



D1.2 Data Management Plan

30/09/2024

Abstract

This first Data Management Plan (DMP) introduces a report that specifies how research data will be collected, processed, monitored, and catalogued, following the FAIR principles. This deliverable is viewed as a living report that will advance throughout the lifetime of the project.



Funded by
the European Union

EOSC Beyond receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101131875

Disclaimer: Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them

Document Description

D1.2 Data Management Plan			
Work Package Number 1			
Due date	30/09/2024	Actual delivery date:	27/09/2024
Nature of document	Report	Dissemination level	Public
Document Status	Under EC Review	Version	1.0
Lead Partner	EGI Foundation		
Authors	Sjomara Spetch (EGI), Hien Bui (EGI)		
Reviewers	Nicolas Liampotis (GRNET), Roksana Wilk (Cyfronet)		
Approved by	Konstantinos Androutsopoulos (OpenAIRE), on behalf of WPLG		
Document link	https://documents.egi.eu/document/4103		
DOI	https://zenodo.org/records/13849918		
Keyword	Data Management Plan, FAIR, GDPR, metadata, reusable data		

Revision History

Issue	Date	Description	Author/Reviewer
V 0.1	02/08/2024	ToC	Sjomara Specht (EGI)
V 0.2	07/08/2024	Initial first input	Sjomara Specht (EGI)
V 0.3	20/08/2024	Finalised and ready for review	Sjomara Specht (EGI)
V 0.4	16/09/2024	External Review	Nicolas Liampotis (GRNET), Roksana Wilk (Cyfronet)
V 0.5	23/09/2024	Incorporated feedback from reviewers	Hien Bui (EGI)
V 0.6	27/09/2024	Approval from the Board	Konstantinos Androutsopoulos (OpenAIRE)
V 1.0	27/09/2024	Final	Hien Bui (EGI)

Copyright and licence info

This material by Parties of the EOSC Beyond Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Table of content

Document Description	2
Revision History	2
Table of content	3
Tables	3
Executive Summary	4
1. Introduction	5
1.1. Purpose and Scope of the document	5
1.2. Structure of the Document	5
2. Data Summary	6
2.1. Potential Data Assets Generated by the Project	6
2.2. Types of Data	6
2.3. Existing data/software and newly generated project outputs	7
3. FAIR Data	10
3.1. Making data findable, including provisions of metadata	10
3.1.1. Metadata	10
3.1.2. Metadata Standards and Implementation	12
3.1.3. Justification for Metadata Choices	12
3.2. Making data openly accessible	12
3.2.1. Repositories	12
3.2.2. Standardised access protocol	13
3.2.3. Metadata availability	13
3.2.4. Making data interoperable	13
3.2.5. Increasing data re-use	13
4. Allocation of Resources	14
4.1. Data Management Responsibilities	14
5. Data Security	15
6. Ethical Aspect	16
7. Conclusion and Next Steps	17
8.1. DMP Change Management	17
Annex A. Privacy Policy	18
Annex B. Data sets	22
Data set Work Packages	22
Data sets use cases (6 use cases)	24

Tables

Table 1 - Repository metadata	10
Table 2 - Document metadata	10
Table 3 - EGI Privacy Policy for the EOSC Beyond project.	23

Executive Summary

The EOSC Beyond Data Management Plan (DMP) is a critical deliverable that outlines the framework for managing the data generated, processed, and disseminated throughout the EOSC Beyond project. This document is structured to ensure that all project outputs adhere to the FAIR principles—Findable, Accessible, Interoperable, and Reusable—and comply with the European Union’s Horizon Europe guidelines for open science and data management.

Key Points of the deliverable include:

- **Data Types and Formats:** The DMP identifies various types of data—textual, numerical, software code, multimedia, and presentations—produced within the project. Each data type will be managed according to its specific requirements to ensure it remains accessible and reusable long after the project’s conclusion.
- **FAIR Data Management:** The DMP provides a detailed plan for making data FAIR, including the use of standardised metadata, open repositories like Zenodo and GitHub, and interoperability standards such as Darwin Core and JSON. The plan also outlines the use of tools such as the F-UJI API to assess and improve data findability.
- **Data Security and Ethical Considerations:** The document addresses the security of both personal and open data, ensuring compliance with GDPR and other relevant regulations. It outlines procedures for anonymising personal data and safeguarding sensitive information, with clear protocols for data breaches and ethical management.
- **Resource Allocation and Responsibilities:** The DMP details the allocation of resources for data management activities and defines the roles and responsibilities of project participants. This includes the task leaders, WP leaders, and data managers, who are responsible for ensuring data quality and compliance with the DMP.
- **Change management for the DMP:** The DMP will be maintained and updated during the project to reflect the most recent developments and conclusions.

The DMP concludes that the robust management of data is essential for the success of the EOSC Beyond project. By adhering to the outlined strategies, the project will not only comply with open science mandates but also enhance the impact and reuse of its research outputs. The DMP is designed to be a living document, with scheduled updates at key project milestones to reflect new developments and ensure ongoing alignment with best practices in data management.

The recommendations include a continuous review of data management practices to adapt to evolving standards and technologies, ensuring that EOSC Beyond remains at the forefront of open science initiatives.

1. Introduction

The current initial version -D1.2 Data Management Plan - is composed of preliminary information and frameworks that will be followed. Hence, it is subject to updates in the future upon developments and changes during the project. In principle, the DMP describes the standards that will be used, and how the project's research data will be stored and published for verification and reuse. EOSC Beyond aims for full open access to data, following the FAIR principles.

1.1. Purpose and Scope of the document

The primary objective of this DMP is to establish a comprehensive strategy for data management that facilitates open access, ensures data security, and promotes the reuse of research outputs. The DMP covers all aspects of data management, from the types of data collected and the methodologies for their storage and dissemination to the ethical considerations and intellectual property rights associated with the data.

1.2. Structure of the Document

This document is comprised of the following chapters:

Section 1: presents an introduction to the project and the document.

Section 2: presents the purpose of data collection, type and format, and origin of the data.

Section 3: outlines EOSC Beyond FAIR data strategies

Section 4: briefly describes the allocation of resources.

Sections 5, 6 and **7**: outline data security, ethical issues.

Section 8: concludes this deliverable.

2. Data Summary

This section provides an overview of the potential data assets generated by the EOSC Beyond project, including the types and origins of data involved. The data generated will adhere to the FAIR principles, ensuring it is findable, accessible, interoperable, and reusable.

2.1. Potential Data Assets Generated by the Project

The EOSC Beyond project will generate and manage various types of data, including software, workflows, publications, and documentation. Data will be stored in a way that adheres to the FAIR principles, ensuring findability, accessibility, interoperability, and reusability.

2.2. Types of Data

The project will manage various types of data, including but not limited to, in addition to publications and documentation:

- **Textual data:** This includes documents, reports, publications, project deliverables, and meeting minutes. These are typically stored in formats such as .doc/.docx, .txt, and .pdf.
- **Numerical Data:** This comprises datasets that involve numerical information, including but not limited to, statistics, measurements, and quantitative research results. These are commonly stored in formats like .csv, .xls/.xlsx, and .json.
- **Geospatial Data:** This category includes data with geographic components, such as coordinates, maps, or location-based information. Geospatial data can be stored in formats like .shp, .geojson, .gml, .kml, and .tiff for raster data. These datasets are typically used for mapping, spatial analysis, and geographic visualisation.
- **Semantic Data:** This includes structured data enriched with meaning, often represented through ontologies, taxonomies, and controlled vocabularies. It supports interoperability and advanced data querying. Semantic data can be stored in formats like .rdf, .owl, and .ttl, which are commonly used for knowledge graphs, linked data, and semantic web technologies.
- **Code and Software:** The project will produce source code and software tools that support data analysis, processing, and visualisation. These are stored in formats such as .py, .R, .m, and other relevant programming languages. Software repositories like GitHub will be used for storage and version control.
- **Multimedia Records:** This category includes images, videos, and audio files that capture visual or auditory data relevant to the project. The formats will include .jpg/.jpeg, .png, .gif for images; .mp4 for videos; and .wav for audio files.
- **Presentations:** Visual aids and slide decks used for project meetings, dissemination events, and educational purposes. These will be saved in formats like .ppt/.pptx for editable presentations and .pdf for final versions.

- **Metadata** This category includes descriptive, structural, and administrative information about the above resource within the project. Metadata helps in organising, searching, and understanding the context of the data. Common formats for metadata include .xml, .json, .yaml, and .rdf, and metadata standards such as Dublin Core, ISO 19115 for geospatial data, or DCAT for datasets.

These formats ensure that the project outputs are accessible, interoperable, and reusable, adhering to the FAIR principles. The selection of specific formats is guided by community standards and the nature of the data or document.

2.3. Existing data/software and newly generated project outputs

The EOSC Beyond project builds on both existing datasets and software as well as newly generated outputs, which are crucial to achieving the project's objectives.

Existing Data/Software

EOSC Beyond will leverage existing data and software components that are foundational to the project's work. These include:

- Existing Data:
 - **Climate Data:** Gridded climate datasets from previous projects, particularly those conforming to CMIP standards, which provide a basis for further research and analysis.
 - **Biodiversity Data:** Information about species distribution, population, taxonomy, and ecosystems. Examples include species occurrence data, ecological community data, and biodiversity assessments.
 - **Geospatial Data:** spatial data that links for example ecological and biodiversity information to geographical locations, or data that enable spatial analysis of social phenomena.
 - **Genomic and Metagenomic Data:** Data related to species' genetic sequences and metagenomic analysis for understanding biodiversity at the molecular level.
 - **Proteomic data:** Data related to protein expression, structures, and functions, including mass spectrometry results and protein-protein interaction studies.
 - **Molecular Biology Data:** Datasets from molecular and cellular biology experiments, including gene expression profiles, molecular signaling pathways, and regulatory networks
 - **Biotechnological and life sciences data:** Data from computational models and simulations of biological systems, including systems biology models, metabolic pathways, and protein folding simulations
 - **Remote Sensing Data:** Satellite and aerial imagery are collected to monitor land use, habitat changes, and environmental phenomena.
 - **Survey Data:** Large-scale datasets from national or international social surveys. These datasets typically include responses to structured

questionnaires covering areas such as demographics, social behavior, attitudes, and opinions.

- **Historical Data:** Data sets covering past social, economic, or demographic trends. This can include archived survey data or reconstructed datasets from older censuses or historical research projects.
- **Microdata:** Anonymized datasets that provide individual-level data, often derived from surveys, censuses, or administrative records. These datasets are used for in-depth analysis of individual or household characteristics.
- **Cross-National Comparative Data:** Harmonised datasets from multiple countries, allowing for cross-national comparisons of social phenomena like employment, income distribution, public opinion, or health outcomes. Examples include data from international surveys like the **European Social Survey (ESS)** and **Eurobarometer**.
- **Food Composition Data:** like nutrient profiles and bioactive compounds.
- **Food Safety Data: Contaminants Data,** Data on the presence of chemical contaminants (e.g., pesticides, heavy metals, mycotoxins), microbial contaminants (e.g., pathogens), and allergens in food products; **Residue Analysis:** Data on residues from veterinary drugs, food contact materials, and other chemicals that may be found in food.
- **Traceability and Provenance Data** and other kinds of data coming from the **METROFOOD-RI:** Data tracking the origin and movement of food products across the supply chain, ensuring authenticity and compliance with safety regulations.
- **Metadata Repositories:** Established metadata repositories such as CESSDA Data Catalogue or LifeWatch ERIC Metadata Catalogue, which will serve as a reference for ensuring that new data generated by EOSC Beyond is discoverable and interoperable.
- Existing Software:
 - **Analysis Tools:** Established software tools: Jupyter Notebooks and Python scripts used for climate data analysis and visualisation, web services, F-UJI API for findability assessment of the content.
 - **Data Management Platforms:** Platforms like the ESGF (Earth System Grid Federation) and existing open-source repositories (e.g., GitHub) where pre-existing code and workflows are maintained and continuously improved.
 - **Catalogues and marketplaces:** Platform offering access to a wide range of research data, services, and tools. Researchers can discover, access, and share data, and also contribute their own resources.
 - **Core Services:** a set of foundational services enabling the integration and interoperability of diverse research resources in the EOSC ecosystem. It includes authentication, authorisation, persistent identifiers, and monitoring tools.

- **HelpDesk:** a centralised support service offering help to users navigating EOSC resources. It provides technical assistance, guidance on using services, and support for onboarding new resources.
- **Other services (software)** like platforms offering training modules and resources aimed at improving researchers' skills in open science practices and data management. EOSC Beyond is not responsible for external data/software but will rely on them as a foundation.

Individual Data Management Plans (DMPs) are maintained by service owners for these existing datasets and software, ensuring they are used appropriately within the project's scope.

Newly Generated Project Outputs

The project will generate a wide range of new data and software outputs, which will be carefully managed and documented to ensure they adhere to FAIR principles. These outputs include:

- **Newly Generated Data:**
 - **Research Data:** This includes newly collected experimental data, survey results, and other forms of primary data specific to the research objectives of EOSC Beyond.
 - **Knowledge Graph Data:** Generated as part of the project's effort to interlink datasets and metadata, facilitating enhanced search and discovery functionalities within the EOSC ecosystem.
 - **Service-Specific Datasets:** Data produced by specific services within the EOSC Beyond project, which will be catalogued and stored following best practices in data management.
- **Newly Developed Software:**
 - **Custom Tools:** Software specifically developed during the project to address unique research needs, including enhanced versions of existing tools or entirely new applications.
 - **Workflows and Scripts:** New workflows, including those for data processing, analysis, and visualization, particularly in Python and other relevant programming languages. These will be documented and shared through platforms like GitHub to ensure reusability.
 - **API Integrations, Connectors:** APIs developed to enhance interoperability between datasets and services, conforming to international standards to facilitate integration within the broader EOSC infrastructure.
- **New Metadata,** EOSC Beyond can produce enhanced metadata that ensures data and services compliance with the **FAIR** principles, which are essential for making data & services easy to find, access, and reuse across different research domains.

All newly generated project outputs will be stored in open repositories, such as Zenodo and GitHub, under appropriate open licences (e.g., CC BY for data and permissive open-source licences for software). This ensures that the outputs are accessible, interoperable, and reusable by the broader research community.

3. FAIR Data

To ensure that data is FAIR, EOSC Beyond will implement recognised standards such as the Darwin Core and JSON for data interoperability, and utilise tools like the F-UJI API to assess and enhance findability scores, particularly for datasets in repositories like CESSDA Data Catalogue. Additionally, OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) will be employed to facilitate metadata sharing across repositories

3.1. Making data findable, including provisions of metadata

Data will be designed to be FAIR and will be made available as openly and as early as possible. Data will only be restricted when required by regulatory or legal constraints or due to the owner's legitimate interests. Raw data from consultations (like user surveys and questionnaires) will be aggregated and anonymised to ensure it can be reused for future purposes.

3.1.1. Metadata

Metadata will adhere to a standardised schema based on community standards, ensuring consistency across all data types.

The metadata required by data repositories will be used as outlined in [Table 1](#). For documents, EOSC Beyond has defined a standard set of metadata to ensure consistent and comprehensive documentation of data assets as shown in [Table 2](#).

Metadata will be created and managed using tools like OpenAIRE Argos to ensure compatibility with FAIR principles and interoperability across repositories

Element	Definition
Title	A name given to the source.
Upload type	e.g., dataset, workflow, project deliverables
Abstract	Describing the document contents and main conclusions
Submitter	The person submitting the document to the repository
Authors	The people involved in contributing to a significant portion of the data
DOI	Provided by the resource
Publication date	The date of first publication
Version	The version number generated by the document repository for the repository identifier. Versioning rule: <ul style="list-style-type: none"> • +0.1 – a new version of the draft • +1.0 – a new version of the approved document

D1.2 Data Management Plan

Language	A language of the intellectual content of the resource.
Keywords	A list of words that will support the search within the repository service
Communities	A specific community in which the upload will appear.
License	Specifies the copyright status under which the upload will be licensed.
Modify	The groups can modify the document. The 'EGI office' SSO group must be always marked.

Table 1 - Repository metadata

Element	Definition
Title	A name given to the source. For milestones and deliverables as described in the Description of Work.
Lead Partner	The recognised short name of the lead partner within the EOSC Beyond project
Authors	The people involved in writing significant portions of the document.
Reviewers	The people involved in reviewing the document.
Approved by	The board involved in approving the document
Copyright status	Public material is licensed under a Creative Commons Attribution 4.0 International License . Confidential material may not be reproduced, distributed, or disclosed in whole or in part without the prior written consent of the EOSC Beyond consortium
Document Type	e.g., deliverable, report, white paper
Status	<ul style="list-style-type: none"> ● Draft - the document is being prepared ● Under EC review - the document is submitted to the EC portal and has not yet been approved by the European Commission ● Approved by EC - the document is approved by the European Commission ● Final - status of the document
Dissemination Level	<ul style="list-style-type: none"> ● Public - can be shared without restrictions ● Confidential - can be shared only with European Commission and project partners
Document link	The URL in the document repository that provides access to the document on the EGI Document Repository (DocDB).
DOI	An identification number is assigned through a repository service.
Keywords	A list of words that will support the search within the document repositories
Abstract	Describing the document contents and main conclusions

Table 2 - Document metadata

3.1.2. Metadata Standards and Implementation

Metadata will be generated in accordance with widely accepted community standards to ensure consistency, interoperability, and ease of discovery across different data types. Specifically:

Darwin Core: This standard will be used for biodiversity data, ensuring that datasets are interoperable with other international biodiversity databases. For instance, species occurrence data collected during the project will include metadata fields such as scientific name, location, and date of observation, following the Darwin Core schema. This will enable seamless integration with global biodiversity platforms like GBIF (Global Biodiversity Information Facility).

JSON-LD: For linked data and semantic web applications, we will utilise JSON-LD (JavaScript Object Notation for Linked Data) to structure metadata in a machine-readable format. This will facilitate the integration of EOSC Beyond datasets with other datasets on the web, enabling advanced data analytics and discovery. For example, metadata describing climate model outputs will be structured in JSON-LD to allow easy linking with external climate data repositories.

3.1.3. Justification for Metadata Choices

The selection of Darwin Core and JSON-LD as metadata standards is based on their wide acceptance and ability to support the specific data types generated by EOSC Beyond and their advantages with interoperability. Darwin Core is a globally recognized standard for biodiversity data, ensuring that our datasets are compatible with international repositories. JSON-LD is chosen for its flexibility in representing structured data and its compatibility with modern web technologies, making it ideal for datasets that require high levels of interoperability and machine-readability.

3.2. Making data openly accessible

Textual outputs such as reports and presentations will be made available to the EOSC Beyond Community on Zenodo upon release, under a permissive licence like CC BY. Sensitive legal, personal, or financial information will be restricted to authorised project partners, with access controls managed through secure platforms like the EGI Document Repository. All data will be accessible via a standard URL or DOI, with metadata openly available to ensure comprehensive documentation and discoverability. Special consideration will be given to personal data, which will be anonymised before sharing, and proprietary data, which will be subject to access restrictions as per legal agreements

3.2.1. Repositories

All documents, presentations and other materials that form an official output of the project (not just milestones and deliverables) are placed in the EGI's trusted [document repository](#) to provide a managed central location for all materials.

In addition, public deliverables and publications will be shared publicly via the [Zenodo platform](#) to increase the discoverability of the project outputs. Zenodo assigns a resolvable DOI to all submissions and exports metadata in standards such as Dublin Core and DataCite Metadata Schema.

All profiles, specifications, configuration files, software, workflows, and code will be deposited in Zenodo. Therefore, EOSC Beyond will use DocDB, Zenodo, and GitHub as their standard and main repositories.

3.2.2. Standardised access protocol

All data will be accessible via a uniform resource locator (URL) or Document Object Identifier (DOI). There will be no restrictions on the use of the research outputs, both during and after the end of this project.

3.2.3. Metadata availability

Metadata containing information to enable users to access the data will be openly available and published together with the data, in the same repositories as listed under [Repositories](#). There is no time limit on metadata and data availability.

The EOSC Beyond project acknowledges the value of documentation for interoperability purposes, increases uptake by different communities, and encourages data owners to document their research data assets. EOSC Beyond does not enforce specific provisions on documentation as long as the data asset is hosted on one of the mentioned repositories and properly curated according to the repository's best practices.

Research data itself should not be considered self-documenting and each published asset must be associated with sufficient documentation resources accessible through a public URL. Documentation must be browsable and include hypertext references to facilitate its fruition. Recommended documentation formats include markdown, HTML, and other markup languages. The inclusion of machine-readable documentation such as OpenAPI where applicable is thoroughly encouraged. If a scientific publication is tied to the research data asset, the publication itself should be referenced and/or made available as part of the documentation.

3.2.4. Making data interoperable

The EOSC Beyond consortium acknowledges the importance of data interoperability: if a data asset cannot be compared, merged, or otherwise integrated with other assets, then its publication would bring little or no value to the community. The consortium is therefore committed to supporting and enforcing data interoperability.

3.2.5. Increasing data re-use

Data and outputs will be licensed under open, permissive licences to maximise reusability. For example, CMIP6 outputs are publicly accessible, and tools will be available through GitHub for further development and reuse.

4. Allocation of Resources

Any expenses associated with the collection/production of FAIR data during the EOSC Beyond activities are included in the project budget. These expenditures will be required to cover a variety of particular data processing and data management operations, ranging from data collection and documentation to storage and preservation to distribution and re-utilization.

These operations are a component of the Work Packages (WPs) that process the relevant data, hence the needed effort will be part of the relevant WP.

The expenses of long-term data preservation are minimal, by using the EGI Online Storage and Google Drive platforms. Using Zenodo and GitHub (both free of charge) ensures that costs for long-term preservation of the data are manageable. Zenodo supports the long-term preservation of data deposits, with the repository projected to be maintained for at least the next twenty years by the host laboratory CERN (see Retention period in Zenodo policies: <https://about.zenodo.org/policies/>). When applicable, a more accurate cost estimate will be provided at a later stage of the project.

4.1. Data Management responsibilities

Within the EOSC Beyond project, the following roles and responsibilities are associated with Data Management, which are defined as follow:

WP leaders are in charge of organising the data processing and quality assurance that take place inside the Work Package they are leading.

Task Leaders are responsible for compiling and producing data within their assigned tasks. In addition to that, they also make sure that the data are properly prepared to be shared among the partners, and made publicly available, when applicable.

Data users are consortium partners who use, e.g., processing operations on the compiled/produced data.

Quality and Risk Manager monitors and supports the WP leaders, and Task Leaders/Use Case leaders in keeping the DMP confluence pages up to date. In addition, reports the changes and processes via milestones and deliverables as specified in the Grand Agreement.

5. Data Security

Any gathered data will be securely handled throughout the entire duration of the EOSC Beyond project, to protect it from loss and unauthorised access. Personal data is only accessible to those who are authorised¹ to access it.

All partners/beneficiaries responsible for processing personal data² have the responsibility to ensure that the data remains protected under all necessary security controls (including backup policies and integrity checks³) and access controls (identification, authentication, authorisation) within their infrastructure. In the unfortunate event of a personal data breach, the project partners will notify without delay their competent national supervisory authorities as well as the data subject(s) that may be affected by the breach. At the same time, they will document any personal data breaches and all related information.

Regarding open data, for security and long-term preservation, EOSC Beyond relies on the EGI Document Repository, Zenodo, Google Drive and GitHub platforms.

¹ Personal contact data are collected during the [EGI SSO](#) (Single Sign On) account creation by the users. After the access is granted, the users can manage their data in autonomy.

² Processing, according to Regulation (EU) 2017/679 of the European Parliament (GDPR), means any operation or set of operations which is performed on personal data or on sets of personal data, whether by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction

³ The integrity check is the process of comparing the current state of stored data and/or programs to a previously recorded state to determine any alteration or change.

6. Ethical Aspect

Ethical and legal issues within the EOSC Beyond project are handled meticulously to ensure full compliance with relevant regulations, including GDPR, and to protect the rights and privacy of all individuals involved in or affected by the research. The following strategies will be employed, with specific examples and procedures to illustrate their application:

- **Anonymisation of Personal Data:** For datasets that include personal data, such as survey responses or interview transcripts, we will implement a two-step anonymisation process. First, direct identifiers (e.g., names, addresses) will be removed. Second, indirect identifiers (e.g., job titles, locations) will be generalised or removed. For example, instead of recording “CEO of a specific company,” we will use “senior management position in the technology sector.” This approach ensures that no identification of natural persons is possible, thus protecting individual privacy and complying with GDPR requirements.
- **Restricted Data Dissemination:** In cases where legal agreements limit the dissemination of certain datasets (e.g., proprietary research data or third-party owned datasets), the data itself will not be made available. However, descriptive metadata, including the type of data, its purpose, and conditions for access, will be published to maintain transparency. For instance, if a dataset is restricted due to intellectual property agreements, we will include metadata that specifies how interested parties can request access under specific conditions.
- **Handling of Personal Data:** Personal data related to EOSC Core Innovation Sandbox users or Providers will be securely stored and will not be shared outside the consortium. For example, user registration data (such as email addresses) will be protected through appropriate technical and organisational measures to ensure data security. These measures include access restricted for authorised personnel only and regular audits to ensure that data handling practices comply with GDPR and other relevant regulations.

7. Conclusion and Next Steps

EOSC Beyond DMP is a complete data management approach that complies with Horizon Europe recommendations and that aims to make data as findable, accessible, interoperable, and reusable (FAIR) as feasible.

Relying on robust technological solutions and standards, such as the EGI Document Repository, Zenodo, GitHub and Google Drive⁴ for the execution of these processes. Additionally, this will ensure that the data created or compiled throughout the EOSC Beyond project, including open data and public publications, will be kept and continue to be usable once the project is completed.

The plan is intended to safeguard the analysis of compiled/created data based on the privacy level and to use an alternate sharing methodology relying upon this level. Confidential information or information that raises ethical problems will not be released.

Finally, the DMP is built on guaranteeing appropriately informed consent and protecting each participant's zone of privacy, while adhering to GDPR guidelines.

8.1. DMP Change Management

This DMP is considered a living document and will be updated during the project to reflect the most recent developments and conclusions. In actuality, this early version of the DMP will be refined and finalised at M36 of the project.

Ad hoc improvements may also be deployed if deemed necessary. In general, changes need to be fully compliant with EU laws and best practices in research data management.

The final version will also include another Annex specifying the dataset produced by each use case (pilot nodes) over the course of the project.

Updates will be entered in the changelog table that is shown on the confluence page of the concerned DMP. Analogously, the distribution of notifications on updates will be realised via regular meetings (WPs, PMO, TCB).

⁴ Google is listed in the DPF and thus, covered by EU-U.S. Data Privacy Framework <https://www.dataprivacyframework.gov/list>. All instances of Jira, GitHub, Confluence are operated on European Servers and compliant with GDPR

Annex A. Privacy Policy

With this privacy policy we, The EGI Foundation, inform the project members about which personal data is collected and processed when using EGI Collaboration Tools. Table 3 outlines the complete policy.

