EOSC Future

D1.7

Open Research Data Pilot

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020, Grant Agreement number 101017536

Version 1.0
November 2023

D1.7 / Open Research Data Pilot

Lead by **TGB**
Edited by Lennart Stoy (TGB)
Reviewed by Paolo Manghi (OpenAIRE) & Athanasia Spiliotopoulou (JNP)

Dissemination Level of the Document

Public

Abstract

The EOSC Future project participated in the Horizon 2020 Open Research Data Pilot. The data management practices of the project were described in D1.4 EOSC Future Data Management Plan (DMP). At the end of the project, EOSC Future has collected and generated data both in the context of the setting up of the EOSC platform as well as in the context of different research activities implemented and/or supported by the project.

Data related to the EOSC portal is available in structured format via the EOSC Marketplace and via APIs created by EOSC Future. These follow structured metadata formats described in the EOSC Profile documentation. Data created in the context of other research activities is openly available for 17 datasets and findable via Zenodo. These datasets are also openly licensed. Other data created by the project was treated as confidential and is only available to the consortium partners.

Version History

Version	Date	Authors/Contributors	Description
V0.1	06/09/2023	Lennart Stoy (TGB)	Initiation – Proposed ToC – First draft
V0.2	10/11/2023	Lennart Stoy (TGB)	Collected input incorporated
V0.3	23/11/2023	Lennart Stoy (TGB)	Draft version available for internal review
V0.4	24/11/2023	Lennart Stoy (TGB)	Comments from internal review incorporated
V1.0	29/11/2023	Lennart Stoy (TGB), Ron Dekker (TGB), Mike Chatzopoulos (ATHENA)	Final Version submitted to EC

Copyright Notice

Table of Contents

List of Abbreviations 3

1 Executive Summary 4

2 Introduction 4

3 Data related to the EOSC Portal 5

4 Data and information related to research activities in EOSC Future 7

Table of Tables

Table 4-1: Open datasets findable through OpenAIRE Graph 7

Table of Figures

Figure 3.1: Service Order Monthly Statistics in the Operations Portal EOSC Instance (Source: M26) 6
Figure 3.2 Search result element with statistical data on downloads and views (Source: M26) 6

List of Abbreviations

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

1 Executive Summary

The EOSC Future project participated in the Horizon 2020 Open Research Data Pilot. The data management practices of the project were described in D1.4 EOSC Future Data Management Plan (DMP). At the end of the project, EOSC Future has collected and generated data both in the context of the setting up of the EOSC platform as well as in the context of different research activities implemented and/or supported by the project.

Data related to the EOSC portal is available in structured format via the EOSC Marketplace and via APIs created by EOSC Future. These follow structured metadata formats described in the EOSC Profile documentation. Data created in the context of other research activities is openly available for 17 datasets and findable via Zenodo. These datasets are also openly licensed. Other data created by the project was treated as confidential and is only available to the consortium partners.

2 Introduction

The EOSC Future Consortium set out with three overarching objectives, namely, to progress 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.

During the implementation of the project, EOSC Future has collected and processed various types of data for research activities supporting the three overarching objectives mentioned above. In line with the overall commitment to Open Science practices, the project laid out a plan to make this data as findable, accessible, interoperable, and re-usable (FAIR) as possible. These efforts have been described in detail in the D1.4 EOSC Future Data Management Plan.[1] As explained in the DMP, this included three major categories of data:

1. Data generated in operating the EOSC platform (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.

The objective of the present deliverable is to provide a succinct report on findability, accessibility, interoperability, and reusability of the data collected and generated by the EOSC Future consortium. Within the scope of this deliverable as summary of the application of the Open Research Data Pilot, we are briefly describing the status of data generated through the project, with a focus on Open Data. Section 3 describes the data generated in operating the EOSC. Data generated from communities connected to EOSC via the EOSC Exchange recorded in the EOSC Research Catalogue and included in this section. Section 4 reports on other data generated, for example in the context of the RDA open calls or the science projects.

During the course of the project, additional categories of data were collected and processed. This includes data generated and collected in the context of the EOSC Observatory and the ten EOSC Future Science Projects. Given the nature of the data as resulting from research activities not related to operating the EOSC platform or to data connected through EOSC via the EOSC Exchange, these are assumed to fall under category 2) and therefore handled according to best practices already in place and reported on in this deliverable under Section 4.

Other EOSC Future project outputs such as software and code, or publications are deemed largely out of scope of this deliverable and reported through their own respective deliverables (referenced, where appliable, in this document) and reporting modules in the Funding and Tender Portal.

[1] https://eoscfuture.eu/wp-content/uploads/2022/12/EOSC-Future-WP1-JNP-D1.4-Data-Management-Plan-2021-09-27.pdf

3 Data related to the EOSC Portal

This section reports on data generated, collected and/or processed as part of the project activities related to operating the EOSC platform (including Portal, supply/demand layers, supporting functions). Generally, this data is provided through the EOSC Portal web interface and dedicated APIs.

Interoperability of this data is supported through structured metadata, specifically for through EOSC Profiles (now in v4.0) which are available through the EOSC Future public wiki.[2]

The **EOSC Service Catalogue** provides a REST API interface for access to the Catalogues Providers, Resources, Controlled Vocabularies and other related entities.[3] Several APIs offer access to:

- Public Data Sources
- Public Interoperability Records
- Public Providers
- Public Resource Interoperability Record
- Public Resources
- Public Services
- Public Training Resources

All entries are stored and can be downloaded through the API as XML objects. Import of specific types of entries is also feasible via API. The complete EOSC API documentation is a living document, generated using frameworks compliant with OpenAPI Specification 2.0. Further information about the APIs offered to access entries was reported in D5.1a[4] and D5.1b.[5]

Entries for research products in the EOSC Research Catalogue are based on the OpenAIRE Graph.[6] As explained in D4.1b[7], this allows both access via APIs, which were adapted to offer search, browse, and navigation functionalities (i.e. link resolution) that take into account services and EOSC Interoperability Framework Guidelines; content in the Graph can be also accessed via regular data dumps.[8] This data set is licensed under Creative Commons Attribution 4.0 International (CC-BY-4.0).

Further **aggregated data on onboarded service resources and providers** is available online. Aggregated data on onboarded service resources is accessible via the URL https://providers.eosc-portal.eu/stats/resources, including, for instance the number of onboarded resources per month, their classification in scientific domains, categories of resources, access types and access modes, target users, as well as language. Information on maturity and funding is available as well. Further aggregated data on onboarded providers can be accessed via the URL https://providers.eosc-portal.eu/stats/providers. This data provides, among others, information on the number of onboarded providers per month, legal status, and classification of providers into scientific ESFRI domains.

Data related to the **portal usage**, such as Resource views and orders are presented as aggregated statistics for each Resource, as reported in Milestone M26.[9] This information is available via the EOSC Operations Portal (Figure 3.1) and data for downloads and views on the page of each resource (Figure 3.2).

[2] https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Portal+Profiles
[3] https://providers.eosc-portal.eu/openapi
[4] https://eoscfuture.eu/wp-content/uploads/2023/10/EOSC-Future-WP5-ICOS-D5.1a-EOSC-Front-Office-Design-Functional-and-Technical-Spec.-2021-09-30.pdf
[5] https://eoscfuture.eu/wp-content/uploads/2023/10/EOSC-Future-WP5-ICOS-D5.1a-EOSC-Front-Office-Design-Functional-and-Technical-Spec.-2021-09-30.pdf
[6] http://graph.openaire.eu
[7] https://eoscfuture.eu/wp-content/uploads/2023/10/EOSC-Future-WP4-EGI-D4.1b-Back-Office-design-functional-and-technical-specifications-2023-06-12.pdf
[8] https://zenodo.org/doi/10.5281/zenodo.3516917
[9] https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Future+Project+Milestones?preview=/47547242/101155230/EOSC-Future-CESSDA-MS26-Usage%20statistics%20for%20datasets%20(views%2C%20downloads)%20will%20be%20collected%20and%20made%20available.-2023-05-16.pdf

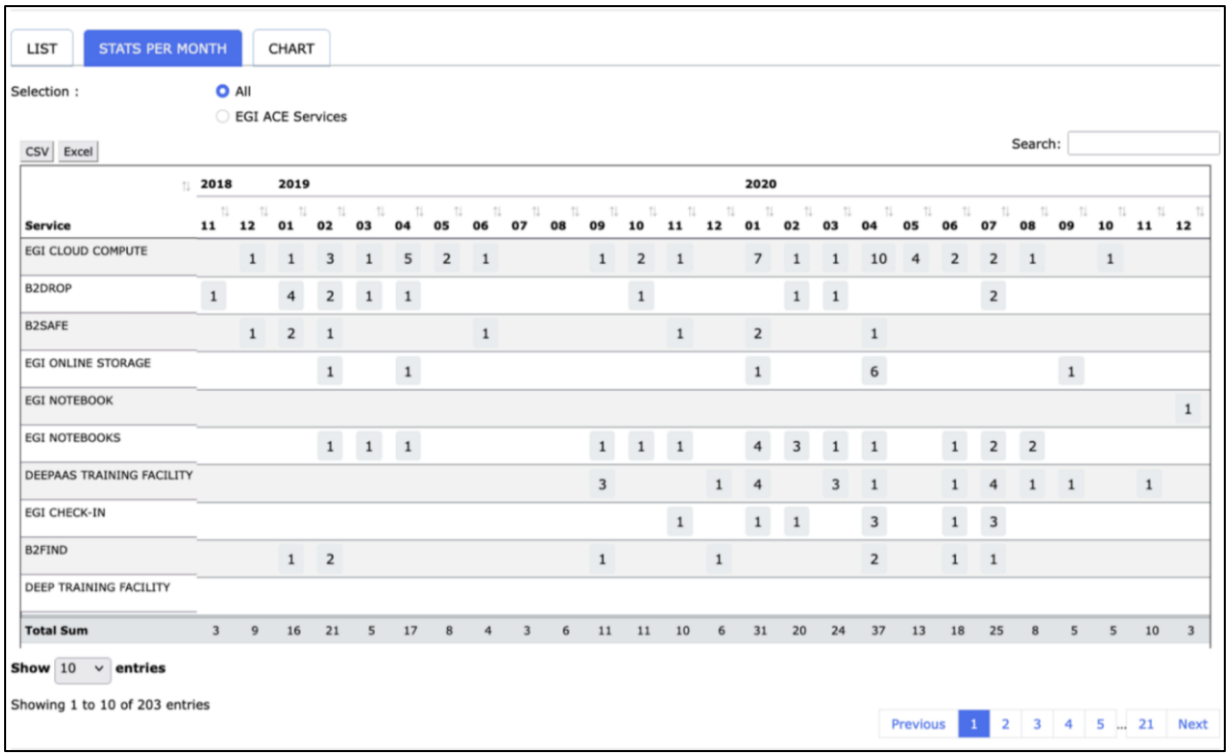

Figure 3.1: Service Order Monthly Statistics in the Operations Portal EOSC Instance (Source: M26)

Figure 3.2 Search result element with statistical data on downloads and views (Source: M26)

The **software and code** generated in the context of the development activities for EOSC platform is generally available through the dedicated GitHub repositories.[10] To further improve findability and re-usability of these results, the consortium is in the process of setting up a central mirror of these repositories as part of the handover process. This process is supported by the TCB and the Innovation Manager.

[10] See overviews in the relevant deliverables 3.3b and 5.3b.

4 Data and information related to research activities in EOSC Future

Research data generated by the EOSC Future Consortium has, where applicable, been made available through the all-purpose repository Zenodo[11]. Zenodo is indexed by the OpenAIRE Graph and therefore provides findability and interoperability of its metadata via the EOSC Catalogue and Marketplace. All data deposited by EOSC Future in Zenodo has been made accessible under open licenses.

Through the OpenAIRE graph also other project outputs (e.g., presentations, posters, reports, publications) are findable and accessible. This includes datasets stored on Zenodo, but also other indexed data repositories, such as NDEx.[12]

After the removal of duplicates in the Open Data reporting module in the EU Funding & Tender Portal, 17 individual datasets were included in the project reporting module at the time of writing this deliverable (Table 4-1 below). This included several updated versions of data sets.

Table 4-1: Open datasets findable through OpenAIRE Graph

#	Title	PID	License	Origin
1	Mpox Knowledge Graph - A comprehensive representation embedding chemical entities and associated biology of Mpox	https://doi.org/10.18119/n9sg7d	CC BY 4.0	WP6
2	Aventa AV-7 ETH Zurich Research Wind Turbine SCADA and high frequency Structural Health Monitoring (SHM) data	https://doi.org/10.5281/zenodo.8229750	CC-BY 4.0	WP6
3	Aventa AV-7 ETH Zurich Research Wind Turbine SCADA and SHM	https://doi.org/10.5281/zenodo.8223010	CC BY 4.0	WP6
4	Data of Survey on National Contributions to EOSC 2021	https://doi.org/10.5281/zenodo.7431678	CC BY 4.0	WP2
5	List of articles resulting from the Google Scholar search "graph based author name disambiguation" published after 1/1/2021	https://doi.org/10.5281/zenodo.8117573	CC BY 4.0	WP4
6	Björkö Wind Turbine Version 1 (45kW) SCADA	https://doi.org/10.5281/zenodo.8213270	CC BY 4.0	WP6
7	FAIR-CHO Citation Model Zotero Group Library Bibliography. Supplementary material	https://doi.org/10.5281/zenodo.8207653	CC BY 4.0	RDA Open Call
8	EU FAR Database: EU Funds absorbed by Romanian Municipalities 2016-2021	https://doi.org/10.5281/zenodo.7408376	CC BY 4.0	RDA Open Call
9	A2.2a Digital repositories data citation practices. Supplementary material	https://doi.org/10.5281/zenodo.8188806	CC BY 4.0	RDA Open Call
10	Enhanced Field Rotor Aerodynamics - ECN	http://dx.doi.org/10.5281/zenodo.10014602	CC BY 4.0	WP6
11	Enhanced Field Rotor Aerodynamics - ECN	http://dx.doi.org/10.5281/zenodo.8218005	CC BY 4.0	WP6
12	Aventa AV-7 (6kW) IET-OST Research Wind Turbine SCADA	http://dx.doi.org/10.5281/zenodo.8192149	CC BY 4.0	WP6
13	Björkö Wind Turbine Version 1 (45kW) high frequency Structural Health Monitoring (SHM) data	http://dx.doi.org/10.5281/zenodo.8229046	CC BY 4.0	WP6

[11] http://zenodo.org
[12] https://explore.openaire.eu/search/project?projectId=corda__h2020::256485716fdb9f5ca69007b7ca5a072b

14	Björkö Wind Turbine Version 1 (45kW) high frequency Structural Health Monitoring (SHM) data	http://dx.doi.org/10.5281/zenodo.8230330	CC BY 4.0	WP6
15	Björkö Wind Turbine Version 1 (45kW) high frequency Structural Health Monitoring (SHM) data	http://dx.doi.org/10.5281/zenodo.8229534	CC BY 4.0	WP6
16	Reproduction package for the paper "A LOFAR sample of luminous compact sources coincident with nearby dwarf galaxies"	http://dx.doi.org/10.5281/zenodo.10018524	CC BY 4.0	WP6
17	Data of Survey on National Contributions to EOSC 2022	http://dx.doi.org/10.5281/zenodo.10155993	CC BY 4.0	WP2

Results for the **EOSC Observatory**, including datasets and code, have been made additionally part of a dedicated Zenodo community, which provides further findability and visibility to these results.[13] This includes datasets for the 2021 Survey on National Contributions to EOSC and the 2022 Survey on National Contributions to EOSC and the respective questionnaires.

Besides research data, **six software packages** are also available through Zenodo, which are results from EOSC Future science project and RDA grants. These packages are in addition to software and code generated in context of the development of the EOSC platform which are being shared via GitHub.

Other activities conducted by the EOSC Future Consortium have collected data in the process of, for example, **gauging user assessments, needs and requirements**, or **feedback on project activities**. These data were treated as **confidential** and **are not released publicly**. As outlined in the original data management plan, 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). Moreover, they may only have minor benefits outside the scope of the project, for example for scientific re-use.

Further project information is made accessible through the project website's repository. This includes, for instance, **public project deliverables** and **communication assets**.[14] As outlined above, other results such as software, reports and publications can be accessed through the OpenAIRE Graph.

[13] https://zenodo.org/communities/eoscobservatory?q=&l=list&p=1&s=10&sort=bestmatch
[14] https://eoscfuture.eu/library/