EOSC Future

EOSC Future

WP3
Working Groups Proposal Template

WP3 Working Groups Proposal Template

Authored by Licia Florio (GÉANT), Christos Kanellopoulos (GÉANT), Rosana Wilk (Cyfronet), Jonathan Tedd (EMBL), Paulo Manghi (OpenAIRE), Diego Scardaci (EGI), Michelle Williams (GÉANT), Owen Appleton (EGI)

Approved by EOSC Future Technical Coordination Board

Abstract

Metadata Working Group
A key aspect of EOSC is the catalog of metadata describing assets that may be used to achieve scientific or business objectives by users and their organisations. Provisioning of the EOSC catalog (s) requires asset suppliers to provide their metadata describing the assets in the EOSC-defined canonical format. This format needs to be:
 (a) rich enough to be utilized by end users for discovery, contextualization, integration and re-use – especially in constructing workflows (however simple or complex) to achieve their processing needs;
(b) rich enough to be a superset of all relevant entities/attributes (objects/properties) of the metadata of the suppliers.
The group would:
   (1) Assess the existing EOSC catalog metadata format with respect to end-user processing requirements;
   (2) Assess the existing EOSC catalog metadata format with respect to metadata formats used by asset suppliers;
   (3) Assess the effort in conversion from asset supplier native metadata formats to that of EOSC (a cost-benefit evaluation);
   (4) Recommend changes to the EOSC metadata format arising from the above actions to increase utilization of EOSC and asset provision in EOSC;
The requirements include the need for metadata covering aspects of AAAI, licensing, curation, provenance, citation/acknowledgement in addition to the end-user processing requirements.

The group would then – with appropriate resourcing – execute the activities planned in the above to ensure provisioning of the EOSC profiles (catalogs) from the research clusters.

Dissemination level of the document

☒Public

Version History

Version	Date	Authors	Description
V1	2021-09-13	Keith Jeffery (ENVRI cluster, EPOS-ERIC, EOSC Future WP3)	Proposal
V2		+ Daan Broeder (SSHOC/CLARIN)	
V3		Keith Jeffery	Revision following meeting 20211201 – basically turning it into a 6-month WG and 6-month mini-project

Copyright Notice

Table of Contents

PROPOSAL FOR A NEW EOSC FUTURE PROJECT WORKING GROUP ON [METADATA]

Refer to the instructions on how to complete this form in **Annex A: Additional Information to Complete the Form (DO NOT AMEND)** below.

Introduction

EOSC Future Project Working Groups (TFs) are established under the auspices of the EOSC Future Project in order to create an open forum where experts from its community exchange information, knowledge, ideas and best practices in view of solving specified technical challenges relevant to the implementation of the European Open Science Cloud. TFs will in particular address technical topics which cross WP boundaries in EOSC Future.

General Information

Working Group Name:	**METADATA WORKING GROUP**
Brief Description of Objective:	[Please describe at a high level the remit, how the TF sits in the rest of the landscape, its core aims and objectives, and what it seeks to achieve within the allotted time] A key aspect of FAIRness is the use of semantically explicit, formal syntax metadata. Metadata also sits within the architectural remit of WP3. The Working Group objective is to improve the FAIRness of community asset metadata in general and esp. interoperability and availability of community metadata schemas by providing a proposal for a registration procedure and an inventory of community metadata schema and practices. Next to that, also existing or required community metadata interoperability solutions e.g.,conversion will be noted e.g., conversions to DataCite. As a second 6-month phase the group will – with appropriate resources – execute the plans developed in the first 6 months.
Intended Impact	By providing high quality information on community metadata schemas and practices, we facilitate other EOSC activities wrt metadata interoperability e.g., the The second 6 months should provide stable ingestion pipelines (harvesting) for provisioning EOSC profiles (catalogs) with asset metadata from the research clusters.
Endorsed by	[Please provide a list of the names, organisations and roles of individuals from the EOSC Future project, Research Infrastructures and/or other related projects that endorse the work of the proposed Working Group] ENVRI: the ENVRI community, including the participants in the current cluster-EOSC project ENVRI-FAIR, will participate. SSHOC[1] the thematic cluster project for the SSH supports the metadata workgroup ambitions. SSHOC currently can contribute involvement from Daan Broeder SSHOC/CLARIN, Mari Kleemola SSHOC/CESSDA both involved with SSH interoperability tasks. Note that Daan would, next to representing SSHOC/CLARIN, also be involved from the interest of the SEMAF initiative[2] EOSC-Life/ELIXIR: <Carole/Jonathon to write something> PaNOSC: <Brian to write something>
Proposed Chair(s)	[Please insert name and role/function in the organisation and/or projects of the Chair(s)] Keith G Jeffery : ENVRI cluster, EPOS-ERIC

---

[1] https://sshopencloud.eu
[2] https://doi.org/10.5281/zenodo.4651421

> As background, my first paper on metadata for interoperability was in 1976, I am co-author of the EU recommendation to Member States named CERIF (Common European Research Information Format), I have worked on/with various metadata standards including DC, DCAT, INSPIRE, I have led WPs or Tasks on metadata/catalogs in several projects and am currently co-chair of the RDA Metadata Interest Group and I am handling metadata interoperability aspects in the IEEE Big Data Governance and Metadata Management Group.

Working Group Aims:

1. [This section should provide more detail on the points listed below:

2. Intended achievements,

3. inputs and support required to start the work and to run the Working Group,

4. dependencies on other activities,

5. required additional expertise and capabilities from other projects and organisations in order to achieve the result,

6. constraints that are relevant to the inputs or outputs.]

