensitive or personal data that cannot be made public.	Data Manager
P11	2025-08-30	Includes sensitive or personal data that cannot be made public.	Data Manager

Supporting | eosc

Co-funded by the European Union

UK Research and Innovation
co-funded by

Datasets P2 and P4 contain sensitive and personal data. They will not be deleted, but moved at the end of the project to a suitable storage operated by the EOSC Association.

6 ELSI

Skills4EOSC takes into account the Ethical, Legal and Social Issues in the project activities in its ELSI meta work package, which is transversal in the methodology. As ELSI are embedded throughout the activities of all work packages as transversal elements, one task from the ELSI Meta-WP is included in all relevant work packages (WP2. 3.4 and 5).

In the period covered by this first DMP version, the ELSI related project activities included discussion on the following aspects:

- Intellectual Property Rights (IPR)
- Public Sector Information (PSI),
- Open Data,
- licensing,
- database rights,
- integrity,
- responsible and participatory research,
- impact on society

In particular the ELSI aspects are currently discussed in WP3 to cover the science for policy interface and in WP5, addressing the different needs of the various research communities and the accessibility of the various types of data they produce.

These discussions address the balance between sharing research results under the principle "as open as possible, as closed as necessary" and the ethical impact on involved communities and individuals, as well as the legal aspects deriving from mandates and regulations, policies and legislative interventions at all levels.

7 Other issues

None. This section will be updated if issues arise.

8 References

No	Description/Link
R1	European Commission (2021). Horizon Europe (HORIZON) - Annotated Model Grant Agreement. Version 0.2, 30 November 2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf
R2	European Commission (2022a). Horizon Europe (HORIZON) - Model Grant Agreement. Version 1.1., 15 April 2022. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf