EOSC Future

D3.4

Working Group Outputs and Mediated Requirements

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

Version 1.0
June 2023

D3.4 /Working Group Outputs and Mediated Requirements

Lead by ELIXIR/EMBL
Authored by Jonathan Tedds (EMBL/ELIXIR), Gavin Farrell (EMBL/ELIXIR), Alessia Bardi (CNR), Enol Fernández (EGI), Keith Jeffery (EPOS ERIC/ENVRI-FAIR), Daan Broeder (CLARIN/SSHOC), Giuseppe La Rocca (EGI)

Reviewed by Paolo Manghi (OpenAIRE / EOSC Future Technical Coordinators Board), Michelle Williams (GEANT / WP3 Lead) and Athanasia Spiliotopoulou (JNP)

Dissemination Level of the Document

Public

Abstract

This deliverable builds upon the document detailing the EOSC Future project Working Group methodologies framework produced by Work Package 3 - Task 3.1.1 Activity. The framework defined the structure for establishing and operating EOSC Future Working Groups. These temporary structures were designed to facilitate the creation of EOSC Interoperability Framework guidelines for horizontal services in areas of importance in the context of the EOSC Future project. Additionally, they were to provide input to EOSC Future project areas such as the EOSC Exchange in relation to their Working Group's area of focus. These Working Groups were to involve and leverage domain experts across the EOSC ecosystem for their input on their areas of expertise supported through regular meetings.

Four Working Groups were set up for areas of importance to produce EOSC Interoperability Framework guidelines for use by the wider scientific community on topics of relevance across the ESFRI defined EOSC Science Project Clusters. The contents of this document highlight these Working Group contributions to the EOSC Interoperability Framework under the four key areas of: Compute, Metadata, Research Product Publishing, and the EOSC Future Science Project Use Cases (Work Package 6).

This document focuses on the four individual Working Groups and details their operations, area landscaping undertaken, and concrete outputs intended to feed into the EOSC Interoperability Framework and wider EOSC Future project. It further details the coordination efforts, support, limitations, and lessons learned while employing the methodology for the Working Groups throughout the project to source content for the EOSC Interoperability Framework.

Version History

Version	Date	Authors/Contributors	Description
V0.1	15/03/2023	Gavin Farrell (EMBL/ELIXIR), Jonathan Tedds (EMBL/ELIXIR),	First draft - core document work
V0.2	21/04/2023	Jonathan Tedds (EMBL/ELIXIR), Gavin Farrell (EMBL/ELIXIR), Alessia Bardi (CNR), Enol Fernández (EGI), Keith Jeffery (EPOS ERIC/ENVRI-FAIR), Daan Broeder (SSHOC/CLARIN), Giuseppe La Rocca (EGI)	Second draft - WG work additions from pre-existing write ups
V0.3	03/05/2023	Jonathan Tedds (EMBL/ELIXIR), Gavin Farrell (EMBL/ELIXIR), Alessia Bardi (CNR), Enol Fernández (EGI), Keith Jeffery (EPOS ERIC/ENVRI-FAIR), Daan Broeder (SSHOC/CLARIN), Giuseppe La Rocca (EGI)	Third draft - WG refinements/new developments to be included e.g.: guideline progression
V0.4	17/05/2023	Jonathan Tedds (EMBL/ELIXIR), Gavin Farrell (EMBL/ELIXIR), Alessia Bardi (CNR), Enol Fernández (EGI), Keith Jeffery (EPOS ERIC/ENVRI-FAIR), Daan Broeder (SSHOC/CLARIN), Giuseppe La Rocca (EGI) + WP3, TCB Members	Fourth draft - whole refinement of document with input from: -Task 3.3 members -Work Package 3 members -Technical Coordinators Board (TCB)
V0.5	20/05/2023	Levente Farkas (EGI)	Data transfer Interoperability Guideline contributions
V1.0	30/06/2023	Jonathan Tedds (EMBL/ELIXIR), Gavin Farrell (EMBL/ELIXIR), Alessia Bardi (CNR), Enol Fernández (EGI), Keith Jeffery (EPOS ERIC/ENVRI-FAIR), Daan Broeder (SSHOC/CLARIN), Giuseppe La Rocca (EGI), Michelle Williams (GEANT)	Final Version submitted to EC

Copyright Notice

Table of Contents

Table of Tables

Glossary

EOSC Future project Glossary is incorporated by reference: https://wiki.eoscfuture.eu/x/JQCK

List of Abbreviations

Acronym	Definition
API	Application Programming Interface
EIAB	EOSC Interoperability Advisory Board
EIAC	EOSC Interoperability Area Chairs
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
IF	Interoperability Framework
FAIR	Findable, Accessible, Interoperable, Reusable
HPC	High Performance Computing
PID	Persistent identifier
RI	Research Infrastructure
SKG	Scientific Knowledge Graph
TCB	Technical Coordination Board (TCB)
TF	Task Force
WG	Working Group
WP3	EOSC Future Work Package 3: Architecture and Interoperability
WP6	EOSC Future Work Package 6: Science Projects

1 Executive Summary

This deliverable builds upon the document detailing the EOSC Future project Working Group methodologies framework defined by Work Package 3 (WP3) - Task 3.1.1 Activity [1] and is complementary to the overarching work of the EOSC Architecture and Interoperability Framework (WP3) [2] .

Section 2 of this document introduces and highlights the need for EOSC Interoperability Framework (IF) guidelines co-developed with scientific experts across the ESFRI defined EOSC Science Project Clusters [3] . It further details the deployment of the EOSC Future project Working Group (WG) methodologies framework to mediate the initial work towards potential EOSC Interoperability Guidelines across key areas of importance to EOSC.

EOSC Interoperability Guidelines were structured to be developed relating to the EOSC-Exchange platform [4] under the areas of: Compute, Metadata, Research Product Publishing, and the EOSC Future Science Project Use Cases (WP6). To facilitate this need, four individual WGs were set up correlating directly to these areas, each with a designated WG chair to champion and progress the group and contribute outputs to the EOSC IF.

The EOSC Interoperability Guidelines, as described in the EOSC Architecture and Interoperability Framework, are envisaged to be guidance on domain standards, APIs, policy frameworks, or any other related community formats to facilitate interoperability with EOSC. The EOSC-Exchange Interoperability Framework implicitly includes all standards, formats, and guidelines used in science, provided they meet the criteria defined by the EOSC IF governance.

Section 3 details the EOSC Future Task 3.3 WG coordination efforts employed to liaise with respective WG chairs and members of the Task 3.3 work.

Section 4 details the Compute Continuum WG [5] set up to create the basis of EOSC IF guidance in the area of scientific computing.

Section 5 details the Metadata [6] WG set up to create the basis of EOSC IF guidance in the diverse area of Metadata standards utilised in the ESFRI Science project Clusters. Of note is section **5.5.3** containing the Technical Options Proposed by the WG.

Section 6 details the Research Product Publishing WG [7] set up to create the basis of EOSC IF guidance on how to publish research products such as Jupyter Notebooks, data sets, R Shiny apps and other related scientific research products.

Section 7 details the Science Use Cases WG [8] set up to create the basis of EOSC IF guidance in relation to the challenges and success of the WP6 science projects and their onboarding for integration with the EOSC platform.

Section 8 details the limitations and learnings of the WG framework methodology employed. It also includes considerations and recommendations for the wider EOSC initiative from this ground work in the context of developing future EOSC Interoperability Guidelines.

Section 9 details areas requiring further EOSC IF horizontal guidelines development and potential groups that could be leveraged to facilitate this. This section also identifies existing scientific and technological expert groups which are currently engaged to develop EOSC Interoperability Guidelines through engagement and dialogue as part of the work of Task 3.3.

Section 10 details summary conclusions and recommendations based on the work of Task 3.3.

The final version of this document has solicited input directly from the members of EOSC Future WP3 - Task 3.3, the four WG area chairs, wider members of EOSC Future WP3, and members of the EOSC Future project Technical Coordination Board (TCB).

2 Introduction

2.1 EOSC Future, Work Package 3 and EOSC Interoperability Framework guidelines

The EOSC Future project's goal is to develop and implement a functional European Open Science Cloud for users and resource providers. Within the EOSC Future project, WP3 focuses on the development of the EOSC technical architecture and the integration of existing services and infrastructures. A key output of WP3 is the EOSC IF. The EOSC IF is intended to host a set of policies and guidelines that enable interoperability of resources and services and will facilitate service composability. These guidelines are intended for use by both the users of EOSC and the scientific resource providers. The planned guidelines that will populate the EOSC IF registry, are diverse in nature but can be categorised into three main guideline areas:

1. EOSC-Core Interoperability Guideline;
2. EOSC-Exchange Interoperability Guidelines (thematic);
3. EOSC-Exchange Interoperability Guidelines (horizontal).

2.2 EOSC Interoperability Framework guidelines explained

2.2.1 EOSC-Core Interoperability Guidelines

This category of guidelines describes specifications for the purposes of interoperating resources with EOSC-Core services [9] . EOSC-Core guidelines should provide context and description in order to provide technical instructions to scientific resource providers that would like to integrate their services and/or resources with (or be interoperable with) one or more EOSC-Core Services which include:

- EOSC Federated AAI;
- Monitoring;
- Accounting;
- Helpdesk;
- Metrics service;
- Resource registries;
- Provider Portal;
- Order management.

These guidelines are created by operators of EOSC-Core Services to provide guidance to scientific service and resource providers who wish to benefit from the functionalities that the EOSC-Core services offer.

2.2.2 EOSC-Exchange Interoperability Guidelines (thematic/community)

This category of guidelines refers to where thematic or community services and resources interoperate with each other, usually at a domain level such as Cluster specific. These are important at the EOSC Exchange level and for service/resource operators across all scientific Cluster domains. Prospective and in-development guideline examples include:

- Covid metadata guideline [10] - guideline in development;
- ARIA data access guideline [11] - guideline in development;
- Bioschemas guideline [12] - prospective guideline.

These guidelines are created by scientific community experts to provide guidance to their communities and highlight the importance of their interoperability efforts at the European wide EOSC level. They are further intended to increase awareness of the existing thematic and community accomplishments to facilitate/increase interoperability relating to specific domains e.g., within the EOSC Science Project Clusters.

2.2.3 EOSC-Exchange Interoperability Guidelines (horizontal)

This refers to where services and resources interoperate with each other across communities and infrastructures. These are important at the EOSC Exchange level to help EOSC facilitate cross-linking of services and resources while staying domain agnostic. These are of wider applicability than the thematic/horizontal guidelines and have a much wider scope. Examples could include:

- Persistent identifiers (PIDs);
- Schema;
- Metadata;
- Container technology;
- High performance compute (HPC).

These guidelines are created by scientific community experts to provide guidance to their own community and highlight their importance at the European wide EOSC level. However, this similarly makes them more difficult to develop and requires cross-Cluster engagement and representation to ensure they are accessible and widely applicable to a diverse range of intended guideline users.

2.3 EOSC Future Task 3.3's scope - Horizontal guideline development & Project input

EOSC Future Work Package 3 - Task 3.3's scope and work remit relate directly back to the EOSC IF. Its goal was to aid the population of the EOSC IF registry with specific guidelines. Specifically, this was through supporting and aiding the development of EOSC-Exchange Interoperability Guidelines (horizontal category). Certain areas of importance were pre-determined to be targeted for EOSC IF horizontal guideline development as underwritten in Task 3.3's initial Deliverable 3.1. 'Science Cases for Development of EOSC Architecture and Frameworks' [13] and from input of the evolving EOSC Future technical roadmap. Additionally, within the remit of the WGs was their interoperability focused input to areas of importance for the EOSC Future project. For example, the Metadata WG's output on metadata requirements and recommendations for the EOSC Exchange platform.

2.4 EOSC Future Task 3.3 - Working Groups Methodology

2.4.1 Working Group Methodology explained

As noted, the sourcing of EOSC Interoperability guidelines would require a large degree of effort from scientific and technological experts from across the EOSC ESFRI defined EOSC Clusters. The identification of a co-development model led to the creation of the EOSC Future project Working Group methodologies framework defined by Work Package 3 (WP3) - Task 3.1.1 Activity. This methodology was established to afford agility and flexibility to timeboxed activities that would bring experts together to discuss and further the development of EOSC Interoperability Guidelines. The WG methodology and open call for WGs led to the creation of four groups in total. Membership of the WGs was open to both members joining internally from within the EOSC Future project and also to those external to the project. Limited funding was originally budgeted to support the WG chairs efforts to lead the WGs where required. However, members external to the EOSC Future project usually contributed to the work in-kind.

2.4.2 Working Groups established

The four WGs were created based on the required outputs of Deliverable 3.1 and the EOSC Future Technical Roadmap to help facilitate the creation of EOSC IF horizontal guidelines were:

- Compute Continuum Working Group (**detailed in section 4);**
- Metadata Working Group (**detailed in section 5);**
- Research Product Publishing Framework Working Group (**detailed in section 6);**
- Science Projects Technical Alignment (**detailed in section 7).**

3 Task 3.3 and Working Group Operation Coordination

3.1 Working Groups overview

Detailed in *Table 3-1* is the high-level overview of the four WGs selected to run from the open call for WGs of Task 3.3.

Table 3-1: Working Groups Overview

Working Group	Chair(s)	Charter document	Planned operational duration	Actual operational duration
Compute Continuum	Enol Fernandez (EGI)	[14]	6 months	13 months
Metadata	Keith Jeffery (ENVRI-Fair) Daan Broeder (SSHOC/CLARIN)	[15]	6 months	12 months
Research Product Publishing Framework	Alessia Bardi (CNR)	[16]	6 months	16 months
Science Projects Technical Alignment	Giuseppe La Rocca (EGI)	[17]	6 months	18 months (ongoing)

3.2 Working Groups Monitoring

The WGs were monitored through the main Task 3.3 monthly calls. A total of 17 monthly meetings have been carried out to date to support the WGs on a monthly basis since their inception and in order to help support the development of prospective EOSC IF horizontal guidelines. All meetings were minuted and WG chairs encouraged to attend and report progress back to the main Task 3.3 group. Support and consultation were offered on the monthly calls to help align and converge on common issues and solutions. Several additional ad-hoc progress monitoring meetings took place between Task 3.3 leadership and the WG chairs as needed. As noted in section 8.1.2-Group duration issue, the operational dates referenced in Table 3-1 reflects the extended operational dates of the WGs.

3.3 Working Groups Engagement & Support

WGs were planned and primarily supported by the WG chairs through their own respective WG meetings. Where appropriate these were aligned with ongoing EOSC Association Advisory Group and Task Force activities of common themes and group representatives attended the meetings. Such meetings facilitated discussion and prioritisation during the development of specific EOSC Interoperability Guidelines through their milestones/goals written in their WG charter documents. Ad-hoc meetings were arranged between Task 3.3 leads and WG chairs as needed to support their activities. Task 3.3 leads also attended various WG sub-meetings as needed and regularly directly supported the Science Use Cases WG, given the wider range of stakeholders involved (typically 30-40 attendees) from across the EOSC Future Science Projects (Work Package 6). Engagement was also bidirectionally facilitated.

3.4 Working Groups Reporting & Output Capture

WGs nearing the end of their life span were tasked with capturing their accomplishments, recommendations and outputs in a written format. Templates were designed and distributed by Task 3.3 leadership in order to uniformly capture this. WGs were also provided with specific presentation templates to present to the EOSC Future Technical Coordinators Board (TCB) at their weekly meetings. The purpose of these WG presentations was to highlight the work accomplished to date or on completion and to receive actionable input from the TCB as to how best to orient their final stage of work.

3.5 Task 3.3 - Progression of EOSC Interoperability Guidelines

Task 3.3 members supported the progression of prospective EOSC Interoperability Guidelines through utilisation of the 'EOSC IF - EIAB & EIAC Charter' [18] which acts as the workflow for EOSC Interoperability

Guideline development through engagement with the relevant governing bodies of the EOSC IF (WP3 and the TCB).

3.6 Supplementary engagements & Actions for underserved areas

As detailed in **section 8** 'Working Group methodologies employed and EOSC contribution engagement learnings', Task 3.3 undertook additional activities beyond the scope of the original WG methodology to source EOSC IF horizontal guidelines not covered by any of the four primary Task 3.3 WGs. These supplementary engagements and actions were effective to support the sourcing of additional guidelines and have led to detailed learnings and recommendations based around the original WG methodology employed.

4 Compute Continuum Working Group

4.1 Working Group Information

Start Date: January 2022
Finish Date: December 2022
Working Group Chair: Enol Fernandez (EGI)
Membership

Table 4-1: Compute Continuum - Working Group Members

Member Name	Country Represented	Cluster Represented	RI/e-Infrastructure Represented	Institute Represented
Christos Arvanitidis	GR	ENVRI-FAIR	LifeWatch ERIC	LifeWatch ERIC
Enol Fernandez	NL	-	EGI	EGI
Juan Miguel González-Aranda	ES	ENVRI-FAIR	LifeWatch ERIC	-
Frederik Coppens	BE	ELIXIR	ELIXIR-BE	VIB
Björn Grüning	DE	ELIXIR	ELIXIR-DE	Uni of Freiburg
Christian Briese	AT	-	-	EODC
Raul Palma	PL	-	-	PSNC
Paolo Manghi	IT	-	OpenAIRE	OpenAIRE
Isabel Campos	ES	-	-	CSIC
Juanjo Dañobeitia	IT	-	EMSO ERIC	-
Ivan Rodero	IT	-	EMSO ERIC	-
Sandro Fiore	IT	-	-	University of Trento
Fabrizio Antonio	IT	-	-	CMCC
Jurry de la Mar	DE	-	-	T-Systems
Jerzy Konarski	PL	-	-	CloudFerro
Pablo Fernández	CH	-	-	CSCS

4.2 WG Background (Pre-operational phase)

4.2.1 General description of the need for the Working Group

The establishment of the Compute Continuum WG within the EOSC Future project was to address the need for the development of EOSC Interoperability Guidelines for improved access to computing resources in the EOSC Exchange. Currently, researchers face challenges in discovering and utilising this catalogue of resources effectively. The WG was set up to develop a standardised framework to enable seamless discovery and access to computing resources across various disciplines, intending to cover topics such as Edge, cloud, and HPC/HTC. By promoting interoperability and developing a better method for compute resource description, the Compute Continuum WG aimed to enhance researchers' ability to leverage computing resources for their scientific needs within EOSC and also develop this work into official EOSC Interoperability Guidelines.

4.2.2 Problems the Working Group was to address

The Compute Continuum Working Group was to address the problem of a lack of widely adopted standards and interoperability among technical solutions for providing access to computing facilities. To tackle this issue, the working group aimed to define a metadata schema as an extension of the EOSC profiles [19] , specifically designed to describe compute resources in the EOSC marketplace resource catalogue [20] This flexible and extensible specification would enable the description of services that provide access to generic computing resources, encompassing the complete compute continuum, including cloud, HTC, HPC, and potentially edge computing. It would also cover access to hardware accelerators like GPUs in these systems whenever available. By establishing these specifications, the Compute Continuum Working Group aimed to mitigate the absence of interoperability standards in the compute service area and enable the discovery and potential automation of compute services usage for user communities. With the aid of this metadata schema, user communities and individual users with specific scientific aims can be triaged and directed to the most suitable compute platform based on their requirements. Importantly, compliance with these specifications was not to necessitate technical changes for compute service providers, but rather the provision of compliant metadata during the EOSC onboarding process. The WG was also to focus on supporting container images and runtimes, as containers have emerged as a prevalent technology for packaging and running software in recent years.

4.2.3 Solutions the Working Group intended to output/provide recommendations on

The WG intended to tackle these challenges through the implementation of practical solutions. Firstly, the group was to focus on enhancing the discoverability of computing resources by developing a standardised framework that enables researchers to easily identify and access the resources they require for their specific research needs. Secondly, the working group planned to establish a technically interoperable specification that facilitates the effective description of computing resources across the compute continuum, targeting areas such as edge, cloud, and HPC resources. By providing researchers with improved access and standardised descriptions, the Compute Continuum working group aimed to empower researchers from various disciplines to leverage computing resources more efficiently within EOSC, ultimately advancing scientific pursuits and collaboration in this space. Where suitable this work was ultimately to translate into an EOSC Interoperability Guideline.

Key objectives included reviewing user requirements, classifying existing implementations, and defining metadata schema(s) for computing services. This would involve extending the EOSC Profiles, specifying interfaces and protocols, and capturing technical characteristics. The group was to publish compute services in the EOSC Resource catalogue and explore compute-storage integration. Through these efforts, the working group intended to improve resource discovery, accessibility, and utilisation in EOSC.

4.3 WG Summary of Actions (During the operational phase)

4.3.1 Working Group - operational activity

The Compute Continuum WG throughout its duration convened through regular monthly calls that served as a venue for collaboration and WG progress towards its planned solutions. These calls provided an opportunity for members to discuss updates, share insights, and address challenges together. Committed participants actively contributed to the discussions and decision-making processes to work towards its outcomes.

4.3.2 Working Group - work undertaken during the operational phase

The Compute Continuum WG undertook a thorough review of the existing computing resources available in the EOSC marketplace and analysed the specific needs expressed by participating communities. This comprehensive assessment encompassed a wide range of computing resources, including IaaS cloud, HPC, HTC, Notebooks, Containers, PaaS, FaaS, and interactive computing. Building upon this review, the working group formulated a set of attributes that are capable of effectively describing these diverse computing resources. To ensure seamless integration and compatibility within the EOSC framework, the group proposed an extension to the EOSC resource profile specifically tailored to accurately depict and characterise computing

resources. These efforts aimed to establish a standardised and comprehensive framework for describing computing resources within the EOSC ecosystem.

4.3.3 Working Group - challenges

The Compute Continuum Working Group encountered several challenges throughout its operation. One significant challenge was the absence of a standard API for accessing computing resources, which posed difficulties in establishing a unified approach. The group acknowledged that defining such an API might not be a desirable outcome, further complicating the process. Additionally, previous attempts to formulate similar schemas, such as GlueSchema, had limited success, emphasising the need for innovative approaches. The WG also faced the challenge of limited dedicated effort from most participants, resulting in slow overall progression and delays in activities. To address the complex nature of computing resources, the group focused on three primary types: HPC/HTC, IaaS Clouds, and Jupyter Notebooks. However, the inherent heterogeneity of these resources, especially when considering federations, presented challenges in modelling and standardisation. Despite these obstacles, the Compute Continuum Working Group remained committed to finding viable solutions and driving forward the development of a comprehensive framework for accessing and utilising computing resources within the EOSC ecosystem.

4.4 WG Highlighted Outputs/Recommendations (Post-operational phase)

The primary output of the WG outputs was the EOSC Compute Continuum Resource Profile Specification [21] . The contents of this document are a metadata schema as an extension of the EOSC profile v4.00 [22] to describe compute resources in the EOSC resource catalogue. This is a flexible metadata schema that allows for describing services providing access to generic computing resources covering as much as possible of the complete compute continuum: cloud, HTC and HPC and edge technology, including access to hardware accelerators (e.g., GPUs) in all these systems whenever available.

4.5 WG Summary Paragraph + Further work needed

The Compute Continuum Working Group successfully developed a functional schema specification for describing compute resources within the EOSC ecosystem. This metadata schema, built upon extensive collaboration and expertise, provides a standardised framework for accurately representing a wide range of computing resources, including HPC/HTC, IaaS Clouds, and Jupyter Notebooks. By establishing a common language and set of attributes, the working group has significantly enhanced resource discovery, accessibility, and interoperability within EOSC. The impact of this achievement on the EOSC ecosystem is significant, as it empowers researchers and user communities to efficiently identify and utilise the most suitable compute resources for their specific scientific aims, thus advancing research capabilities. Additionally, this standardised schema facilitates automated usage and enables seamless integration of compute services, ultimately promoting the overall effectiveness and efficiency of the EOSC infrastructure. It is planned as a next step to form an official EOSC Interoperability Guideline based around the specification document in order to increase its visibility and disseminate the importance of the metadata schema in the context of EOSC Exchange compute service provision.

5 Metadata Working Group

5.1 Working Group Information

Start Date: May 2022
Finish Date: May 2023
Working Group Chairs

- Keith Jeffery (EPOS)
- Daan Broeder (CLARIN)

Membership

Table 5-1: Metadata - Working Group Members

Member Name	Country Represented	Cluster Represented	RI/e-Infrastructure Represented	Institute Represented
Keith Jeffery	UK	ENVRI	EPOS	UKRI/BGS
Daan Broeder	NL	SSHOC	CLARIN	CLARIN ERIC
Mari Kleemola	FI	SSHOC	CESSDA	FSD/TAU
Carole Goble	UK	EOSC-Life	ELIXIR-UK	Manchester U
Nick Juty	UK	EOSC-Life	ELIXIR-UK	Manchester U
Enrique Garcia	CH	ESCAPE	CERN	CERN
Alasdair Gray	UK	EOSC-Life	ELIXIR-UK	Heriot Watt U
Jean-Francois Perrin	FR	PaNOSC	ESRF	ESRF
Wolmar Nyberg Akerstrom	SE	EOSC-Life	ELIXIR-SE	Uppsala U
Paul Millar	DE	PaNOSC	DESY	DESY
Brian Matthews	UK	PaNOSC	STFC/RAL	STFC/RAL
Tony Burdett	UK	EOSC-Life	ELIXIR-EBI	EMBL-EBI
Susanna Sansone	UK	EOSC-Life	ELIXIR-UK	Oxford U
Helen Parkinson	UK	EOSC-Life	ELIXIR-EBI	EBI Cambridge
Stian Soiland-Reyes	UK	EOSC-Life	ELIXIR-UK	Manchester U

5.2 WG Background (Pre-operational phase)

5.2.1 General description of the need for the Working Group

EOSC Profiles are metadata schemas describing EOSC Resources (services, research products, etc.), in order to enable convergence across communities, stakeholders, and types of resources. Accordingly, EOSC profiles have been defined, and are currently being maintained, via community and participatory efforts which aim at identifying the optimal metadata descriptions for the purpose of discovery, interlinking, and science monitoring. The approach is pragmatic, aiming at capturing metadata common across communities, disciplines, and across different types of resources, to ensure a fair coverage of actual metadata values across those while enabling the implementation of the aforementioned capabilities. The boundaries of this approach are, on the one hand by the misalignment of schemas, technical platforms, and practices in the different domains or regions; on the other hand, in the misalignment of the "local" roadmaps (research infrastructures, clusters, countries, ministries, libraries) with a European or global need of convergence.

The EOSC Thematic Science Project Clusters during the early stages of the project expressed the view that the EOSC Profiles should be further extended to enable enhanced forms of Finding, Accessing, Interoperating and

Re-using (FAIR) for the research communities. Such extensions could and should be regarded as a stimulus to highlight the need to invest on commons for the science good.

Initially the EOSC Profiles v3 [23] was primarily structured toward the description of services. In the subsequent update to EOSC Profile v4 this was extended to include Data Sources, Service Catalogues, Research Products, and Training resources. Research Products in the context of EOSC Future and EOSC Profiles can include various forms of research outputs, such as processed datasets, analytical results, visualisation products, models, algorithms, software, and other resources that have been derived from science.

Given the goal of the initially constructed EOSC Profiles to design a concise common data model, this group identified the need for additional metadata attributes to be provided by the EOSC Profiles.

Once implemented by the communities in their platforms and be made part of their practices and publishing workflows, these attributes would then be available to the EOSC Resource Catalogue and be leveraged by researchers in a way that would allow them to discover and utilise digital assets when using the EOSC Portal [24] .

5.2.2 Problems the Working Group aimed to address

The Metadata WG aimed to explore how the EOSC Profiles could be extended to enhance EOSC resource descriptions for the benefit of research communities in the Clusters and beyond. In particular, for when the EOSC Portal and its catalogue of offerings is to be presented as a (one stop shop) solution for resource discovery by researchers for research. The requirements with regards to specificity of the metadata was to be extended beyond what was offered by the current EOSC Profiles.

Secondly the WG was to act in a manner to bridge the gap between the highly diverse reality of metadata between infrastructures (e.g., Research Infrastructures), and address the possibility to develop and garner support for a common federated metadata infrastructure of use to research communities that could be offered by EOSC.

5.2.3 Solutions the Working Group intended to output/provide information and recommendations on

The first solution the WG intended to report on was a landscaping of the current situation of the diverse community metadata infrastructures, their metadata usages and their metadata practices characterised by the many differences with regards to their own organisation. On an organisational level this means understanding the consolidation efforts of metadata catalogues at different levels of the organisation, and to further understand the high diversity of the metadata schema used.

The second solution the WG intended to work on was ways to bridge the gap between that highly diverse reality, and the possibilities with regards to supporting a common federated metadata infrastructure, useful for research communities that could be offered by EOSC.

Further, it was intended to improve and extend the current metadata usage of the EOSC Portal and EOSC catalogue through recommendations for an enhanced metadata schema suitable for:

a. Research communities' users to utilise digital assets in EOSC;

b. Research community Research Infrastructures (RIs) to be able to onboard such metadata.

This was intended to facilitate the perspective for community RIs to join with their thematic catalogues of resources to form a federation of catalogues within the EOSC Portal and its overarching catalogue. This was to help build towards the EOSC vision to create for researchers a cross-disciplinary centralised resource access point and be the first stop for researchers to explore for resources from all disciplines.

5.3 WG Summary of Actions (During the operational phase)

5.3.1 Working Group - operational activity

The Metadata WG maintained regular communication through monthly meetings, supplemented by emails and regular organisational teleconferences led by the chairs. The group actively engaged with the RDA Metadata Interest Group [26] and its subgroups to ensure ongoing alignment and collaboration.

5.3.2 Working Group - work undertaken during the operational phase

The WG undertook a comprehensive survey to gather use cases and explored various metadata architectures across the stakeholders of EOSC. This work was to ensure that the EOSC metadata inventory was thorough and inclusive. The group remained open to receiving additional input for the inventory and incorporated community metadata infrastructure descriptions, referring to relevant reports such as from EOSC-Enhance [28] , to enhance the completeness of the survey inventory.

The WG collected use cases and examined metadata architectures, revealing significant heterogeneity. While ENVRI produced a Hub with a canonical rich metadata catalogue, it was acknowledged that this solution might not be optimal for other Clusters. Notably, SSHOC and EOSC-Life had multiple poles of aggregation with varying purposes and standards. However, there was some convergence towards accepting the concept of Scientific Knowledge Graphs (SKGs) as a basis for flexible interoperability and a federative metadata infrastructure. SKGs can be defined as fully connected graphs with relationships between base entities also as first-class objects. This was shown to have significant overlaps with the work of OpenAIRE in this area of knowledge graphs.

5.3.3 Working Group - challenges

The WG encountered major challenges in sourcing the relevant information needed for this work due to the voluntary nature of participation for most of its members, who were not able to be remunerated by EOSC-Future due to their participation from outside of the project. Attempts were made to facilitate transfer of budget to remunerate these members for their efforts, however, due to rules relating to cost claim eligibility it was not possible to progress this avenue of incentivisation. Consequently, this issue proved difficult to attract individuals with the necessary expertise to join the group and contribute to the relevant surveys and landscaping documents. Additionally, the perception of the temporary six-month WG activities as opposed to the longer-term and ostensibly impactful engagement activities such as within the EOSC Association Task Forces on similar topics such as Metadata [29] added to the complexities faced by the WG to progress their work to the intended outputs.

5.4 WG Highlighted Outputs/Recommendations (Post-operational phase)

5.4.1 Working Group summary outputs

The main output of the Metadata WG was the comprehensive EOSC stakeholders' metadata landscaping via collection of Science Project Cluster use cases and metadata architectures [30] . This output provides valuable insights into the diverse range of metadata practices and requirements across different Science Project Clusters within EOSC. By analysing these use cases and architectures, the WG gained a deeper understanding of the challenges and opportunities in developing a federated metadata infrastructure. This information now serves as a foundation for further discussions and actions aimed at enhancing interoperability and facilitating effective metadata management within the EOSC Future project.

Additionally, the Metadata WG provides summary observations and recommendations on metadata in the context of EOSC Profiles, EOSC Future and some potential solutions that can be considered for incorporation.

5.4.2 Working Group summary observations

1. EOSC resource catalogue: the current EOSC 'Profiles' of providers, services, and research products have a set of attributes which in their current form were designed for describing:

a. Services according to a service metadata profile designed to embrace the requirements of research infrastructures and e-Infrastructures;
b. Research products according to known standards (e.g. DataCite, DublinCore, JATS, etc.), to describe publications, research data, research software, etc.

Such profiles were designed to identify the most convenient (completeness, coverage) crosswalks from local communities to build a cross-domain EOSC Knowledge Graph. In some cases, such choices do not fully match the richness seen in research organisations or ESFRI RIs for their digital assets (software, services, data, data products, workflows). This WG therefore identifies the need to investigate the possibility to build a further overlay of the knowledge graph that takes into account community flavours and meets the fine-grained requirements of end-users of the EOSC Portal (researchers).

2. The EOSC catalogue (Marketplace) content is and will continue to be heterogeneous and, alongside enabling integrations to EOSC Core functionality, should also include entries for direct access to RI portals and assets.
3. The metadata of the EOSC catalogue is addressing the complexities of a diverse set or researcher requirements and ambitions. EOSC resources and interfaces need to facilitate the meeting of the FAIR Principles metric tests as detailed below:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.
R1. (meta)data have a plurality of accurate and relevant attributes.

5.4.3 Working Group summary recommendations

1. A full range of EOSC stakeholders should be closely involved in determining the metadata required for the various functionalities including discovery, contextualisation, utilisation (e.g., in workflows), intentionally including research, advice to public authorities, and commercial. Such stakeholders would include some of the relevant EOSC Association Advisory Groups and their Task Forces: the Semantic Interoperability Task Force is particularly relevant here.
2. The EOSC resource profiles (metadata schemas) of EOSC services and research products in the EOSC Resource Catalogue should be rich enough to be useful for interdisciplinary resource discovery and contextualisation, even if not all providers populate and update all the entities/attributes (objects/properties).
3. The EOSC Core service metadata schemas must be findable and interoperable such that they are reusable from the perspective of scientists willing to perform cross-disciplinary discovery.
4. Suitable crosswalks and semantic mappings should be enabled by EOSC and the communities, e.g. as a community exploitable service, to maximise interoperability with the major community approaches.

5.5 WG Summary Paragraph + Further work needed

5.5.1 Summary Paragraph

The WG successfully engaged into the initial delivery of a metadata inventory through surveying and landscaping activities. The inventory compensates missing information regarding community metadata infrastructures from previous studies such as EOSC Enhance: D4.3 Analysis of existing research data cataloguing efforts towards integrated discovery [31] . Additionally, some of the EOSC Future WP6 technical science projects are producing outputs related to metadata infrastructure. The WG chairs aim to provide feedback and align recommendations with these outputs where useful. Furthermore, the WG co-chairs had a productive meeting with members of the EOSC Future TCB and WP3, during which a proposal for potential solutions from the Metadata WG was requested. This proposal is presented immediately below in **section 5.5.2** and **section 5.5.3**.

5.5.2 Further work required - Introduction

Following a meeting with representatives of EOSC Future WP3 and the project's TCB, the Metadata WG was asked to produce proposals for architectural options that would be appropriate for the Scientific Clusters. Below are the proposals from Metadata WG to WP3 and the EOSC Future project more generally. The technical options are outlined below, and they could be progressively scheduled in different combinations (i.e., start with A1, then A2, then B3, B4...).

It is important to emphasise that the efforts required for the operation of thematic catalogues and the efforts required to onboard resources in the EOSC Resource Catalogue as a cross-disciplinary entry-point are both important for science and they cannot replace each other. Hence ESFRI RI and Community research services need the EOSC Portal to support the integration of thematic catalogues alongside any central catalogue where only singular resources can be on-boarded (with the associated administrative overheads).

5.5.3 Further work required - Technical Options Proposed

Currently, the EOSC Catalogue aggregates metadata profiles of services and research products from (human) providers, thematic and national catalogues, and other data sources (repositories, archives, SKGs, entity registries, etc.). The following technical extensions are recommended in the EOSC Catalogue and EOSC Portal.

A: EOSC Portal would provide a 'shopping mall' where 'shops' are individual Research Infrastructure resource portals

1. New EOSC Portal interface in the EOSC Marketplace, labelled e.g. "portals", that would contain metadata descriptions (service profile format) and the URL to a RI user portal.

2. An extension with (a) a faceted search interface for finding suitable portals and (b) an API to this zone of the EOSC portal so that programmatic access is possible. This could be improved if the portals all had in their metadata records terms from a controlled vocabulary. It was suggested those metadata vocabularies terms can be used e.g., to locate all portals that have/support "salinity" as a metadata term.

B: EOSC provides federated search

Beyond the current discovery capabilities over the EOSC Resource aggregation, the EOSC Core could support a federated search that exploits the capabilities in 1. and 2. to enable users to select several thematic catalogues and perform a distributed query via the provided APIs.

3. Clusters define a vocabulary of terms. All Cluster RIs describe their services locally (at the RI level) using this vocabulary in addition to whatever other metadata they may have. EOSC provides federated search over all the cluster RIs using APIs (to be developed and agreed) and accessed as URLs which are an attribute in the metadata describing the RIs from 2., returning the metadata records 'in full' where the common vocabulary is homogeneous, but the detailed other entities/attributes are heterogeneous. Note that this vocabulary of terms can be the superset of all RI specific metadata in a Cluster or community.

4. Utilizing this level of metadata at each RI for each service enables the federation of SKGs. The rich metadata available in the thematic catalogues can be exchanged between catalogues on a SKG entity basis between those catalogues that implement the SKG federation guidelines. The EOSC Catalogue will be compatible with SKG data models.

C: EOSC Resource Catalogue to enhance thematic catalogues and data sources

5. Based on 4., it becomes possible to harvest all the metadata and enable homogeneous search over the agreed terms and with application profile type extensions to the extended (heterogeneous) metadata.

6. Progressively enrich and harmonise the EOSC resource catalogue metadata schema to include thematic, FAIRness, reproducibility, interoperability flavors and stimulate the evolution of RI thematic

catalogues (one entity/attribute at a time). This will increase the number of entities/attributes that are homogeneous across all metadata of all RIs and all of their underlying data sources.

6 Research Product Publishing Framework Working Group

6.1 Working Group Information

Start Date: November 2021

Finish Date: March 2023

Chairs:

- Alessia Bardi (CNR)
- Jose Benito Gonzalez Lopez (CERN)
- Paolo Manghi (OpenAIRE)

Membership

Table 6-1: Research Product Publishing Framework - Working Group Members

Member Name	Country Represented	Cluster Represented	RI/e-Infrastructure Represented	Institute Represented
Chris Ariyo	FI	EOSC-Life	-	CSC
Andreas Czerniak	DE	-	OpenAIRE-Nexus	Bielefeld University Library
Paul Gondim van Dongen	NL	-	-	SURF
Georgios Kakaletris	GR	-	NEANIAS	National and Kapodistrian University of Athens
Raul Palma	PL	-	-	PSNC
Silvio Peroni	IT	-	OpenCitations	University of Bologna
Hans van Piggelen	NL	-		SURF
Mark van de Sanden	NL	-	EUDAT	EUDAT
Diego Scardaci	-	-	EGI	EGI
Jochen Schirrwagen	DE	-	OpenAIRE	Bielefeld University
Debora Testi	IT	-	-	CINECA
Raphaël Tournoy	FR	-	-	CNRS
Irena Vipavc	SI	-	-	University of Ljubljana
Deborah Grbac	IT	-	-	Università Cattolica del Sacro Cuore di Milano
Carl-Fredrik Enell	SE	-	-	EISCAT Scientific Association
Guido Aben	SE	-	-	SUNET
Ivan Heibi	IT	-	-	OpenCitations
Jorik van Kemenade	NL	-	-	SURF

6.2 WG Background (Pre-operational phase)

Open Science principles emphasise the timely publication of various research products to enable transparent assessment, review, reproducibility, and recognition of researchers' work in all its facets. However, the publishing process has become burdensome for scientists as they are frequently required to invest considerable time in publishing their articles, data, software, and other products across numerous institutional or thematic repositories. This includes scenarios like the initial publication of new resource products or the double-

publishing of research products to meet both institutional mandates and community practices. Unfortunately, this demanding task often remains incomplete, resulting in some products remaining unpublished and others displaying incomplete or imprecise metadata.

For the EOSC to act as enabler for Open Science practices, its IF should guide services of RIs and the Science Clusters of EOSC on how to implement (semi-)automated workflows for the deposition and consumption of research products. The benefits of this are that EOSC will:

- Reduce the burden of work on scientists, who can instead focus on their experiments and bypass the hurdle of the manual publishing process or download of existing data to reuse.
- Ensure publishing of science in a structured, complete, FAIR, reproducible, and undertaken in a positive and community-driven manner.
- Ensure high-quality metadata and fully-fledged monitoring/accounting of science, by systematically interlinking research products between themselves and with the services of which they are related to.
- Increase scientific repository visibility and usage to facilitate Open Science.

6.2.1 Problems the Working Group was to address

After communities were investigated, it was identified that there was a need for an integration approach for research performing services (from RIs and the Science Clusters) with repositories for research product deposition. The integration needed could ensure that research outcomes of such services are deposited automatically (with prior authorization of the users) into a given repository. This would realise a live end-to-end scientific workflow, from experimentation to publishing.

The limit of existing approaches is due to their bindings to a specific repository APIs and formats and when introducing multiple repositories as potential targets of deposition for a service this multiplies the problem, as bilateral interactions with the respective repository API must be established. For example, the Zenodo deposition API and the B2SHARE API are similar but differentiate in many ways. Therefore, a service willing to automate publishing into either repository would require the implementation and maintaining two different workflows, a key issue the WG hoped to address.

Another important aspect of Open Science is the possibility to re-use existing research products (e.g., research data), deposited in repositories and accessible via their persistent identifiers (e.g., handle, DOI, ark). However, there is no standard way a service can access the actual content behind persistent identifiers, as these typically resolve to the landing pages of the research products. This is another key issue faced in the research product publishing space. The lack of a standard for accessing the actual content identified by persistent identifiers makes the automatic consumption of research products difficult to implement and when possible, is limited to the persistent identifiers issued by a specific repository (e.g., the first prototype of the Data Transfer Service integrated in the EOSC EXPLORE Portal supports only DOIs from Zenodo).

6.2.2 Solutions the Working Group intended to output/provide recommendations on

The WG intended to recommend guidelines for the EOSC IF that would include protocols for research product deposition and content consumption, so as to enable EOSC services to seamlessly integrate with any compliant repositories.

6.3 WG Summary of Actions (During the operational phase)

6.3.1 Working Group - operational activity

The Research Product Publishing WG operated through online calls. The WG held seven online meetings since November 2021, involving participants from INFRA-EOSC projects, thematic, and horizontal research infrastructures. To facilitate communication among group members, a dedicated mailing list was established by OpenAIRE. The WG chairs created a collaborative working agenda document via Google Docs, which served as the main platform for collecting contributions from the WG members. These past activities showcased the

group's commitment to collaboration and finding innovative solutions to enhance the publication processes of research products within the EOSC ecosystem.

6.3.2 Working Group - work undertaken during the operational phase

The WG studied the status quo of (semi-)automatic publishing workflows by reviewing existing experiences about the design and implementation of (semi-)automatic deposition workflows of research assets produced in a RI to deposition in a repository service. The landscape analysis study revealed 15 implementations, which were then classified according to the following axes:

- Maturity level (design, prototype, beta, production);
- Type of submitter/receiver (repository to repository, analysis/research tool to repository, scholarly service to repository, publishing platform to repository);
- Implemented API/protocol (Zenodo API [32] , B2Share API [33] , SWORD [34]).

The WG went through a brainstorming session for the identification of scenarios, even beyond those implemented by the use cases reported for the landscape study, that would benefit from an IF guideline for research product deposition. A final set of five scenarios has been defined as a result of that session and the collaboration of the EOSC Future cross-infrastructure use cases on the Data Transfer Service.

The existing protocols and frameworks for deposition were scanned and a survey was conducted among the members to reach an agreement for proposal to the EOSC IF. From this activity two options were identified as relevant and are proposed for inclusion into the EOSC IF: SWORD protocol v3 [35] and a combination of COAR Notify [36] and Signposting [37] .

The group further agreed to develop EOSC Interoperability Guidelines for research product onboarding (i.e., the OpenAIRE guidelines) as a suitable metadata exchange format. In addition, SignPosting (in particular the part of "Publication Boundaries") is also proposed to be included to support the 5th scenario about the direct access to actual content behind a persistent identifier (proposed by the group working on the EOSC cross-infrastructures use cases).

All of the above listed work is currently in the progress to be refined into usable EOSC Interoperability Guidelines for wider use by the EOSC researcher community. Two guidelines are in draft stage as of May 2023 and are detailed as follows:

- Guidelines for accessing content via PIDs [38]
- Guidelines for research product deposition [39]

6.3.3 Working Group - challenges

The Research Product Publishing Working Group encountered variations in the boundaries of the topics to target due to diverse perspectives and areas of expertise among the participants. Some participants emphasised protocols for file transfers, while others focused on protocols for metadata exchange. Additionally, there were discussions regarding domain-agnostic versus domain-specific metadata formats. After deliberations, the WG reached a consensus not to recommend domain-specific formats or formats tailored to specific types of products, such as Jupyter Notebooks. Instead, the group decided to concentrate on general-purpose formats and defer community-specific decisions to the respective communities.

As only a few WG members possessed direct hands-on experience in implementing the suggested protocols, additional time was required to involve colleagues with relevant expertise. This collaborative effort aimed to foster a more comprehensive understanding of the potentials and limitations of the protocols under consideration.

6.4 WG Highlighted Outputs/Recommendations (Post-operational phase)

6.4.1 Working Group summary of outputs and recommendations

1. Landscape study undertaken and scenarios document [40] [40]

   a. The document noted provides an overview of existing experiences, at any maturity level, about the design and implementation of (semi-)automatic deposition workflows of research assets and describes five generic scenarios that would benefit from an EOSC Interoperability Guideline for research product publishing.
2. Recommended protocols for research product deposition in push mode:
   a. SWORD Protocol v3.0.
3. Recommended protocols for research product deposition in pull mode:
   a. COAR Notify + Signposting: the service can inform the repository that something new is available at a given accessible location (with COAR Notify) and the repository can then use the Signposting protocol (implemented by the service) to know where to get the content and metadata for the deposition.
4. Recommended protocols for accessing the content behind a PID:
   a. SignPosting (publication boundary) [41]
5. Recommended metadata exchange formats:
   a. Latest versions of the OpenAIRE guidelines [42] - these are EOSC guidelines for onboarding of research products;
   b. Other community specific format may be adopted in addition to these OpenAIRE guidelines.

Additional details and examples available in the WG's summary recommendation document [43]

6.5 WG Summary Paragraph + Further work needed

The WG successfully undertook landscaping with regards to research product publishing across EOSC stakeholders. The landscaping study revealed several implemented use cases of various maturities and generic scenarios that could benefit from the existence of an EOSC Interoperability Guideline for aiding their research product publishing. Two options were identified as relevant and are in development as EOSC Interoperability Guidelines for inclusion into the EOSC IF registry. These include the SWORD protocol v3 and a combination of COAR Notify and Signposting. The EOSC guidelines for research product onboarding (i.e., the OpenAIRE guidelines) are further in these guidelines suggested as metadata exchange format. In addition, SignPosting is also proposed to be included to support another research product publishing scenario for the direct access to actual content behind a persistent identifier. The WG after the publication of its guidelines is keen to monitor and gather suggestions, comments and feedback from EOSC partners and research communities about the proposed frameworks and their known limitations (e.g., complexity of SWORD v3, low adoption, partial implementation).

Finally, further work is needed from the implementation point of view of these EOSC Interoperability Guidelines for their refinement. Detailed are the progression of some implementations:

- OpenAIRE is including the implementation of the framework in its roadmap to improve the metadata integration with Zenodo and other repositories.
- HAL and Episciences have implemented the COAR Notify and Signposting protocols in production, while other use cases, including the one between HAL and Peer Community In are planned to be implemented during summer 2023.
- Episciences and Software Heritage are working on the implementation of COAR Notify in the context of the FAIRCORE4EOSC [44] EC funded project (first version planned for Dec 2023).

7 Science Projects Technical Alignment

7.1 Working Group Information

Start Date: September 2021
Finish Date: On-going (until project end)
Chairs:

- Giuseppe La Rocca (EGI)
- Jonathan Tedds (ELIXIR/EMBL-EBI)

Members
The Technical Alignment with Science Projects Working Group was composed of:

- Representatives from all the Science Clusters (ENVRI-FAIR, ESCAPE, Life Science, PANOSC, SSHOC)
- Members from the INFRAEOSC-07 projects including C-SCALE, DICE, EGI-ACE, GEANT, OpenAIRE-Nexus and RELIANCE
- Technology providers from EOSC Future WP4 and WP5

Table 7-1: Science Projects Technical Alignment - Working Group Projects

Project Name	Cluster Represented	Cluster Participants	E-Infrastructure Participants
Dark Matter	ESCAPE	Caterina Doglioni Ian Bird	Gianni Dalla Torre
Extreme Universe	ESCAPE	Ian Bird Elena Cuoco	Gianni Dalla Torre
Smart Cities	SSHOC ENVRI FAIR	Knut Kalgraff	Zdenek Sustr
ARIA	SSHOC EOSC Life	John Shepherdson Claudia Alén Amaro Marcus Povey	Mark van de Sanden
COVID-19 metadata findability and interoperability in EOSC	EOSC Life SSHOC	Antje Keppler Carazo Jose Maria Wolfgang Fecke	Alessia Bardi Raul Palma
NIS-Impact	ENVRI FAIR	Christos Arvanitidis Miguel González	Stefan Reimond
Dashboard on the State of the Environment	ENVRI FAIR	Dick Schaap Alex Vermeulen	Stefan Reimond
Tracing Bio-Structures	PANOSC	Gianluca Santoni Patrick Fuhrmann Franck Schluenzen Miguel Gonzalez	Giuseppe La Rocca
Dynamics of Biological Processes	PANOSC	Gianluca Santoni Patrick Fuhrmann Franck Schluenzen Miguel Gonzalez	Giuseppe La Rocca

7.2 WG Background (Pre-operational phase)

7.2.1 General description of the need for the Working Group

RIs, such as the ones on the ESFRI roadmap [45] , are mainly characterised by the huge data volumes they generate and handle, and further by the innovative applications and tools that they bring together in order to ensure effective research data exploitation by thousands of researchers across scientific disciplines and research projects. In the EOSC ecosystem, the ESFRI RIs are organised in five thematic Science Clusters. These Science Clusters play an important role for the provision of data and services to their research communities.

The EOSC Future project has selected Science Projects from the thematic Science Clusters to demonstrate how, by facilitating their integration with the EOSC Core and other horizontal services offered by EOSC, they can address major science and societal challenges and promote cutting-edge research, innovation and knowledge transfer in Europe, and as a consequence, support Horizon Europe's missions within the EOSC. From a technical perspective, the integration of these Science Projects with the EOSC core and horizontal services will provide important assets to enrich the EOSC Exchange through research data, workflows, thematic data analytics tools and advanced data management solutions.

7.2.2 Problems the Working Group was to address

The main objectives of the Technical Alignment with Science Project WG were to create a discussion forum between the technical partners of the project and the Science Projects to:

- Assess the technical requirements of the Science Projects;
- Identify gaps in the project solutions and EOSC architecture;
- Pilot integrations with EOSC Core and horizontal services;
- Keep Science Projects informed about the technical achievement of the project (from all technical WPs);
- Oversee how the science projects adopt project technical outputs;
- Where useful to do so based on the work of their projects, produce an EOSC Interoperability Guideline.

7.2.3 Solutions the Working Group intends to output/provide recommendations on

With regards to tangible EOSC Interoperability Guidelines in the context of Task 3.3 the WG has identified three initial EOSC IF horizontal guidelines planned for development.

- SSHOC - Access Management for Distributed RIs: ARIA ADAM guideline - developed and awaiting onboarding by John Shepherdson (CESSDA) & Marcus Povey (Instruct).
- EOSC-Life, MetaCOVID project: Proposal for a COVID metadata interoperability guideline (expected during June/July 2023) as proposed by Christian Ohmann (ECRIN).
- ENVRI-FAIR & SSHOC, SmartCities project: under consideration via development request, with hoped delivery during Q3 of 2023.

There is scope for further guideline development from some of the other projects once they reach maturity phase.

7.3 WG Summary of Actions (During the operational phase)

7.3.1 Working Group - operational activity

The Technical Alignment with Science Projects WG entered its operational phase in September 2021. To streamline communication, a dedicated mailing list was established for WG members. Additionally, a shared folder hosted on a Google Drive was created to store relevant slide decks and presentations used during meetings. To ensure the smooth technical integration of the EOSC Future Science Projects with the EOSC Core and the INFRAEOSC-07 projects' horizontal services, technical experts known as e-Infrastructure contacts were assigned to each Science Project. These experts took on the responsibility of liaising with the Science Projects, aligning use cases with appropriate service providers, overseeing integration plans, and reporting on achievements, and lessons learned. Monthly meetings provided a platform for Science Projects and e-infrastructure contacts to present progress with integration plans and address any blocking issues.

Throughout the project's duration, the WG's scope expanded to include members from T6.2 - "Integration of EOSC-Core Services into European Research Practice" who provided regular updates on the status of service onboarding in EOSC. Additionally, members of EOSC Future WP10 offered updates on the dissemination plans. Since the initiation of the WG, a total of 19 online meetings have been organised to facilitate collaboration and progress within the group.

7.3.2 Working Group - work undertaken during the operational phase

The WG, in collaboration with the assigned e-Infrastructure experts and Technology providers (from EOSC Future WP4 and WP5), provided assistance to support the technical integration with the EOSC Core and horizontal services provided by the INFRA-EOSC-07 projects.

7.3.3 Working Group - challenges

The main challenges faced by the WG during the operation phase are listed below:

EOSC Integration plans:

- Different levels of maturity from the various Science Projects.
- For some of EOSC Future Science Projects, their integration plans to integrate with the EOSC Core/horizontal services were not clear, or not foreseen in the short term.

Amendments:

- The administrative delays with the ratifying and signing of the project amendments has impacted on the official kick-start of the work plan of the SSHOC Science Projects. This has led to delays in their work plan execution.

On-boarding in EOSC:

- Several issues were raised during the onboarding of the Science Cluster services in the EOSC Portal. In particular the main issue was related to the upload of the service providers' privacy policy and terms of use which were considered optional in the previous release of the Portal, but that were requested in order to complete with the registration workflow. These issues have been solved with the recent new release of the onboarding procedure and the update of the workflow.
- The overall procedure for onboarding services in EOSC still remains complex and difficult to follow for most smaller providers. More flexibility on the category classification (data resource vs resource catalogue) is needed and under development to resolve this blocker.
- The services developed by many of the Science Clusters are the main result of collaborative efforts between different partners sometimes formalised in a Memorandum of Understanding (MoU) but without a specific legal entity formally responsible for the service. For this reason, the on-boarding of new services in EOSC under the umbrella of a single legal entity is still an issue and a show-stopper.
  - An example of this issue can be seen in the case of the ELIXIR Europe RI [46] of the EOSC Life Cluster which hosts approximately 500-member state provided bioinformatics services [47] which range from databases, to registries, tooling, and beyond. All of this content could be in theory onboarded to EOSC for researcher discoverability and usage. However, as ELIXIR's legal status is provided by the EMBL [48] it makes this action difficult to directly onboard services due to this requirement.
  - Further RIs may provision shared service from between several member institutions such as the RDMkit [49] which is a service supported and provided by multiple ELIXIR RI member states. This growing federated service delivery makes it even more difficult to attribute sole legal entities.

Onboarding as a certain resource category/type:

- EOSC Profiles were dynamic and updated as per needs by service/resource providers during the project to accommodate data sources This changed the requirements and integration/onboarding of certain Science Projects that were newly categorised as data source.

7.4 WG Highlighted Outputs/Recommendations (Post-operational phase)

7.4.1 Working Group summary of outputs and recommendations

The main outputs produced by the WG, grouped by categories, are listed below:

1. **Showcase Integration Story (SIS)**:
   a. ENVRI - Climate Change Impact on Biodiversity and Ecosystems in Europe (1) [50]
   b. ENVRI - Dashboard on the State of the Environment (2) [51]

c. EOSC-Life - Imaging Data in EOSC - COVID-19 as Demonstrator (4) [52]
d. ESCAPE - Dark Matter SIS (5) [53]
e. ESCAPE - Extreme Universe and Gravitational Waves (6) [54]
f. PaNOSC - Tracing Biostructures (7) [55]
g. PaNOSC - Serial Crystallography SIS (8) [56]
h. SSHOC - Access Management for Distributed RIs (10) [57]

2. **Status of the integration with the EOSC Core/horizontal services**:
   a. **PaNOSC - Serial Crystallography:**
      i. AAI: completed;
      ii. Cloud resources (offered by EGI-ACE): completed;
      iii. Presented during the first project review in June 2022.

   b. **ENVRI-FAIR - Dashboard on the State of the Environment:**
      i. AAI: completed;
      ii. Helpdesk, Monitoring: in progress;
      iii. Cloud resources (offered by EGI-ACE): completed;
      iv. Demo during the 2nd. project review in Nov. 2022.

   c. **ESCAPE - Dark Matter:**
      i. ESCAPE AAI federated with the EOSC AAI: completed;
      ii. Cloud resources (offered by EGI-ACE): completed;
      iii. Demo during the 2nd. project review in Nov. 2022 [58]
   d. **EOSC-Life - EU-OpenScreen:**
      i. EGI Notebooks (offered by EGI-ACE): completed.

   e. **EOSC-Life – MetaCOVID:**
      i. Performed interviews with the RIs on the use of contextual metadata.
      ii. Completed (preparatory TC + interview):
         1. ECRIN;
         2. BBMRI;
         3. CESSDA (TAU/FSD);
         4. CLARIN (& UA);
         5. EATRIS;
         6. EU-OPENSCREEN.
      iii. EOSC-Life - MetaCOVID: Proposal for metadata interoperability guideline based on interview landscaping listed above (e.ii.1-6) (approved by EIAB & EIAC - expected delivery June/July 2023).
      iv. OpenAIRE and ECRIN defined a crosswalk between the ECRIN model and the OpenAIRE guidelines (see this spreadsheet) [59] ECRIN is implementing an API that could be used by OpenAIRE to harvest the metadata records.
      v. EuroBioImaging in contact with the OpenAIRE Aggregation team.

   f. **EOSC-Life – EuroBioImaging:**
      i. Developed Tool and Workflow Development to Support Next Generation File Format OME-Zarr [60] ;
      ii. The tool has been onboarded to the WorkflowHub.

   g. **EOSC-Life - Instruct-ERIC:**
      i. Completed the integration of the 3DBionotes-WS and COVID19 Structural Hub with the Helpdesk;

ii. The integration with the Monitoring service is on-going.

h. **SSHOC - ARIA Science Project:**
   i. Initial pool of resources has been allocated at IN2P3-IRES[1].

3. **On-boarding in EOSC:**
   a. PaNOSC:
      i. VISA platform (completed) [61]
   b. ESCAPE:
      i. ESCAPE Virtual Observatory (in progress);
      ii. ESCAPE Open-source Scientific Software and Service Repository onboarded as DataSource (completed);
      iii. CERN/HEP Open Data Portal (in progress).
   c. ENVRI-FAIR - Dashboard on the State of the Environment:
      i. EOSC Future Dashboard for Environmental Indicators (completed) [62]
   d. EOSC-Life - Instruct-ERIC:
      i. 3DBionotes-WS on-boarded in the EOSC Marketplace (completed);
      ii. BioModels Monkey Pox integrated as Research Data Source (completed).
   e. EU-OpenScreen:
      i. KG resources onboarded to EOSC services (completed):
         1. Monkeypox Knowledge Graph: A semantic resource embedding biological and chemical entities [63]
         2. COVID-19 Knowledge Graph: A semantic resource embedding biological and chemical entities [64]

4. **Dissemination:**
   a. ENVRI-FAIR - Dashboard on the State of the Environment:
      i. Dissemination at ENVRI-week and EGU2023 event [65]
   b. EOSC-Life - MetaCOVID:
      i. Internal workshop was organised on 25 April 2023 with participation from ECRIN, EATRIS, BBMRI, EU-OPENSCREEN, CESSDA, CLARIN. The main objective of this workshop was to facilitate the alignment on the use of metadata, and finalise the Deliverable - "Characterising contextual metadata in selected Research Infrastructures: A summary report from semi-structured interviews".
   c. Smart Cities:
      i. Dissemination at EGU2023.

7.5 WG Summary Paragraph + Further work needed

Since the beginning of the Technical Alignment with Science Projects WG, it has organised and continues to run online meetings involving representatives from all of the Science Clusters, members from the INFRAEOSC-07 projects, and technology providers (from WP4 and WP5). During the online meetings, organised on a monthly basis, members of the Science Projects provide a round-table status report of the integration plans of the Science Projects with the EOSC Core and horizontal services provided by the INFRAEOSC-07 projects. Experts from the projects (WP4, and WP5) and service providers from the INFRAEOSC-07 projects provide assistance during the piloting activities. After 19 months of running activities, the WG achieved several accomplishments:

- Various integrations with the EOSC core services;
- Various integrations with the horizontal services;
- On-boarding;
- Showcase Integration Stories (SISs) were produced;
- Planned development of three EOSC Interoperability Guidelines - in various stages of completion:

[1] https://www.in2p3.cnrs.fr/fr

- ARIA ADAM guideline (close to EOSC IF registry onboarding);
- MetaCOVID guideline (early stages of development);
- SmartCities guideline (under consideration of guideline request).

However, as noted in **section 7.3.3** Working Group - challenges, there still remains work to streamline the Cluster service onboarding and integrations. Through highlighting these challenges, it is hoped to help guide the development of solutions to mitigate such blockers to service onboarding and integration.

8 Working Group methodology employed - challenges, learnings & recommendations for sourcing EOSC Interoperability Guidelines

8.1 Challenges of using WGs as vehicles for EOSC IF Guideline development

As highlighted throughout Deliverable 3.4, despite various successes in accomplishing EOSC IF horizontal guideline development, the WGs methodology framework employed was an imperfect solution for the complex and pioneering task. Below are the key limitations of the WG methodology employed and learnings/recommendations to mitigate such factors in the future when considering methods of further population of the EOSC IF registry with prospective horizontal guidelines.

8.1.1 Pioneers of EOSC-IF horizontal guidelines, guideline template flexibility & diverse WG aims

The WGs of Task 3.3 were kicked off during the project at the same time as the EOSC IF of WP3 was maturing. The exact remit and scope of the guidelines to be produced by the WGs was not set in stone and as a result each WG had its unique perspective and intentions to produce useful interoperability focused outputs. This pathfinding and pioneering work have led to several varied and useful outputs but not horizontal EOSC Interoperability Guidelines in all cases.

Flexibility of EOSC Interoperability Guidelines template adherence where fields are not applicable/conducive to guideline development has also been noted through WG usage. It is now considered during guideline development for the template in use to be flexible during its development under the guidance of the EOSC Future EIAC and EIAB. This helps to avoid one-size fits all approach where not useful and limiting to EOSC Interoperability Guideline development.

Finally, EOSC Interoperability Guidelines were not the sole focus or scope of the WGs and for example in the case of the Metadata WG, recommendations and bidirectional dialogue with the project TCB was a key focus and output of the groups work. In this case an EOSC Interoperability Guideline document for public dissemination was not the intended output.

8.1.2 Group duration issue

As noted in *Table 3-1* the intended six-month duration of the WGs was insufficient. In most cases at least twelve months were needed, and this further highlights the challenges to undertake such work requiring diverse input from several stakeholders to align on a guideline recommendation for any topic and to then further develop this into a shareable format such as an EOSC Interoperability Guideline. This further impacted the ability to set up new WGs for any underserved areas (**section 9.1**) as support was still required to ensure maturation of current WGs until the point of production for a WG output.

8.1.3 Funding EOSC Interoperability Guideline development efforts & Need for ESFRI RI Partner Beneficiaries

Attempts were made to fund expert participants of the Metadata WG with available project underspend who were not internal to the EOSC Future project. However, due rules surrounding cost claim eligibility, it was not possible to do so. This limitation affected the willingness and participation of prospective members in this particular WG as a result. Further, certain EOSC Future beneficiaries could allot time to work on guidelines if on the project side while those external to the project could not do so. However, even though this was possible to do, most EOSC Future project participants were preoccupied with their own demanding WP commitments to

participate in additional work. Only those external to the project with a compelling interest and time to develop EOSC Interoperability Guidelines were regular contributors as a result.

ESFRI Research Infrastructures (RIs) were largely not members of the EOSC Future consortium [66]. This led to a disjoint in the expertise available for guideline development with researcher side input lacking and further those who did join external to the project with in-kind effort had a lack of needed EOSC Future knowledge and purpose of the EOSC Interoperability Guideline initiative. Key people from ESFRI RIs potentially capable of participating in guideline development were oftentimes already working at full capacity and therefore could not undertake regular additional unfunded and in-kind work.

8.1.4 Competing funded EOSC projects

Several EOSC Projects which began subsequent to EOSC Future project affected participation of various prospective and previously interested members of the EOSC community to engage in the in-kind WGs. Where members were involved in one or more projects, they also lacked time to contribute in the WGs due to funded EOSC project constraints and commitments. Further, certain funded EOSC project deliverables were already working towards similar outputs to proposed areas for guidelines development at a more focused and engaged whole project level as opposed to that of an unfunded and transient WG.

8.1.5 Competing Working Groups/Task Forces/Initiatives at EOSC Level

There are other pre-existing EOSC related WGs and task forces requiring in-kind effort at the EOSC level. This has led to EOSC community fatigue and a lack of enthusiasm for further WG engagements or the set-up of new WGs for underserved subject areas such as the case of PIDs detailed later in **section 9.2.1.**

A comprehensive and well-maintained listing of such in-kind EOSC community groups would be a useful addition to monitor competing interests and align initiatives for EOSC Interoperability Guidelines development. Existing groups should be taken into consideration when underwriting project needs as opposed to recreating the same group.

8.1.6 Re-use of current expert groups e.g.: RDA/RIs/EC Projects/EuroHPC/+...

Section 9.3 further in this document details a diverse range of pre-existing groups from other initiatives and organisations with grounded area focus which could be useful to and to further avoid overlaps with EOSC groups and recreation. Hybrid groups could also be considered for such efforts to leverage multiple membership pools. However, the efforts in coordinating and linking such efforts synthetically will need to be considered with correct incentivisation mechanisms to ensure participation be it at financial or recognition level.

8.2 Recommendations to help boost EOSC-IF guidelines development

8.2.1 Promotional campaign targeted at specific bodies groups

The EOSC IF registry during the course of Task 3.3 work was in its development stages. This was too early to run a promotional campaign of the EOSC IF and call on standards bodies to start proposing EOSC IF horizontal guidelines for the EOSC IF. However, now that the project is nearing its close, example guidelines will begin to populate the EOSC IF registry, templates are available to use and the UI was released in May 2023 it can now be considered to run such a promotional campaign to advertise and encourage the creation of further required guidelines.

8.3 Alleviating administrative burden on WG chairs

WG area chairs flagged the administrative burden when coordinating such WGs with a diverse set of members to coordinate internationally. A sample solution could be sourced from the consideration of using the framework of the Task Forces of the EOSC Association. For their groupings they employ paid coordinators to support their Task Forces in the form of Task Force Support Officers [67] and this could similarly be considered if ever the WG methodology is reused with adaptations.

8.3.1 Flexible financing of experts - open calls/bounties model

Agile open calls and lightweight funding transfer mechanisms to support the development of needed EOSC Interoperability Guidelines should be closely considered in the future when sourcing specific EOSC Interoperability Guidelines. Potential EOSC IF horizontal guideline open calls could be considered in this regard or a similar bounty programme as is regularly employed in the software development space.

8.3.2 Non-financial incentives - Impact, Credit, Recognition and Visibility

Credit, recognition and visibility for EOSC community members is important supplementary consideration in addition to financial compensation for their time to develop EOSC Interoperability Guidelines. Mechanisms for addressing this need should be considered such as EOSC champion awards, young researcher engagement rewards and so on to address various groups to engage who are looking beyond the financial barrier to garner their engagement. Clear outputs and directionality also assist with this when those getting involved understand how they will be credited and what they are contributing to. Prospective citation mechanisms of EOSC Interoperability Guidelines and further ORCID integrations to recognise their contributions should also be considered.

9 Further areas of engagement for sourcing EOSC Interoperability Guidelines

9.1 Underserved EOSC IF horizontal guideline development

Four WGs were created around the requirements output of Deliverable 3.1 and the EOSC Future Technical Roadmap to help facilitate the creation of EOSC IF horizontal guidelines. However, as noted in **section 8** limitations were encountered in employing this WG methodology to support guideline development. Therefore, below each sections details underserved EOSC Interoperability Guideline areas requiring cross-collaboration to facilitate their creation from ESFRI RIs, E-Infrastructures, current/future EOSC Projects and the efforts to date undertaken through Task 3.3 to highlight their need and link relevant expert groups to the EOSC Future body for their creation. Further to this, groups capable of developing.

9.2 Underserved areas & actions taken to develop EOSC Interoperability Guidelines

Section 9.2 details the underserved areas requiring further EOSC Interoperability Guidelines developments and actions taken during the EOSC Future project duration to fulfil these through linking in with expert parties of these areas.

9.2.1 Persistent Identifiers (PIDs)

Persistent identifiers (PIDs) were identified as an important component to facilitate interoperability between data, resources and services across all of the scientific domains. This is due to PIDs and their capability as a long-lasting reference to a digital resource. Unlike URLs, which may break, a persistent identifier reliably points to a digital entity [68] . For this reason, it was determined to focus on the sourcing of an EOSC IF horizontal guideline relating to PIDs. However, several barriers made it difficult to form a EOSC Future WG surrounding this topic area. Firstly, was that there was a pre-established overlapping EOSC Association Task Force on PIDs known as 'PID policy and implementation' [69] with relevant experts from across EOSC already engaged in this area of work. During later stages of the Task 3.3 work, contact was made with a co-chair of the PIDs Task Force to build links to the group and the EOSC Future EOSC IF to explore the creation of a PIDs guideline. The second barrier was the beginning of EOSC funded sister FAIR related projects FAIR-IMPACT [70] and FAIRCORE4EOSC, both of which had EOSC experts involved and dealing with PIDs (e.g., full WP on PIDs). For this reason, Task 3.3 made efforts to align with these sister projects. A member of the FAIRCORE4EOSC Technical Steering Group positively engaged and confirmed the planned delivery of a comprehensive PIDs which would fit in the EOSC Interoperability Guideline registry in early 2025 at Month 34 of the FAIR-IMPACT project.

9.2.2 Artificial Intelligence & Machine Learning (AI/ML)

Artificial Intelligence (AI) and Machine Learning (ML) were areas determined to be of importance to source an EOSC IF horizontal guideline. Given the diversity and expert knowledge required for developing an intuitive user focused EOSC Interoperability Guideline on AI/ML efforts were made to run a workshop to engage such knowledgeable bodies across Europe. However, the limited availability and possible engagement opportunities for these groups made it difficult to find a suitable date to run such a workshop and the required additional guideline support in the aftermath of such a workshop was deemed too large of an overhead to support without a continued structure to support this guideline development.

Therefore, a second direction for this EOSC Interoperability Guideline development was deemed to be through the expert grouping of the AI4EOSC project [72] . A member of AI4EOSC was approached and brought the proposition of the guideline development to the AI4EOSC, project management board. The proposal was approved and a guideline is under current development through the active support of EOSC Future Task 3.3.

9.2.3 Cloud compute and HPC/HTC

Cloud computing, High Performance Computing (HPC) and High Throughput Computing (HTC) have widespread usage across ESFRI RIs to facilitate scientific discovery. They support overcoming requirements for extended computing capabilities not currently available on one's own work machine (e.g., laptop/PC) through accessing larger machines offering expanded capabilities such as peak processing, extended throughput and storage capabilities.

This area was a subtopic of the Compute Continuum WG but given the scope of the WG was not fully developed into an EOSC IF Guideline. These three areas alone are diverse in nature and would require several expert operator and user groups to develop comprehensive and usable EOSC Interoperability Guidelines in each area. Expert computing E-Infrastructures such as GEANT [73] and EGI [74] could potentially be leveraged to develop collaborative guidelines in this space. Further to this, ESFRI RIs from across the EOSC Thematic Clusters with expert user groups in areas of cloud, HPC and HTC services such as the ELIXIR Compute Platform [75] (for the life sciences) could act as the reciprocal user bodies for these guideline developments. Finally, the EuroHPC Joint Undertaking [76] which serves as a complementary parallel of the EOSC initiative but for the European computing space could be leveraged to develop such compute oriented horizontal guidelines.

9.2.4 Containerisation & Orchestration

Digital containers (e.g., Docker/Singularity) and their management through orchestration software (e.g., Kubernetes) were identified to be an area where EOSC IF horizontal guidelines could be developed. This area was noted to fall under the remit of the working charter of the Compute Continuum WG but given the scope of the WG was again not at the capacity to fully develop an EOSC Interoperability Guideline on this topic.

9.2.5 Data Transfer

Additionally, EOSC IF horizontal guidelines for the transfer of data e.g.such as using File Transfer Protocols (FTP) were identified as potentially useful. A diverse body of work on these areas was undertaken through the EOSC Future project in Work Packages 4 and 5. This materialised into a data transfer EOSC Interoperability Guideline [78] which is currently under review after its submission by a member of the EOSC Future project.

9.3 Prospective groups capable of EOSC IF horizontal guideline developments

Once taking into consideration the learnings from the WG methodology employed to source EOSC Interoperability Guidelines in **section 8**, and further the recommendations in **sub-section 8.2 (**Recommendations to help boost EOSC-IF guidelines development), the following groupings could be considered as sources for EOSC Interoperability Guidelines development (sections **9.3.1 -9.3.6)**. However, the correct framework for incentivisation of these guidelines' developments must be carefully considered with a clear remit of the guidelines needed and their related focus/purpose.

9.3.1 EOSC Projects

Current EOSC EC funded projects could be utilised for sourcing EOSC Interoperability Guidelines. Generally, near the start of these projects there is a level of agility possible to plan for guideline development. Further, towards the end of projects where underspend arises there may be further scope to leverage such projects for EOSC Interoperability Guideline development as an effective utilisation of this funding while still effectively contributing to the EOSC initiative. Such cases of leveraging current EOSC projects to help support EOSC Interoperability Guideline development includes **section 9.2.1** and **section 9.2.2** where guidelines development is underway for PIDs by the FAIR-IMPACT project and AI/ML by the AI4EOSC project after direct request, engagement and support from Task 3.3.

Future EOSC projects, where relevant, should underwrite the initiative to form EOSC Interoperability Guidelines related to their area of work. This would allow access to funding for agile internal working groups of experts in a lightweight manner within the projects in a predetermined natural grouping of area experts (e.g., AI/PIDs) to fulfil EOSC Interoperability Guideline delivery underwritten by the project proposers within the project.

9.3.2 EOSC Association ESFRI Working Group & ESFRI RIs

The EOSC Association led ESFRI WG [78] was recently set up in 2023. This WG gives a specific role to EOSC-Association's ESFRI RIs to better define their needs relative to EOSC and their positioning in the EOSC ecosystem. Ultimately the goal with this WG is to facilitate uptake of Open Science and FAIR data practices by the RI's large and largely coherent scientific communities. This top-level venue of coordination between RIs could be leveraged by EOSC and the EOSC Association to petition the population of the EOSC IF registry with specific EOSC guidelines, with a focus on cross-Cluster horizontal EOSC Interoperability Guidelines. However, an incentivised funding mechanism would need to be established to avoid delayed EOSC Interoperability Guideline developments and avoid the previously detailed in **section 8** limitations of the EOSC Future WG methodology.

Further, existing collaborations of ESFRI RIs through the EOSC Thematic Cluster projects or prospective future joint inter-Cluster projects could be leveraged to develop the most comprehensive and widely applicable EOSC Interoperability Guidelines for areas of need.

9.3.3 EOSC Association Task Forces

EOSC Association effectively supports a diverse range of thirteen EOSC Task Forces with well-established groups of experts of various topics. These Task Forces, where relevant, could be leveraged to develop EOSC Interoperability Guidelines. However, an incentivised funding mechanism would need to be considered/established to avoid delayed EOSC Interoperability Guideline developments and avoid the previously detailed in **section 8** limitations of the EOSC Future WG methodology as these groups run in a very similar manner and mainly rely on in-kind contributions of effort.

9.3.4 E-Infrastructures

E-Infrastructures [79] such as GEANT & EGI host a body of experts knowledgeable in technical aspects important to researchers. They are further involved in the EOSC projects and instrumental in the design and operation of the EOSC portal as evidenced in the EOSC Future project. Involvement from e-Infrastructures has also been evident in the production of certain EOSC Interoperability Guidelines such as the AARC [80] EOSC Interoperability Guideline. Potential future technical oriented guidelines could be developed for the EOSC IF in association with user groups for example from the ESFRI RIs.

9.3.5 Research Data Alliance

The Research Data Alliance [81] is a well-established international organisation with over 13,000 members from 148 countries. It hosts a diverse range of well supported RDA Interest Groups and RDA Working Groups which exist to help form standards and guidelines for scientific research. This organisation and its European arm [82] are in a strong and experienced position to answer the call for future EOSC Interoperability Guidelines development of a diverse nature. There is an ongoing positive development of the EOSC initiative in effectively leveraging the RDA in this regard is evident through the recently funded EC EOSC project - RDA Tiger [82] and

efforts of EOSC Future Work Package 3 through the Open Call for RDA / EOSC Future Domain Ambassadors [83] . Efforts are currently underway from Task 3.3 leadership to best leverage these ambassadors for EOSC Interoperability Guideline development and they have been recently engaged on their monthly calls with regards to the EOSC IF and guidelines.

9.3.6 EuroHPC

The European High Performance Computing Joint Undertaking (EuroHPC JU) is a legal and funding entity, created in 2018 and located in Luxembourg to lead the way in European supercomputing. Its main goals include:

- To make Europe a world leader in supercomputing;
- Boost European scientific excellence and industrial strength;
- Focus on European technological sovereignty.

This initiative could serve as an excellent grouping of knowledgeable experts to develop cloud, HPC and HTC guidelines for the EOSC IF. Future efforts to expand the collaboration between EOSC and EuroHPC initiatives could include such an action. EuroHPC's Centres of Excellence [84] which relate back to EOSC Cluster focus areas could be explored as a starting point for such engagement.

10 Conclusions and Summary Recommendations

Task 3.3 of WP3 successfully employed the WG methodology across four diverse domains of relevance to the ESFRI EOSC Cluster project communities. The four WG targeted the areas; Compute, Metadata, Research Product Publishing, and the EOSC Future Science Project Use Cases (WP6). From these WG the basis for EOSC Interoperability Guidelines were developed through several initial WG specific outputs. These outputs are in the process of maturing into EOSC Interoperability Guidelines and evolving into actionable internal project technical recommendations.

Further to the above noted guideline development, engagements of Task 3.3 with other initiatives of domain experts is leading to the development of an AI/ML guideline, a PIDs guideline, a COVID-19 Metadata guideline and more.

In contrast to the successes of the Task 3.3, the limitations of the original WG methodology were identified and lessons learned which can be leveraged to ensure future EOSC projects can build upon the use of such a WG framework and for the EOSC initiative when sourcing EOSC Interoperability Guidelines. Key limitations included budget transfer for WG participation issues, conflicting project commitments, time constraints and being the first to pioneer the development of such horizontal EOSC Interoperability Guidelines.

Further work is required under the cross-Scientific Cluster areas of high-performance computing (HPC) and containers. Existing organisations (e.g., ESFRI RIs/e-Infrastructures), organisational frameworks such as the RDA Interest Groups and Working Groups, EOSC projects, EuroHPC and the EOSC Association Task Forces must be engaged and best leveraged to ensure the EOSC IF registry is successfully populated through co-development of guidelines by experts from the EOSC Clusters as the development of EOSC continues.

References

[1] WP3-T3.1.1 internal Activity 3.1.1 document on Working Groups Methodology (Version 1.0 July 2021). [online] Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Working+Groups?preview=/1805596/1805597/Activity%203.1.1%20Task%20Force%20Methodology.pdf>

[2] EOSC Architecture and Interoperability Framework. [online] Available at: <https://eosc-portal.eu/sites/default/files/EOSC%20Future-WP3-EOSC%20Architecture%20and%20Interoperability%20Framework-2021-12-22%5B17%5D%5B6%5D-2.pdf>

[3] ESFRI EOSC Science Project Custers. [online]. Available at: <https://eosc-portal.eu/events/esfri-science-clusters-long-term-commitments-open-science>

[4] EOSC-Exchange information. [online]. Available at: <https://eoscfuture.eu/ker/eosc-exchange/>

[5] Compute Continuum Working Group public Wiki page. [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Compute+Continuum+Working+Group>

[6] Metadata Working Group public Wiki page. [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Metadata+Working+Group>

[7] Research Product Publishing Framework Working Group public Wiki page. [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Research+Product+Publishing+Framework+Working+Group>

[8] Science Projects Technical Alignment Working Group public wiki page. [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Science+Projects+Technical+Alignment>

[9] EOSC Core information <https://eoscfuture.eu/wp-content/uploads/2022/04/EOSC-Core.pdf>

[10] Covid metadata guideline [online]. Available at: <https://eoscfuture.eu/ker/science-projects/>

[11] ARIA data access guideline [online]. Available at: <https://eoscfuture.eu/data/access-management-for-distributed-research-infrastructures/>

[12] Bioschemas guideline [online]. Available at: <https://bioschemas.org/>

[13] EOSC Future Deliverable 3.1 <https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Future+Project+Deliverables?preview=/18940097/47547210/EOSC%20Future-WP3-EMBL-D3.1%20Science%20Cases%20for%20Developement%20of%20ESOC%20Architecture%20and%20Framework-2021-09-24.pdf>

[14] Compute Continuum WG charter [online]. Available at:<https://wiki.eoscfuture.eu/display/PUBLIC/Compute+Continuum+Working+Group?preview=/16875989/16875990/Compute%20Continuum%20WG-v2-final.pdf>

[15] Metadata WG charter [online]. Available at: <https://docs.google.com/document/d/14B-4tGMle_aCL9ECgrBdfU3VDfLM_Iu_/edit>

[16] Research Product Publishing Framework WG charter [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Research+Product+Publishing+Framework+Working+Group?preview=/9045406/12683498/WP3%20Working%20Group%20-%20Research%20Product%20Publishing%20Framework%20WG.pdf>

[17] Science Projects Technical Alignment WG charter detailed on EOSC Future Public wiki <https://wiki.eoscfuture.eu/display/PUBLIC/Science+Projects+Technical+Alignment>

[18] EIAC and EIAB <https://eosc-portal.eu/eosc-interoperability-framework/eiab-and-eiac-charter>

[19] EOSC Profiles [online]. Available at: <https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles>

[20] EOSC marketplace [online]. Available at: <https://marketplace.eosc-portal.eu/>

[21] EOSC Compute Continuum Resource Profile Specification [online]. Available at: <https://docs.google.com/document/d/1CAAZ07Gz7EJ9svNv_6noZVkprWTpYHu52gDqNAvO2lo/edit#heading=h.wjjv1un4kxbx>

[22] EOSC Profile v4 [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/B.+v4.00+EOSC+Resource+Profile>

[23] EOSC Profiles v3 [online]. Available at: <https://eosc-portal.eu/sites/default/files/EOSC-Profiles-v3.00.pdf>

[24] EOSC Portal [online]. Available at: <https://eosc-portal.eu/>

[25] The FAIR Guiding Principles for scientific data management and stewardship [online]. Available at: <https://doi.org/10.1038/sdata.2016.18>

[26] RDA Metadata Interest Group [online]. Available at: <https://www.rd-alliance.org/groups/metadata-ig.html>

[27] Scientific Knowledge Graphs [online]. Available at: <https://direct.mit.edu/qss/article/2/4/1296/108052/New-trends-in-scientific-knowledge-graphs-and>
[28] EOSC Enhance [online]. Available at: <https://eosc-portal.eu/about/enhance>
[29] EOSC Association Task Forces [online]. Available at: <https://eosc.eu/eosc-task-forces>
[30] EOSC Metadata WG survey of Cluster metadata usage [online]. Available at: <https://drive.google.com/drive/folders/1cAPD8Rc-9GplCTlfJO3x8Xgl54OFUT0J>
[31] EOSC Enhance: D4.3 Analysis of existing research data cataloguing efforts towards integrated discovery [online]. Available at: <https://eosc-portal.eu/sites/default/files/Final-D4.3%20Analysis%20of%20existing%20research%20data%20cataloguing%20efforts%20towards%20integrated%20discovery%281%29.pdf>
[32] Zenodo API [online]. Available at: <https://developers.zenodo.org/>
[33] B2Share API [online]. Available at: <https://eudat.eu/services/userdoc/b2share-http-rest-api>
[34] SWORD API [online]. Available at: <https://guides.dataverse.org/en/latest/api/sword.html>
[35] SWORD Protocol v3. [online]. Available at: <https://sword.cottagelabs.com/swordv3/>
[36] COAR Notify Initiative [online]. Available at: <https://www.coar-repositories.org/notify/>
[37] Signposting API [online]. Available at: <https://api.signpost.com/v2/docs>
[38] Draft Guidelines for accessing content via PIDs [online]. Available at: <https://docs.google.com/document/d/19D5Eqct3qNuqqLFt2DZp7sbvSsQwIyEV/edit>
[39] Draft Guidelines for research product deposition [online]. Available at: <https://docs.google.com/document/d/1ipSQAAQjVtWOIeuQdvlTOGKRoqd4_OMo/edit>
[40] Landscape study undertaken and scenarios document [online]. Available at: <https://docs.google.com/document/d/1dyfyJLDjyZiHVeRq1aIUMc9p9bw85zJ5P7eVmUW68lU/edit?usp=sharing>
[41] SignPosting [online]. Available at: <https://signposting.org/publication_boundary/>
[42] OpenAIRE guidelines [online]. Available at: <http://guidelines.openaire.eu>
[43] WG's summary Recommendation document [online]. Available at: <https://docs.google.com/document/d/1EuVFTijAC5ebG5nitoaYaF7Iih40dH7qqP48c5doEM4/edit?usp=sharing>
[44] FAIRCORE4EOSC Project [online]. Available at: <https://faircore4eosc.eu/>
[45] ESFRI Roadmap [online]. Available at: <https://www.esfri.eu/esfri-roadmap>
[46] ELIXIR Europe [online]. Available at: <https://elixir-europe.org/>
[47] ELIXIR services list [online]. Available at: <https://elixir-europe.org/services>
[48] ELIXIR's legal status FAQ 17[online]. Available at: <https://elixir-europe.org/about-us/governance/faqs#:~:text=ELIXIR%20is%20not%20established%20as,the%20context%20of%20ELIXIR%20activities.>
[49] ELIXIR RDMkit [online]. Available at: <https://rdmkit.elixir-europe.org/>
[50] ENVRI - Climate Change Impact on Biodiversity and Ecosystems in Europe (1) [online]. Available at: <https://docs.google.com/document/d/1HYj6u54V7etfSybFlzDBaQ66N_3mS3eGGnXRLYf-l6M/edit# >
[51] ENVRI - Dashboard on the State of the Environment (2) [online]. Available at: <https://docs.google.com/document/d/1QMM2viUSb6sptv4A_5v1uW03Zf_cJsPd30Jnrt91lbo/edit# >
[52] EOSC-Life - Imaging Data in EOSC - COVID-19 as Demonstrator (4) [online]. Available at: <https://docs.google.com/document/d/1xyJ3zLE-iPRfWuuiFb7zpjlDydKpXXf1AMisls58KuM/edit# >
[53] ESCAPE - Dark Matter SIS (5) [online]. Available at: <https://docs.google.com/document/d/19WnD4m-n75ZsgDFsQZyAWvDcjtnmNW02modDOW09Dyo/edit >
[54] ESCAPE - Extreme Universe and Gravitational Waves (6) [online]. Available at: <https://docs.google.com/document/d/1knoMrJ_ekRSNRdCJshBuD1juinNl3P5bEGe97P4SKjQ/edit# >
[55] PaNOSC - Tracing Biostructures (7) [online]. Available at: <https://docs.google.com/document/d/1LpTYmuzU4kKqC-7kS3xfgR_R752iGPeuDvN_FrImu6M/edit#>
[56] PaNOSC - Serial Crystallography SIS (8) [online]. Available at: <https://docs.google.com/document/d/14BUA8kip4dGXv5-RmZZaHoDNzrEvlsfrhXzFch76nz8/edit >
[57] SSHOC - Access Management for Distributed RIs (10) [online]. Available at: <https://docs.google.com/document/d/1BHUq1xGGrFW1FxlvIbt9ahsAMkm1jQ38SXj_V6ApiKg/edit# >
[58] ESCAPE Dark matter project demo[online]. Available at: <https://drive.google.com/file/d/1FUFkWGdrfJxKdKNH-6d_ltSWlly3wTeH/view>
[59] OpenAIRE and ECRIN defined a crosswalk between the ECRIN model and the OpenAIRE guidelines. [online]. Available at: <https://docs.google.com/spreadsheets/d/1N23xSyzFkbi-cmpEUT1BE4k69gYZUG7NLnUAV1uuv3Q/edit#gid=25715578>

[60] EOSC Life - EuroBioimaging project - Developed Tool and Workflow Development to Support Next Generation File Format OME-Zarr. [online]. Available at: <https://github.com/Euro-BioImaging/BatchConvert>
[61] PaNOSC - VISA platform [online]. Available at: <https://marketplace.eosc-portal.eu/services/virtual-infrastructure-for-scientific-analysis>
[62] EOSC Future Dashboard for Environmental Indicators [online]. Available at: <https://marketplace.eosc-portal.eu/services/eosc.emso_eric.eosc_future_environment_dashboard>
[63] Monkeypox Knowledge Graph: A semantic resource embedding biological and chemical entities [online]. Available at: <https://explore.eosc-portal.eu/search/software?pid=10.5281%2Fzenodo.7351393>
[64] COVID-19 Knowledge Graph: A semantic resource embedding biological and chemical entities [online]. Available at: <https://explore.eosc-portal.eu/search/software?pid=10.5281%2Fzenodo.7351221 >
[65] ENVRI-week and EGU2023 event [online]. Available at: <https://meetingorganizer.copernicus.org/EGU23/session/45572>
[66] EOSC Future Consortium. [online]. Available at: <https://eoscfuture.eu/about/consortium>
[67] EOSC Task Force support officers [online]. Available at: <[https://eosc.eu/secretariat>
[68] PIDs explained by ORCID [online]. Available at: <https://info.orcid.org/documentation/integration-guide/orcid-and-persistent-identifiers/#:~:text=A%20persistent%20identifier%20(PID)%20is,persistent%20identifier%20for%20a%20person>
[69] EOSC Association Task Force on PIDs known as 'PID policy and implementation' [online]. Available at: <https://www.eosc.eu/advisory-groups/pid-policy-implementation>
[70] FAIR-IMPACT project [online]. Available at: <https://fair-impact.eu/>
[71] FAIRCORE4EOSC Technical Steering Group [online]. Available at: <https://fair-impact.eu/technical-bridging-team>
[72] AI4EOSC [online]. Available at: <[https://ai4eosc.eu/>
[73] GEANT cloud services [online]. Available at: <https://clouds.geant.org/>
[74] EGI cloud services [online]. Available at: <https://www.egi.eu/service/cloud-compute/>
[75] ELIXIR Compute Platform [online]. Available at: <https://elixir-europe.org/platforms/compute>
[76] EuroHPC JU [online]. Available at: <https://eurohpc-ju.europa.eu/index_en>
[77] EOSC IF data transfer guideline [online]. Available at: <https://wiki.eoscfuture.eu/display/PUBLIC/Endorsement+Request+-+EOSC+Data+Transfer%3A+Architecture+and+Interoperability+Guidelines>
[78] EOSC Association led ESFRI WG [online]. Available at: <https://www.eosc.eu/news/eosc-general-assembly-convenes-5th-meeting>
[79] E-Infrastructures [online]. Available at: <https://wayback.archive-it.org/12090/*/https://ec.europa.eu/digital-single-market/en/e-infrastructures>
[80] AARC guideline [online]. Available at: <https://wiki.geant.org/display/AARC/AARC+Architecture>
[81] Research Data Alliance [online]. Available at:<https://www.rd-alliance.org/ >
[82] RDA Europe. [online]. Available at: <https://www.rd-alliance.org/rda-europe>
[83] RDA Tiger project [online]. Available at: <https://www.rd-alliance.org/rda-tiger>
[84] EuroHPC COEs [online]. Available at: <https://eurohpc-ju.europa.eu/kick-10-centres-excellence-hpc-support-transition-towards-exascale-2023-01-26_en>