Intended Achievements

Structured documentation on community metadata schemas and practices that is a useful input for activities in EOSCf e.g., EOSCf activity 3.2.1 and other EOSC projects e.g., the HORIZON-INFRA-2021-EOSC-01-03 call. Associated mappings/conversions to EOSC profiles (catalogs) will also be documented and validated.

The inventory should also be immediate useful for communities that look to improve the interoperability of their metadata with others

This will all improve utilisation of EOSC by end-users and improve provision of assets to EOSC – both by evolving the metadata format of the catalog(s) to achieve the above and the specification of converters.

In the second 6 months the documented mappings and conversions will be engineered into ingestion pipelines (harvesting) so that the EOSC profiles (catalogs) are provisioned from the research clusters.

inputs and support required to start the work and to run the Working Group

A major input will be – from each domain (e.g., each cluster) a specification of their metadata schema and practices and to achieve their objectives (for both end-users and asset suppliers); their current state in metadata provision; detecting potential problems foreseen in mapping/converting to the EOSC metadata profiles; any proposals for improving the EOSC catalog metadata standard. Much relevant information is already available as documents from the cluster project activities and the communities themselves.

A second input will be from the RDA Metadata Interest Group and the associated Metadata Standards Catalog Working Group which provide a landscape analysis of the current international metadata scene.

dependencies on other activities

There is a strong dependency on current projects within the clusters, and other domains, which aim to improve FAIRness including the provision of rich metadata according to the FAIR principles.

required additional expertise and capabilities from other projects and organisations in order to achieve the result

As indicated in the inputs section, additional expertise will be required from clusters and other domains to understand their metadata requirements and capabilities. In the second 6 months, and subject to appropriate resourcing, the documented schemas and mappings/conversions from the first 6 months will be engineered into ingestion pipelines (harvesting).

constraints that are relevant to the inputs or outputs

a 6-month duration will certainly allow progress but the proposed task for the working group will need several 6-month iterations and next iterations can/should liaise with future projects such as HORIZON-INFRA-2021-EOSC-01-03and HORIZON-INFRA-2021-EOSC-01-05

The first 6 months is mapping the landscape and particularly the mappings between metadata schemas and associated conversion services. The second 6 months is engineering those specifications as ingestion pipelines.

Expected Outputs/Results of the Working Group

Please list the expected outputs that the Working Group should achieve.

First 6 months:

1. An inventory and analysis of community metadata schemas, practices and metadata interoperability solutions (conversion/mapping) requirements of users and suppliers concerning metadata by domain;

2. A proposal for an optimal community metadata schema registration procedure and framework;

3. Information for other EOSCf activities such as 3.2.1 working on the evolution of the EOSC catalog metadata standard;

Second 6 months:

1. Engineered ingestion pipelines for each cluster to provision the (amended) EOSC profiles

Duration:

Initial plan for 6 months, with mid-point review in month 3.

#	Planned Activities	Outputs	Due Month (indicative)
1	Generation of questionnaires for the landscape surveys and an inventory of relevant existing documents	Distributed questionnaires and relevant document listings	2
2	Interacting with persons completing the questionnaires leading to receipt of questionnaire results	Conversations with persons in the various domains, discuss optimal registration procedure and framework	4
3	Analysis of the results	Analysis of available information, possible some recommendations to other EOSC activities (3.2.1)	5
3	Provision of recommendations	structured documentation listing the metadata schema and additional interoperability solutions and requirements	6
4	Prototype ingestion pipelines from 2 clusters	Services configured to ingest metadata from catalogs of clusters	9

		to the amended EOSC profiles (catalog)	
	Operational ingestion pipelines from all clusters	Services at appropriate TRL	12

Background Information

[Please insert any additional supporting information as to why the Working Group is proposed; any pre-existing work; a problem statement should be included in order to ensure that all stakeholders understand precisely the problem that the Working Group wants to address]

[TEMPLATE ENDS]

Annex A: Additional Information to Complete the Form (DO NOT AMEND)

Please refer to the Working Groups Methodology for more information.

Initiation

Working Groups can commence as soon as the Technical Coordination Board has approved them.

The initiation of the Working Group will be announced publicly by WP3, and the charter, minutes and results of the Working Groups will be made publicly available in the EOSC Portal.

Once the Working Group has been created, EOSC Future WP3 Task 3.3 will provide continuous support.

Operations, composition, and roles

If it is the opinion of the Working Group proposer that additional expertise and capabilities are required to optimise results, the proposer should expressly set this out in the 'Aims' section along with how it proposes to source the required support. Subsequent to the approval of the Working Group, the members may then source the appropriate expertise.

A Working Group is led by its Chair(s) who shall be appointed for the term of the relevant Working Group's cycle by way of the approval of the Working Group by the Technical Coordination Board.

The Chair is responsible for the business of the Working Group. The Working Group takes decisions by consensus.

The Chair is also responsible for maintaining information on the TF in the project wiki and the monthly briefing during the WP3 calls.

Resources and support

As long as the Working Group is supported by the Technical Coordination Board, the Project Co-Ordinator will provide the following services:

- mailing list(s)
- wiki space.
- online survey tool - on request.
- Support for outreach and promotion (in collaboration with WP10)

Official minutes keeping, meeting recording or other archiving is the responsibility of the Working Group.

Working Group Chairs, if belonging to organisations in the EOSC Future project, are eligible for additional effort accounted in WP3. For other participants, Working Groups are seen as a vehicle to achieve the already planned project outcomes rather than an additional activity.