



**D1.4**

# Data Management Plan



Version 1.0  
September 2021

# D1.4 / Data Management Plan

Lead by JNP

Authored by Yannis Mitsos (JNP), Athanasia Spiliotopoulou (JNP)

Reviewed by Hilary Hanahoe (RDA) & Matthew Viljoen (EGI)

## Dissemination Level of the Document

Public

## Abstract

This deliverable outlines the Data Management Plan for the data collected and generated by the EOSC Future consortium, to ensure that the project complies with the Open Research Data Pilot requirements. The document follows the Guidelines on FAIR Data Management in Horizon 2020 and provides the initial description of the data and its management plan as recommended by the template annexed to the said guidelines. This includes data related to the EOSC Portal, a web portal facilitating the access to and use of the EOSC Resources, as well as to the activities during the project. The description covers the type, origin, scale, and security of data, its curation, preservation, sharing as well as the handling of the metadata and the standards applied.

## Version History

Version	Date	Authors/Contributors	Description
Vo.1	01/06/2021	Yannis Mitsos (JNP), Athanasia Spiliotopoulou (JNP)	Initiation – Proposed ToC – First draft
Vo.2	16/07/2021	Yannis Mitsos (JNP), Antonis Zervas (JNP)	Interim draft of the deliverable
Vo.3	31/08/2021	Yannis Mitsos (JNP), Athanasia Spiliotopoulou (JNP), Jorge-A. Sanchez-P. (JNP), Nikos Vogiatzis (JNP).	Complete Final Draft available for internal review
Vo.4	15/09/2021	Hilary Hanahoe (RDA), Matthew Viljoen (EGI)	Comments and revisions from internal quality review
Vo.5	23/09/2021	Yannis Mitsos (JNP)	Consolidated draft delivered to Project Leader
Vo.6	24/09/2021	Yannis Mitsos (JNP)	Final Document for submission
V1.0	27/09/2021	Yannis Mitsos (JNP), Ron Dekker (TGB), Mike Chatzopoulos (ATHENA)	Final Version submitted to EC

## Copyright Notice



This work by Parties of the EOSC Future Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020, Grant Agreement number 101017536.

## Table of Contents

List of Abbreviations .....	3
<b>1 Executive Summary .....</b>	<b>5</b>
<b>2 Introduction .....</b>	<b>6</b>
<b>3 Data Related to the EOSC Portal .....</b>	<b>7</b>
<b>3.1 Data Summary .....</b>	<b>7</b>
3.1.1 EOSC Portal Provider Profile .....	7
3.1.2 EOSC Portal Resource Profile .....	8
3.1.3 Usage activity and user-collected information .....	10
<b>3.2 FAIR Resources .....</b>	<b>11</b>
3.2.1 Making data findable, including provisions for metadata .....	11
3.2.2 Making data accessible .....	11
3.2.3 Making data interoperable .....	11
3.2.4 Increase data re-use .....	12
<b>3.3 Allocation of Resources .....</b>	<b>12</b>
<b>3.4 Data security .....</b>	<b>12</b>
<b>3.5 Ethics .....</b>	<b>12</b>
<b>4 Data related to EOSC Future Project Research Activities .....</b>	<b>14</b>
<b>4.1 Data Summary .....</b>	<b>14</b>
<b>4.2 FAIR Outputs .....</b>	<b>14</b>
4.2.1 Making data findable, including provisions for metadata .....	14
4.2.2 Making data accessible .....	14
4.2.3 Making data interoperable .....	14
4.2.4 Increase data re-use .....	15
<b>4.3 Allocation of resources .....</b>	<b>15</b>
<b>4.4 Data Security .....</b>	<b>15</b>
<b>4.5 Ethics .....</b>	<b>15</b>
<b>5 IPR Management .....</b>	<b>16</b>
<b>6 Conclusions .....</b>	<b>17</b>
<b>Appendix A EOSC Future Confluence and Jira Privacy Policy .....</b>	<b>18</b>
<b>References .....</b>	<b>22</b>

## Table of Tables

Table 3-1: EOSC Portal Provider Profile .....	7
Table 3-2: EOSC Portal Resource Profile .....	9

## List of Abbreviations

Acronym	Definition
<b>ACL</b>	Access List
<b>API</b>	Application Programming Interface
<b>CA</b>	Consortium Agreement
<b>CC-BY</b>	Creative Commons attribution
<b>DESCA</b>	Development of a Simplified Consortium Agreement
<b>DIH</b>	Digital Innovation Hub
<b>DMP</b>	Data Management Plan
<b>DOI</b>	Digital Object Identifier
<b>EC</b>	European Commission
<b>EOSC</b>	European Open Science Cloud
<b>ESFRI</b>	European Strategy Forum on Research Infrastructures
<b>EU</b>	European Union
<b>FAIR</b>	Findable Accessible Interoperable Reusable
<b>GAG</b>	Grant Agreement
<b>GNU</b>	GNU's Not Unix
<b>GDPR</b>	General Data Protection Regulation
<b>GPL/A</b>	General Public License Affero
<b>IdP</b>	Identity Provider
<b>IP</b>	Intellectual Property
<b>IPR</b>	Intellectual Property Rights
<b>IaaS</b>	Infrastructure as a Service
<b>IT</b>	Information Technology
<b>JSON</b>	JavaScript Object Notation
<b>PID</b>	Persistent Identifiers
<b>PMB</b>	Project Management Board
<b>RDF</b>	Resource Description Framework
<b>RDFS</b>	RDF Schema
<b>REST</b>	REpresentational State Transfer
<b>RI</b>	Research Infrastructure
<b>SOB</b>	Strategy and Oversight Board
<b>SKOS</b>	Simple Knowledge Organization System
<b>SME</b>	Small Medium Enterprise
<b>TRL</b>	Technology readiness level
<b>XML</b>	Extensible Markup Language

<b>XSD</b>	XML Schema Definition
<b>WP</b>	Work Package

## 1 Executive Summary

The Data Management Plan (DMP) provides an overview of data to be managed by EOSC Future and identifies solutions, both 'existing/already planned' and 'possibly to be developed' for meeting the FAIR principles.

Within EOSC Future, the following three categories of data to be managed are foreseen:

1. Data generated in operating the EOSC (including Portal, supply/demand layers, supporting functions);
2. Data generated in DIH Pilots and other work supported by RDA open calls;
3. Data from communities connected to EOSC via the EOSC-Exchange.

Data from (1) will be kept and analysed to support better understanding of the EOSC ecosystem, to support added value services to users and to support tracking of impact through usage.

For (2), this will be handled according to best practices already in place within the Science Clusters and reported on in the DMP.

For (3), EOSC Future will not own or control these data as they only pass through EOSC, but it will encourage good research data management and adoption of FAIR principles.

The DMP also specifies the recommended licensing schemes following project IPR management approaches.

EOSC Future will comply with the EU and national regulations on data handling and publishing, in line with the consortium's exploitation plans. The project will carefully address privacy and copyright issues prior to publicly releasing any data. All users will be made aware of the GDPR regulations.

The document includes the following sections: Section 3 provides a FAIR analysis on the data generated by operating the EOSC Portal, while section 4 elaborates on the data generated by other activities anticipated in the project's lifetime. Section 5 approaches the IPR uses and finally section 6 concludes this deliverable.

The DMP will be updated, if required, over the course of the project lifetime.

## 2 Introduction

EOSC Future project targets to develop an operational, common and interoperable EOSC Platform ('System of Systems') with an integrated execution environment consisting of data, professionally provided services, and open research products and infrastructure that will be accessed and used by the European Researchers who will be engaged, facilitated, trained and supported to utilise the EOSC Resources and solutions. Such platform will serve as a web of scientific resources that are open and/or Findable, Accessible, Interoperable and Reusable (FAIR) and will be discoverable, accessible, and composable.

To achieve this, EOSC Future will aggregate services provided by Research Infrastructures (RIs), e-Infrastructures, Science Clusters, and Research Organizations leveraging, enhancing, expanding, integrating, and optimizing, where necessary and possible, the outputs of past and current EOSC projects (including EOSC-hub<sup>1</sup>, OpenAIRE-Advance<sup>2</sup>, EOSC Enhance<sup>3</sup>, eInfraCentral<sup>4</sup>, Science Cluster projects, and the INFRAEOSC-07 projects). In the EOSC ecosystem, the ESFRI Research Infrastructures, organized in five thematic Science Clusters [7], which are ENVRI-FAIR for environmental research, PaNOSC for multidisciplinary scientific analysis, ESCAPE for astronomy and particle physics, SSHOC for social sciences and humanities and EOSC-Life for life sciences, are critical for the provision of data and services to their research communities and integrating these data and services into the use of the emerging EOSC infrastructure, is of paramount importance.

The Science Clusters, along with the projects that will be funded under the INFRAEOSC-07 call, will provide the scientific data and services needed for interdisciplinary research and will be showcasing the benefits of the EOSC Platform.

Building on previous efforts and projects' results, EOSC Future will focus on a) the realization of EOSC-Core and EOSC-Exchange with interoperable data and Resources; b) the integration of data and Resources from the Science Cluster communities into the EOSC Platform; c) the direct involvement of users in the co-design and implementation of the EOSC Platform.

EOSC Future project implementation plan is structured in three main technological iterations (by Month 6, 18, and 30) to achieve the following eight general objectives:

1. Deliver and operate the EOSC-Core;
2. Expand EOSC-Exchange with Resources across disciplines;
3. Scale up capabilities and deliver an EOSC Execution Framework;
4. Increase European scientific impact with EOSC integrated and interoperable cross-domain scientific Resources, and collaboration;
5. Enable innovation with SMEs and industry through procurement of commercial services and EOSC Digital Innovation Hub (DIH);
6. Support and train users and Providers of EOSC;
7. Reach out and engage with EOSC communities and end users;
8. Align implementation with the strategic vision of EOSC.

The project commencement date is the 1st of April 2021, and its execution plan will be aligned with the base-line technical roadmap linked to deliverables and milestones to be further developed and specified during project implementation.

This document presents the Data Management Plan (DMP) for the data to be collected and/or created during the lifecycle of the EOSC Future project. The DMP complies with the guidelines of FAIR Data Management in Horizon 2020 [1].

---

<sup>1</sup> <https://www.eosc-hub.eu>

<sup>2</sup> <https://www.openaire.eu/advance/>

<sup>3</sup> <https://eosc-portal.eu/enhance>

<sup>4</sup> <https://cordis.europa.eu/project/id/731049>



## 3 Data Related to the EOSC Portal

### 3.1 Data Summary

The overarching goal of EOSC Future project is to build an EOSC platform that takes a 'System of Systems' approach linking to other research portals and offering added value resources to a wide range of users.

One of the main objectives in achieving such goal is the expansion of the EOSC-Exchange with Resources across disciplines. Specifically, the onboarding of INFRAEOSC-07 and Science Clusters Resources will be facilitated by onboarding procedures and specifications to be developed while the e-Infrastructures and Research Infrastructures will collaborate to jointly deliver common services supporting scientific workflows.

The EOSC Portal collects, manages, and visualizes content about Resources including services, datasets, assets, research products and in general scientific Resources as well as commercial Resources through the integration of commercial providers into EOSC, in order to enable members of the Research and Academic community to easily access them via the EOSC Portal.

For each Resource, a set of metadata is collected and presented by the EOSC Portal as an integrated and harmonised catalogue to enable Resources discoverability and composability.

The data types managed by the EOSC Portal also include information about the Providers and Users of the EOSC Portal, activity data produced by the interaction of the users within the EOSC Portal and aggregated statistics produced by analysing the EOSC Portal content and used for visualization purposes (e.g., visualise Resources per country on map, etc).

Three main entities will comprise the EOSC Portal data types as described in the following sections.

#### 3.1.1 EOSC Portal Provider Profile

Information for the Providers, offering EOSC Resources, will be collected and stored in the EOSC Portal upon successful completion of the organisations' onboarding process that will be conducted by authoritative representatives. This information includes:

- Basic information about the Provider, such as the Provider ID, name, legal status, etc.;
- Marketing information, such as logo and multimedia;
- Classification information, such as the scientific domain where the Provider offers Resources;
- Location information, details of the address of the Provider;
- Contact information, details of the main and public contact of this Provider;
- Maturity information, certifications received, etc.;
- Other information, affiliations, ESFRI domain, etc.

Table 3-1 provides an overview of the EOSC Portal Provider Profile data structures with the attributes' names and example values.

Table 3-1: EOSC Portal Provider Profile

Attribute Name	Example
<b>Basic Information</b>	
<b>ID</b>	embl-ebi
<b>Name</b>	Earth Observation Data Centre for Water Resources Monitoring
<b>Abbreviation</b>	Carl Zeiss
<b>Website</b>	<a href="https://jelastic.com/">https://jelastic.com/</a>
<b>Legal Entity</b>	Y
<b>Legal Status</b>	Non-Profit Partnership (NPP)
<b>Marketing Information</b>	
<b>Description</b>	EnhanceR is a nationally and internationally recognized network for Swiss research IT expertise. It is an association according to Swiss law.
<b>Logo</b>	<a href="https://jelastic.com/wp-content/themes/salient/img/logo.svg">https://jelastic.com/wp-content/themes/salient/img/logo.svg</a>
<b>Multimedia</b>	<a href="https://vimeo.com/XYZ123">https://vimeo.com/XYZ123</a>
<b>Classification Information</b>	

<b>Scientific Domain</b>	Information Science and Technology
<b>Scientific Subdomain</b>	Complex Data Facilities
<b>Tags</b>	Open Data, Open Science, Publications
<b>Location Information</b>	
<b>Street Name and Number</b>	Street 10
<b>Postal Code</b>	GR-12345
<b>City</b>	Athens
<b>Region</b>	Attica
<b>Country</b>	Greece
<b>Contact Information</b>	
<b>Main Contact/Provider Manager</b>	
<b>First Name</b>	Jack
<b>Last Name</b>	White
<b>Email</b>	jack.white@example.com
<b>Phone</b>	+30123456789
<b>Position</b>	Coordinator
<b>Public Contact</b>	
<b>First Name</b>	Alan
<b>Last Name</b>	Smith
<b>Email</b>	alan.smith@example.com
<b>Phone</b>	+30123456789
<b>Position</b>	Manager
<b>Maturity Information</b>	
<b>Life Cycle Status</b>	Operational
<b>Certifications</b>	ISO-27001
<b>Other Information</b>	
<b>Hosting Legal Entity</b>	LegalEntityName
<b>Participating Countries</b>	Greece
<b>Affiliations</b>	ELIXIR
<b>Networks</b>	GEANT
<b>Structure Type</b>	Virtual
<b>ESFRI Domain</b>	Data, Computing and Digital Research Infrastructures
<b>ESFRI Type</b>	Not and ESFRI project or landmark
<b>MERIL Scientific Domain</b>	Humanities & Arts
<b>MERIL Scientific Subdomain</b>	Collections
<b>Areas of Activity</b>	Applied research
<b>Societal Grand Challenges</b>	Wellbeing
<b>National Roadmaps</b>	Yes

### 3.1.2 EOSC Portal Resource Profile

Another entity collected by the EOSC Portal is a Resource. A Resource is described by a well-defined schema, which contains:

- Basic information about the Resource, such as the resource ID, name, description, provider, etc.;
- Marketing information related to the target users, multimedia related to the Resource, etc.;
- Classification information, such as the category and the scientific domain related to the Resource;
- Location information, i.e., countries and languages the Resource is available;
- Contact information, i.e., contact details for individuals responsible for user request for the Resource;
- Maturity information, i.e., TRL, version, phase, etc.;
- Dependencies on other EOSC Resources;
- Attribution to funding bodies and projects;
- Management information, i.e., terms of use, helpdesk, and other operational information;
- Access and order information;
- Financial information, such as price or financial model.

Table 3-2 provides an overview of the EOSC Portal Resource Profile data structures.

Table 3-2: EOSC Portal Resource Profile

Attribute Name	Example
<b>Basic Information</b>	
ID	openrisknet.squonk_computational_notebook
Name	PaaS Orchestrator
Resource Organisation	openair
Resource Providers	cines
Webpage	https://cite.vamdc.eu
<b>Marketing Information</b>	
Description	Connect spatial information communities and their data using a modern architecture, which is at the same time powerful and low cost, based on International and Open Standards for services and protocols (a.o. from ISO/TC211 and OGC).
Tagline	The open repository for high performance computing code samples
Logo	http://openminted.eu/wp-content/uploads/2018/08/catalogue-of-TDM-components.png
Multimedia	https://vimeo.com/XYZ456
Use Cases	https://example.com/UseCases.html
<b>Classification Information</b>	
Scientific Domain	Natural Sciences
Scientific Subdomain	Earth and related environmental sciences
Category	Processing and analysis data management
Subcategory	Aggregators and integrators aggregators and integrators data
Target Users	Researchers, Research group, Research community, Research project
Access Type	Virtual
Access Mode	Excellence-based, Peer reviewed
Tags	Open Science, data, dataset, data archive, library, repository
<b>Geographical and Language Availability Information</b>	
Geographical Availability	World
Language Availability	English
<b>Resource Location</b>	
Geographic Location	EU
<b>Contact Information</b>	
<b>Main Contact/Provider Manager</b>	
First Name	Jack
Last Name	White
Email	jack.white@example.com
Phone	+30123456789
Position	Director
Organization	Openedition
<b>Public Contact</b>	
First Name	Alan
Last Name	Smith
Email	alan.smith@example.com
Phone	+30123456789
Position	Manager
Organization	Organization
<b>Other Contact</b>	
Helpdesk Email	helpdesk@Provider.com
Security Contact Email	secutiry@Provider.com
<b>Maturity Information</b>	
Technology Readiness Level	TRL8
Life Cycle Status	Production

<b>Certifications</b>	ISO27001
<b>Standards</b>	OpenStack API OAuth
<b>Open Source Technologies</b>	OpenStack
<b>Version</b>	0.1
<b>Last Update</b>	31/8/2021
<b>Change Log</b>	SparQL endpoint added
<b>Dependencies Information</b>	
<b>Required Resources</b>	egi-fed.check-in
<b>Related Resources</b>	egi-fed.cloud_compute
<b>Related Platforms</b>	MetaCentrum
<b>Attribution Information</b>	
<b>Funding Body</b>	European Commission
<b>Funding Program</b>	Horizon 2020
<b>Grant/Project Name</b>	1234567890
<b>Management Information</b>	
<b>Helpdesk Page</b>	<a href="https://about.west-life.eu/network/west-life/documentation">https://about.west-life.eu/network/west-life/documentation</a>
<b>User Manual</b>	<a href="https://github.com/eubr-bigsea/lemonade">https://github.com/eubr-bigsea/lemonade</a>
<b>Terms of Use</b>	<a href="https://www.openaire.eu/infrastructure-acceptable-use-policy">https://www.openaire.eu/infrastructure-acceptable-use-policy</a>
<b>Privacy Policy</b>	<a href="https://www.openaire.eu/infrastructure-acceptable-use-policy">https://www.openaire.eu/infrastructure-acceptable-use-policy</a>
<b>Access Policy</b>	<a href="https://doi.org/10.17882/42182">https://doi.org/10.17882/42182</a>
<b>Resource Level</b>	<a href="https://auth.west-life.eu/aai/AUP.pdf">https://auth.west-life.eu/aai/AUP.pdf</a>
<b>Training Information</b>	<a href="https://www.seadatanet.org/Events/Training-courses">https://www.seadatanet.org/Events/Training-courses</a>
<b>Status Monitoring</b>	<a href="https://status.ror.org/">https://status.ror.org/</a>
<b>Maintenance</b>	<a href="http://Resource-name.Pro-vider.eu/Maintenance">http://Resource-name.Pro-vider.eu/Maintenance</a>
<b>Access and Order Information</b>	
<b>Order Type</b>	Open access
<b>Order</b>	<a href="https://Resource-name.Provider.eu/order">https://Resource-name.Provider.eu/order</a>
<b>Financial Information</b>	
<b>Payment Model</b>	<a href="http://Resource-name.Pro-vider.eu/PaymentModel">http://Resource-name.Pro-vider.eu/PaymentModel</a>
<b>Pricing</b>	<a href="http://Resource-name.Provider.eu/Price">http://Resource-name.Provider.eu/Price</a>

The above information will be collected from the registered EOSC Providers through the EOSC onboarding process (web interface or API) and updated either manually or automatically (via an API) by the EOSC Providers.

### 3.1.3 Usage activity and user-collected information

Users of the platform are unauthenticated or authenticated end-users who access and browse the EOSC Portal and provide feedback related to a Resource and Provider Managers/Resource Managers—members of a Provider organisation, who are authorised to upload and update Resource related content at the EOSC Portal. Usage events and statistics are collected by the EOSC Portal and then analysed to produce consolidated statistics that are visualised in the EOSC Portal.

The following information about users' visits is collected in the EOSC Portal database which are fetched either from the Identity Provider (IdP) or from the user's clicks while exploring the Portal:

- User unique identifier;
- First name;
- Last name;
- Email;
- Reason for accessing EOSC services;
- Country of origin;
- Customer typology, (1) Researchers, (2) Research groups, (3) Research communities, (4) Research projects, (5) Research networks, (6) Research managers, (7) Research organisations, (8) Students, (9) Innovators, (10) Businesses, (11) Resource Providers, (12) Funders, (13) Policy Makers, (14) Research Infrastructure Managers, (15) Resource Provider Managers, (16) Resource Managers, and Others;

- Organisation/company;
- Country of collaboration;
- Webpage of the EOSC Portal accessed;
- Ordered Resources;
- Messages (communication with support team);
- User rating information, etc.

The users' data is used in the ordering process and provided to the EOSC support team to complete resource orders and support users. This data is not publicly displayed in any place of the EOSC Portal.

Furthermore, the content will be stored and managed in XML format. Access to the EOSC Catalogue and Resource metadata for all interested users will be provided through the EOSC Portal [2] and through APIs.

The software and data related to the EOSC Future project may be of interest to all users who need a uniform way of accessing, browsing, searching, comparing, and visualizing the EOSC Resources, a single point of reference for researchers and the broad community to provide access and make Resources discoverable.

## 3.2 FAIR Resources

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

Access to the EOSC Portal Catalogue is provided through the EOSC Portal web interface and dedicated APIs. The main entities in the EOSC Portal, the Provider and the Resource are described by a structured set of metadata, which are called EOSC Profiles and are publicly available in the EOSC Portal documentation section [3], [4]. The current specifications are in v3.00.

According to the currently implemented functionality at the EOSC Portal, each Resource is identified by an ID, which is generated by the EOSC Portal based on unique characteristics of the Resource, such as the Provider, the internal ID and the version. However, in the future, external PID providers will be used for assigning PIDs to Resources.

The EOSC Portal will also offer faceted browsing and keyword search functionality for enabling users to discover and find Resources in a most intuitive manner. The EOSC Portal will use the open-source software Elastic Search to index the whole set of metadata of the EOSC Resources and enable the above two features.

In addition, the EOSC Portal will support two means of versioning; firstly, the Resource model holds information about the Resource version as identified by the Provider; secondly, the EOSC Portal itself will keep an internal logging and versioning system for tracking of updates of Resource characteristics.

Each Resource update will generate a new version which will be the one accessible through the platform, while previous versions of a Resource will be archived for future use. Finally, all actions will be recorded, and provenance information related to the user and the action performed will be stored in the platform registry.

### 3.2.2 Making data accessible

The EOSC Portal will provide continuous access to the EOSC Catalogue via the web platform as well as via standard APIs. Data, available in the EOSC Portal will also be available for bulk download in the form of XML and JSON formats.

Data related to the portal usage, such as Resource views and orders will be analysed internally but not shared in raw format. They will be presented as aggregated statistics (e.g., average) for each Resource.

The description of the Provider and Resource Model (e.g., the EOSC Profiles specifications) is documented in the platform and is available for download in various open formats, such as XSD, JSON and RDFS.

### 3.2.3 Making data interoperable

Each Resource collected in the EOSC Portal is stored as a separate XML object. For metadata interoperability, Resource schema will reuse terms from widely known vocabularies and ontologies such as the Dublin Core terms and SKOS.

Taxonomies and classifications for enumerated attributes of a Resource (e.g., categories, providers, languages, countries, etc.) will be modelled as SKOS concept schemes.

Interoperability will be supported by the provision of dedicated APIs that allow the import and export of the content in standard formats (XML, JSON) and standard data models.

Also, specific APIs will be provided on request to serve interoperability with custom needs and requirements from relevant projects that act as Resource Aggregators.

#### 3.2.4 Increase data re-use

The **public** content within the EOSC Portal will be available for download and re-use with no restrictions. Access to data sets characterised as 'sensitive', is provided only to the appropriate and earlier agreed target groups and not via public unauthorised means. The public content will be available under permissive licenses, (CC-BY 4.0[5], CC-o or comparable) but certain conditions (e.g., Non-commercial use) and/or exceptions may also apply.

The EOSC software will be available under the GPL/A license. In the cases where specific resource information cannot be publicly shared, the reasons will be mentioned in their metadata descriptions (e.g., ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Concerning the time for which the data will remain re-usable, the platform will be maintained, and the content will be reusable for at least three years after the end of the project.

In addition, the EOSC Portal Catalogue will be uploaded, on a regular basis, in Zenodo, as well as in EU Open Data portal, thus, ensuring the long-term access and reusability of the data.

### 3.3 Allocation of Resources

ATHENA leads the implementation of the Provider layer of the EOSC Portal and the onboarding functionality for Providers. ATHENA is also responsible for the design, implementation, and evaluation of the AI software of the EOSC Portal.

CYFRONET is focusing on the development of the End-User layer and related components by providing the main components of service ordering system and will implement the majority of functionality related to demand side of EOSC Portal.

EGI, EUDAT, GÉANT, GRNET, JNP, OpenAIRE, OpenNode and INFN have been assigned the role to connect Providers and Resources to the EOSC and enlist them in EOSC Registry including the EOSC Portal thus focusing on operational aspects.

The developed software will be hosted in national-wide e-Infrastructure facilities. As such, the EOSC Portal will be hosted in GRNET's ViMa [6] service that offers IaaS resources primarily to the Greek research and academic community. No commercial hosting is anticipated.

### 3.4 Data security

Concerning data security, the platform content will be archived in a secure manner, via regular database backups, in the project's repository that will serve as the main backup area (the periodicity of the backup procedures will be defined according to the frequency of content updates).

Preservation and back-ups of the data are ensured by the provider's data centre preservation policies: daily backups, security updates, tightly controlled administrative access.

Both live and backed-up data are hosted in infrastructures offered by the consortium members and hosted in EU-located facilities.

Moreover, for long term preservation, data access and reuse after the end of the project, the EOSC Portal data will be available in Zenodo.

### 3.5 Ethics

Ethical and legal issues are tackled in the following way:

- In case of datasets containing personal data (e.g., from surveys), only anonymised data will be made openly available. The anonymisation procedure will ensure that no identification of natural persons is possible at all.
- In case of datasets containing data bound by specific agreements that prohibit further data dissemination, data will not be made available (but metadata will).

In accordance with legal restrictions, personal data concerning the EOSC Portal users' or Providers' data, are not shared in any manner.

## 4 Data related to EOSC Future Project Research Activities

### 4.1 Data Summary

Various research activities undertaken in the framework of the EOSC Future project (e.g., the EOSC Knowledge Hub) are expected to generate data which will be collected within the following context:

- Data collected via online surveys for the elicitation of user needs: these data and their metadata are processed and retained via trusted providers, controlled by consortium partners. The data privacy policy is available in the project's website [9].
- Interviews held for user requirements, minutes and notes of meetings, validation workshops, focus group meetings and webinars: these involve documents held and shared among the consortium partners according to their institutional practices and on the project's collaborative platform (Microsoft Office 365).
- Desk research and analysis: documents gathered during background research about best practices and comparable service typologies, platforms, repositories, and marketplaces.
- Project dissemination material: documents and publications presenting the findings, mappings, analyses, training material, tutorials and visualisations of the project, and text and images on project website<sup>5</sup> and social media channels.
- Project website and the information contained therein.
- Deliverables and financial reports produced by the Consortium: documents required by the project funder (European Commission) presenting the progress of the project.
- Consortium meeting notes, financial information, and other internal administrative data in common document formats (spreadsheets, text documents).
- User assessment results gathered from a targeted user base of stakeholders actively engaged in the process in the form of answers in questionnaires and notes on interviews.

### 4.2 FAIR Outputs

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

Documents and code underlying all EOSC Future research activities will be accompanied by standard general metadata and receive a DOI (Digital Object Identifier) number.

Access to the project's output will be provided through the EOSC Future project's site. Scientific publications will be also disseminated through the project website as well as through scholarly communication channels, e.g., Publishers/Journals web sites, Institutional Repositories, scholarly communication networks (ResearchGate, Google Scholar).

#### 4.2.2 Making data accessible

Depending on each deliverable's dissemination level, deliverables will be publicly available via the project's website, while others will be restricted.

Scientific papers and publications will be mainly available / accessible via the Publishers' web site according to the associated access method.

EOSC Future promotes 'Open Access' thus a machine-readable electronic copy of every publication is expected to be deposited in a suitable Open Access repository.

In the case of surveys containing any kind of personal data, only anonymised versions of the datasets will be openly available; raw data will be archived in closed or restricted access.

#### 4.2.3 Making data interoperable

Interoperability of the data collected by surveys, deliverables and publications will be achieved by the standard metadata required by OpenAIRE [8].

---

<sup>5</sup> <https://eoscfuture.eu/>



#### 4.2.4 Increase data re-use

The project's output will be made available under Creative Commons attribution license CC BY 4.0 for documents, reports, presentations etc., unless otherwise agreed and for justified reasons.

### 4.3 Allocation of resources

TGB is responsible for the collaborative platforms adopted by the project. Namely, EOSC Future relies on a suite of collaboration tools that supports the organization and management of the project. It comprises several dedicated mailing lists, a collaboration platform (including event calendar) as well as cloud-based storage solutions for the collaborative creation, management, and versioning of documents.

Microsoft Teams is used as collaborative space for documents that are 'living documents' or are still in status of draft and need therefore revisions by other partners. Once documents have been finalised and officially adopted, they will be archived in a project repository.

Project-related documents include, among others:

- Contracts (e.g., GAG, CA and Accession Forms, procurement contacts, etc.);
- Agenda and minutes of meetings;
- Dissemination material (e.g., press releases and clippings);
- Collection of templates for the generation of various materials (e.g., presentations, deliverables, reports, etc.).

In addition, Confluence and Jira have been provided and may be used by the project partners as agreed. The privacy policy governing the latter products is available in Appendix A EOSC Future Confluence and Jira Privacy Policy

The responsibility for managing data underlying EOSC Future research activities lies with the authors of the individual research studies who will reference or include the sources to where the data are stored.

### 4.4 Data Security

Deliverables, reports, or any other produced document will be primarily stored in the project's collaboration online server hosted by Microsoft's Office 365 suite that adheres to all common security IT standards thus minimizing the risk of unintended exposure to non-authorized users.

### 4.5 Ethics

Only anonymised data will be made openly available if datasets contain any personal data. However, if certain datasets contain data under specific restrictions that prohibit further data dissemination, they will be prevented from being publicly available either by requiring only authenticated and authorised access or by imposing Access Lists (ACLs) to the API endpoint(s) serving them.

## 5 IPR Management

While the EOSC Future workplan does not anticipate focused development activities, the service integration and interdisciplinarity support activities will likely generate valuable, identifiable IPR. The detailed strategy for knowledge management, protection and the exploitation of results is defined in the Consortium Agreement and further elaborated in Deliverable D1.3 -Project Implementation Report and Key Exploitable Results Report.

This will be devised with the focus on the objective of practical development and exploitation through provision of EOSC-Core services, while considering a) the principle of open-source licencing of the foreground and b) the legitimate interests of the beneficiaries where there may be good reasons for wider dissemination and academic publication of aspects of the work.

It is anticipated by the partners that the requirements for access to background IP and the protection and exploitation of foreground IP can generally be met by the approach contained in the Commission's Model Grant Agreement and the latest DESCA Model Consortium Agreement for Horizon 2020 projects. The plan for the use and dissemination of foreground IP will detail the intended approach beyond the lifetime of the project to make service delivery and development possible.

A specific aspect of the EOSC-Core is the anticipated need to integrate project results and the background and foreground IP with third-party components (access libraries, external services) that will become known only during the project lifetime. The issues arising from these kinds of developments will be evaluated, as part of the overall Exploitation and Innovation Management activity (WP1, Task 1.4). To simplify management of the different open-source licenses, they are divided into three categories:

1. Apache-style licenses requiring only an acknowledgement;
2. Linux-style IPR regime - GNU-style license, requiring that derived works be released under the same license and the ownership cannot be accurately determined (no alternative licensing is possible);
3. GNU-style license with known ownership.

It is expected that most of EOSC Future foreground would be licensed under variation 1, to minimise the barriers to uptake. Option 3 balances the principles of open development and legitimate interests of beneficiaries. Option 2 is acceptable primarily if the approach is 'inherited' from a crucial background or side-ground component. To ensure access for third parties, project results that are essential for implementing the interoperability of the EOSC Portal will be licensed under an open licence (Apache 2.0 or CC-BY 4.0, except where there are justified reasons to choose another approach).

The management of the IP generated by EOSC Future will be carefully monitored by the Strategy & Oversight Board (SOB) and the Project Management Board (PMB), to ensure that it complies with the obligations of the beneficiaries under the grant agreement about IPR, dissemination and use issues.

The PMB will ensure that the requirements for data collection and storage, community access to foreground IP and possible transfer of foreground IP are in place.

The Consortium Agreement is be formulated to support these obligations, with a particular emphasis on responsibilities for internal management of IP, identification of access to background IP for the purposes of carrying out the project and the arrangements for the use of the foreground IP generated. TGB via their role in WP1 will ensure that all relevant project outputs have appropriate IPR management and are licensed so that they can be exploited by the EOSC community and dealt at project end.

## 6 Conclusions

The DMP aims at providing a framework to safeguard the data collected, processed and/or generated during the project's activities across their entire lifecycle, while also making them FAIR.

In the framework of EOSC Future, the DMP is a living document and may be updated throughout the course of the project, considering its latest developments and available results. If necessary, additional ad hoc updates may be realised to include new data (e.g., on the allocation of resources), additional details (e.g., on the Ordering Process), and reflect modifications in the methodologies applied or other aspects relevant to the data management changes.

## Appendix A EOSC Future Confluence and Jira Privacy Policy

# EOSC Future Confluence and Jira Privacy Policy

Version 1.0, June 9th, 2021

<b>Name of the service</b>	EOSC Future Confluence and Jira
<b>Description of the service</b>	<p>The EOSC Future Confluence (wiki) and Jira (ticketing system) (hereinafter referred to as: "the service" ) support the EOSC Future project's activities. Personal data is used to provide access to the service with the proper access levels.</p> <p>This privacy notice describes how we, Technopolis Group Belgium, (hereinafter referred to as "we" or "TGB" or "the Data Controller"), collect and process data by which you can be personally identified ("Personal Data") when you use the service.</p>
<b>Data controller</b>	<p>Technopolis Group Belgium          Avenue de Tervueren 188A          1150 Brussels          Belgium</p>
<b>Data processors</b>	<ul style="list-style-type: none"> <li>• EGI Foundation: service operator and manager</li> <li>• CESNET: resource provider, sub-contracted data processor of EGI Foundation</li> </ul>
<b>Data protection officer</b>	<p>Agis Evrygenis, TGB Managing Partner  <a href="mailto:agis.evrygenis@technopolis-group.com">agis.evrygenis@technopolis-group.com</a>          Avenue de Tervueren 188A          1150 Brussels          Belgium</p>
<b>Jurisdiction and supervisory authority</b>	<p>Jurisdiction: BE, Belgium</p> <p>TGB's lead supervisory authority is the Belgian Data Protection Authority "Gegevenbeschermingsautoriteit – Autorité de protection de données". They can be contacted at <a href="http://www.dataprotectionauthority.be">www.dataprotectionauthority.be</a></p>

<b>Personal data processed</b>	<p>The service may process the following personal data:</p> <p><b>Identification data:</b></p> <ul style="list-style-type: none"> <li>● Name</li> <li>● Identification numbers (as provided by identity providers)</li> <li>● E-mail address</li> <li>● Affiliation</li> <li>● IP address</li> </ul> <p><b>Behavioural data:</b></p> <ul style="list-style-type: none"> <li>● Usage data</li> <li>● Technical logs with timestamps</li> </ul> <p><b>Data allowing conclusions on the personality:</b></p> <ul style="list-style-type: none"> <li>● Membership information on roles and groups</li> </ul>
<b>Purpose of the processing of personal data</b>	<p>The purpose of the collection, processing and use of the personal data mentioned above is:</p> <ul style="list-style-type: none"> <li>● To provide the service functions, i.e. to identify users and provide access to Confluence and Jira with the proper access levels.</li> <li>● To monitor and maintain service stability, performance and security</li> </ul>
<b>Legal basis</b>	<p>The legal basis for processing personal data is: Legitimate interests pursued by the controller or by a third party according to Art. 6 (1) (f) GDPR.</p>
<b>Third parties to whom personal data is disclosed</b>	<p>Personal data will not be used beyond the original purpose of their acquisition. If a 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 in a contract when using the respective function or service. In particular, the data you provide to us will not be used for advertising purposes.</p> <p>For the purpose given in this privacy policy, personal data may be passed to the following third parties:</p> <p><b>Within the EU / EEA:</b></p> <ul style="list-style-type: none"> <li>● Confluence and Jira users, space and project managers</li> <li>● 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</li> </ul>

	<p>necessary to provide the incident response capability where doing so is likely to assist in the investigation of suspected misuse of Infrastructure resources.</p> <p><b>Outside the EU / EEA:</b></p> <ul style="list-style-type: none"> <li>• Confluence and service users, space and project managers</li> </ul> <p>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>
<p><b>Your rights</b></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"> <li>• Information about your data stored with us and their processing</li> <li>• Correction of incorrect personal data</li> <li>• Deletion of your data stored by us</li> <li>• Restriction of data processing, if we are not yet allowed to delete your data due to legal obligations</li> <li>• Objection to the processing of your data by us</li> <li>• Data portability</li> </ul> <p>To access and rectify the data released by your home organisation (e.g. your university, research institute or any other identity provider), you should contact them.</p> <p>You can complain at any time to the supervisory Data Protection Authority (DPA) responsible for you. Your responsible DPA depends on your country and state of residence, of your workplace or of the presumed violation. A list of the supervisory authorities with addresses can be found at <a href="https://edpb.europa.eu/about-edpb/board/members_en">https://edpb.europa.eu/about-edpb/board/members_en</a>.</p> <p>You can contact the Data Controller's lead supervising authority using the contact details provided in the Jurisdiction and Supervisory Authority section.</p>
<p><b>Data retention and deletion</b></p>	<p>Your personal data associated with your account is kept as long as your account is active. Your account can be deactivated on request.</p> <p>The records of your use and technical log files produced by the service components will be deleted or anonymised after, at most, 18 months. As required by EC-funded projects, data added to Jira/Confluence will be retained for 5 years from the start of the project.</p>

<b>Data protection code of conduct</b>	<p>Your personal data will be protected according to the Code of Conduct for Service Providers, a common standard for the research and higher education sector to protect your privacy:</p> <p><a href="http://www.geant.net/uri/dataprotection-code-of-conduct/v1">http://www.geant.net/uri/dataprotection-code-of-conduct/v1</a></p>
<b>Security</b>	<p>We have put in place appropriate security measures to prevent your personal data from being accidentally lost, used or accessed in an unauthorised way, altered or disclosed. In addition, we limit access to your personal data to those employees, agents, contractors and other third parties who have a business need to know. They will only process your personal data on our instructions and they are subject to a duty of confidentiality.</p> <p>We have put in place procedures to deal with any suspected personal data breach and will notify you and any applicable regulator of a breach where we are legally required to do so.</p>

## References

- [1] H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. (2016). [online] Available at: [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).
- [2] eoscfuture.eu. (n.d.). EOSC Portal. [online] Available at: <https://eoscfuture.eu/>.
- [3] eoscfuture.eu, 2021, *Provider data model*. Available at: <https://eoscfuture.eu/sites/default/files/EOSC-Profiles-v3.00.pdf>.
- [4] eoscfuture.eu, 2021, *Resource data model*. Available at: <https://eoscfuture.eu/providers-documentation/eosc-provider-portal-resource-profile>.
- [5] Creativecommons.org. (2016). Creative Commons — Attribution 4.0 International — CC BY 4.0. [online] Available at: <https://creativecommons.org/licenses/by/4.0/>.
- [6] vima.grnet.gr. (n.d.). ViMa - About the service. [online] Available at: <https://vima.grnet.gr/about/info/en>.
- [7] www.eosc-hub.eu. (n.d.). EOSC-hub and the ESFRI cluster projects | EOSC Hub. [online] Available at: <https://www.eosc-hub.eu/eosc-hub-and-esfri-cluster-projects> [Accessed 23 Sep. 2021].
- [8] issue.openaire.research-infrastructures.eu. (n.d.). Core Data Model - OpenAIRE2020 - Wiki - D-Net project tracking tool. [online] Available at: [https://issue.openaire.research-infrastructures.eu/projects/openaire2020-wiki/wiki/Core\\_Data\\_Model](https://issue.openaire.research-infrastructures.eu/projects/openaire2020-wiki/wiki/Core_Data_Model) [Accessed 23 Sep. 2021].
- [9] Future, E. (n.d.). Privacy Policy. [online] EOSC Future. Available at: <https://eoscfuture.eu/privacy-policy/> [Accessed 23 Sep. 2021].