Name of the Service	Collaboration tools for the EOSC Beyond project
Description of the Service	<p>The EGI Collaboration Tools services for the EOSC Beyond (hereinafter referred to as: "the service" or "Collaboration Tools") support the EOSC Beyond project's activities. Personal data is used to provide access to the service with the proper access levels. Personal data is collected as part of the project's activities.</p> <p>This privacy notice describes how we, the EGI Foundation (hereinafter referred to as "we" or "the Data Controller"), collect and process data by which project members can be personally identified ("Personal Data") when the service is used.</p>
Data controller	<p>The EGI Foundation Science Park 140 1098 XG Amsterdam The Netherlands.</p>
Data protection officer	<p>The EGI Foundation Data Protection Officer Science Park 140 1098 XG Amsterdam The Netherlands E-mail: dpo@egi.eu</p>
Jurisdiction and supervisory authority	<p>Jurisdiction: NL, The Netherlands</p> <p>EGI Foundation's lead supervisory authority is the Dutch Data Protection Authority. They can be contacted at https://autoriteitpersoonsgegevens.nl/en/contact-dutch-dpa/contact-us</p>

<p>Personal data processed</p>	<p>The service may process the following personal data:</p> <p>Identification data:</p> <ul style="list-style-type: none"> ● Name ● Identification number ● E-mail address ● Phone number ● Address ● Bank details ● Other: affiliation, IP address <p>Behavioural data:</p> <ul style="list-style-type: none"> ● Usage data ● Data on purchase or payment transactions ● Working time data ● Other: technical logs with timestamps, attendance at meetings <p>Data allowing conclusions on the personality:</p> <ul style="list-style-type: none"> ● Other: membership information on groups, roles, and communities <p>Biographical data:</p> <ul style="list-style-type: none"> ● CV data <p>Sociodemographic data:</p> <ul style="list-style-type: none"> ● Gender
<p>Purpose of the processing of personal data</p>	<p>The purpose of the collection, processing, and use of the personal data mentioned above is:</p> <ul style="list-style-type: none"> ● To provide the service functions, to coordinate and manage the project according to applicable requirements including projects' contracts, guidelines, funding policies, and legal requirements. ● To keep evidence for audit needs. ● To monitor and maintain service stability, performance, and security.
<p>Legal basis</p>	<p>The legal basis for processing personal data is compliance with a legal obligation or legitimate interests pursued by the controller or by a third party according to Art. 6 (1) (f) General Data Protection Regulation (GDPR).</p>

Third parties to whom personal data is disclosed

Personal data will not be used beyond the original purpose of their acquisition. If forwarding to third parties should be necessary to answer an inquiry or to carry out a service, the consent of the data subject is considered to have been given by entering a contract when using the respective function or service. In particular, the data provided will not be used for advertising purposes.

For the purpose given in this privacy policy, personal data may be passed to the following third parties:

Within the European Union (EU) / European Economic Area (EEA):

- CESNET: resource provider, sub-contracted data processor of EGI Foundation
- Google Ireland Limited: resource provider, sub-contracted data processor of EGI Foundation
- [Zenodo](#): public deliverables and publications
- European Commission
- Project's partners
- Individuals responsible for managing projects, Work Packages, and tasks.
- The records of your use and technical log files produced by the service components may be shared for security incident response purposes with other authorised participants in the academic and research-distributed digital infrastructures via secured mechanisms, only for the same purposes and only as far as necessary to provide the incident response capability were doing so is likely to assist in the investigation of suspected misuse of Infrastructure resources.

Outside the EU / EEA:

- Project's partners
- Individuals responsible for managing projects, Work Packages, and tasks.

Any data transfer to a third country outside the EU or the EEA only takes place under the conditions contained in Chapter V of the GDPR and in compliance with the provisions of this privacy policy and any related policies adopted by the EGI Federation.

<p>Your rights</p>	<p>You can exercise the following rights at any time by contacting our Data Protection Officer using the contact details provided in the Data Protection Officer section:</p> <ul style="list-style-type: none"> ● Information about the data stored with us and their processing ● Correction of incorrect personal data ● Deletion of the data stored by us ● Restriction of data processing, if we are not yet allowed to delete the data due to legal obligations ● Objection to the processing of the data by us ● Data portability <p>Project members can complain at any time to the supervisory data protection authority (DPA). The responsible DPA depends on the country and state of residence, of the project member's workplace, or of the presumed violation. A list of the supervisory authorities with addresses can be found at https://edpb.europa.eu/about-edpb/board/members_en.</p> <p>You can contact EGI Foundation's lead supervising authority using the contact details provided in the Jurisdiction and Supervisory Authority section.</p>
<p>Data retention and deletion</p>	<p>As per EC Grant agreements' requirements, data should be kept for at least 5 years after the end of the project.</p> <p>The data are deleted or anonymised as soon as retention periods have passed, and the data are not required anymore for any of the purposes listed above.</p> <p>The records of the project members' use and technical log files produced by the service components will be deleted or anonymised after, at most, 18 months as documented in EGI-doc 2732: Policy on the Processing of Personal Data.</p>
<p>Security</p>	<p>We take appropriate technical and organisational measures to ensure data security and protection against accidental or unlawful destruction, accidental loss, alteration, unauthorised disclosure, or access.</p> <p>A comprehensive overview of the technical and organisational measures taken by EGI Foundation can be found at EGI Document 3737: EGI Foundation Technical and Organisational Measures (TOM)</p>
<p>Data Protection Code of Conduct</p>	<p>EGI Foundation is conforming to GEANT Code of Conduct and project members' personal data will be processed in accordance with the Code of Conduct for Service Providers and the EGI-doc-2732-v3: Policy on the Processing of Personal Data.</p>
	<p>This policy is based on AARC Policy development kit (licenced under CC BY-NC-SA 4.0)</p>

Table 3 - EGI Privacy Policy for the EOSC Beyond project.

Annex B. Data sets

Data set Work Packages

Data Summary	
Data description: Types of data	1. Project Documentation <ul style="list-style-type: none"> ● Metrics ● Risks ● Procedures ● Plans ● Meetings agenda ● Meetings participation list ● Meeting minutes ● Presentations ● Deliverables ● Mailing list archive ● External Feedback (surveys, events, etc) ● Services information ● Stakeholders information ● Promotional material (flyers, posters, branding materials, etc.) ● Other communication & dissemination material (multimedia, video, etc) ● Report
Data description: Origin of data	All the data will be produced and provided by project members.
Data description: Scale of data	<1GB
Standards and metadata	plain text such as .docx, .txt, .rtf, .pdf, .pptx, xml, .xls, .html . Multimedia such as jpg/jpeg, gif, tiff, png or video formats
Data sharing: Target groups	All project members and the EC Project office. Communication activities; will be publicly available focusing on the target audiences of the project: including users, technology providers and infrastructure providers
Data sharing: Scientific Impact	Scientific Publications in peer-reviewed journals, conferences & events aiming to engage stakeholders

Data sharing: Approach to sharing	<ol style="list-style-type: none"> 1. Shared within the consortium and European Commission <ul style="list-style-type: none"> ● Presentations: Public presentations are made public via Indico portal or external conference pages ● Deliverables: All deliverables are shared within the consortium and also with the European Commission. Public deliverables are accessible to everyone via the project website and Zenodo portal. ● Mailing list archive: only accessible by the mailing list members. ● Publications will be available via the project website and EOSC Beyond community on Zenodo Repository. ● Promotional and other audio-visual material will be available via the project website. 2. Shared with the Project office and management boards to support work, as well as with the European Commission. <p>Unless otherwise stated all content will be available under the CC BY 4.0 licence and metadata under the CC0 licence. Any consortium-restricted content is shared via access-protected confluence space</p>
Archiving and preservation	<p>Once the project is finished, all the information will be preserved by the EGI Foundation for at least 5 years as well on the EC funding portal. Publications will also be kept in the Zenodo Community.</p>
Allocation of resources	
Who will be responsible for data management in your WP/Task?	<p>Work Package Leaders and Task Leaders.</p>
How will long-term preservation be ensured?	<p>Long-term preservation is not needed, except from the contractual 5 years after the project. A copy of all the documentation of the project is kept by the European Commission in the funding portal. Deliverables, publications and other dissemination material are shared via the Zenodo portal, which grants long-term preservation</p>
Data Security	
What provisions are or will be in place for data security (including data recovery as well as secure	<p>To access the data shared only within the consortium, an EGI SSO account is required. Accounts and access management is the responsibility of the coordinator.</p>

<p>storage/archiving and transfer of sensitive data)?</p>	
<p>Will the data be safely stored in trusted repositories for long-term preservation and curation?</p>	<p>For security and long-term preservation, EOSC Beyond relies on EGI Document Repository, Zenodo and Google Drive platforms</p>
<p>Other issues</p>	
<p><i>Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?</i></p>	<p>EGI Foundation will take care of the data according to the ISO 27000 standard for Information security management and GDPR.</p>