

EOSC-IF

Interoperability Guideline: Research Product Deposition

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



EOSC-IF / Interoperability Guideline: Research Product Deposition

Lead by **CNR**

Authored by the EOSC Future Working Group on Research Product Publishing: Alessia Bardi (CNR), Paolo Manghi (OpenAIRE), Jose Benito Gonzalez Lopez (CERN), Chris Ariyo (EUDAT), Andreas Czerniak (Bielefeld University Library), Paul Gondim van Dongen (SURF), Georgios Kakaletis (NKUA), Raul Palma (PSNC), Silvio Peroni (University of Bologna), Hans van Piggelen (SURF), Mark van de Sanden (EUDAT), Diego Scardaci (EGI), Jochen Schirrwagen (Bielefeld University Library), Debora Testi (CINECA), Raphaël Tournoy (CNRS), Irena Vipavc (Social Science Data Archive, University of Ljubljana), Deborah Grbac (Library of Università Cattolica del Sacro Cuore di Milano), Carl-Fredrik Enell (EISCAT Scientific Association), Guido Aben (CS₃MESH₄EOSC), Ivan Heibi (University of Bologna), Jorik van Kemenade (SURF)

Reviewed by Michelle Williams (GÉANT)

Dissemination Level of the Document

Public

Abstract

Open Science calls for researchers to publish as soon as possible any type of research product in such a way their research activity can be transparently assessed, reviewed, reproduced, and rewarded in all its aspects. However, the publishing process has become more and more a burden for scientists, who must, most of the time, spend time to publish their articles, data, software, and other products in the many institutional or thematic repositories of reference. Scenarios include first-time publishing of new resource products or double-publishing of research products, to satisfy institutional mandates and community practices. Such tedious work is often incomplete, with some products ending up unpublished and others showing incomplete or imprecise metadata.

Some communities investigated and realised the integration of their research performing services, from research infrastructures and clusters, with repositories for research product deposition. The integration ensures that outcomes of such services are deposited automatically, prior authorization of the users, into a given repository, giving life to an end-to-end scientific workflow, from experimentation to publishing.

The limit of existing approaches is to be bound to a specific repository API and format; introducing multiple repositories as potential targets of deposition for the service, 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 different in many ways; a service willing to automate publishing into either repositories would require implementing and maintaining two different workflows.

For the EOSC to act as enabler for Open Science practices, its Interoperability Framework should guide services of research infrastructures and clusters of the EOSC on how to implement (semi-)automated workflows for the deposition and consumption of research products. To support different integration options, two modalities are supported by these guidelines: SWORD protocol v3 for push mode and a combination of COAR Notify and Signposting for pull mode. The EOSC guidelines for research product onboarding are suggested as metadata exchange format.

Version History

Version	Date	Authors/Contributors	Description
V0.1	08/05/2023	Alessia Bardi (CNR)	Initiation
V1.0	28/06/2023	Alessia Bardi (CNR)	Version for submission
V2.0	07/09/2023	Alessia Bardi (CNR)	Addressed comments from the reviewers: some formatting and improvement to the text in section 2 and 6

Copyright Notice



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

Table of Contents

1.	Glossary	2
2.	List of Abbreviations	2
1	Intended Audience	3
2	Description and Response to Community Need	3
3	Licensing Information	4
4	Related Standards	4
5	Integration Options	5
6	Interoperability Guidelines	5
7	Examples of solutions implementing this specification	6

Glossary

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

List of Abbreviations

Acronym	Definition
API	Application Programming Interface
COAR	Confederation of Open Access Repositories
EOSC	European Open Science Cloud
EOSC IF	European Open Science Cloud - Interoperability Framework
HAL	Hyper Article en Ligne
PID	Persistent identifier
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WG	Working Group

1 Intended Audience

These guidelines target the technical experts responsible for services used by researchers for generating or publishing research products. Examples include, but are not limited to: e-infrastructure services, thematic services, data repositories, institutional repositories, data archives, overlay journals, publishing platforms.

2 Description and Response to Community Need

Open Science calls for researchers to publish as soon as possible any type of research product in such a way their research activity can be transparently assessed, reviewed, reproduced, and rewarded in all its aspects.

The publishing process has become more and more a burden for scientists, who must, most of the time, spend time to publish their articles, data, software, and other products in the many institutional or thematic repositories of reference. Such tedious work is often incomplete, with some products ending up unpublished and others showing incomplete or imprecise metadata. As a solution to these problems, some communities investigated and realised the integration of their research performing services, from RIs and Clusters, with repositories. The integration ensures that outcomes of such services are deposited by the services, when authorised by the users, into a given repository, giving life to an end-to-end scientific workflow, from experimentation to publishing. The limit of existing approaches is to be bound to a specific repository API and format. Solving this limitation by introducing multiple repositories as potential targets of deposition for the service 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 different in many ways; a service willing to automate publishing into either repositories would require implementing and maintaining two different workflows.

With these guidelines, research infrastructures will be able to easily integrate any compliant deposition platform without the need of developing and maintaining any platform-specific API. Fig. 1 illustrates four scenarios that would benefit from the adoption of the guidelines for research product deposition.

1. Scenario 1 - Thematic service and deposition options

A researcher uses a thematic service and produces a research product (e.g. a dataset) that is stored in a storage resource of the underlying e-infrastructure. The thematic service can propose to the researcher several deposition options, targeting repositories that are known to be compatible with the EOSC Interoperability Framework for research product publishing.

2. Scenario 2 - Automatic deposition into repositories of reference

A researcher deposits a research product (e.g. a publication) on a thematic repository. By knowing the affiliation of the researcher, the thematic repository can automatically deposit the publication also in the institutional repository of reference. In a similar way, if the researcher deposits the publication in the institutional repository and provides information about the co-authors, the repository can automatically deposit the publication in the institutional repositories of the co-authors.

3. Scenario 3 - Automatic Open Access deposition from a service of an e-infrastructure

A researcher uses a thematic service and produces a research product (e.g. a dataset) that is stored in a storage resource of the underlying e-infrastructure. The researcher wants to deposit the product on an Open Access repository, but has to wait for an embargo period. The storage resource, based on a configuration provided by the researcher, can automatically deposit the product on an Open Access repository when the embargo period ends (so the researcher does not need to remember that the asset prepared months before can now be made available in Open Access).

4. Scenario 4 - Seamless submission process via publishing platform and overlay journals

Overlay publishing platforms and overlay journals typically suggest their users to deposit their research assets in repositories and get back with the identifier/PID/URI of the deposition so as to proceed with the publication workflow. With the EOSC IF for research product publishing, the platforms and overlay journals could offer their users an integrated experience.

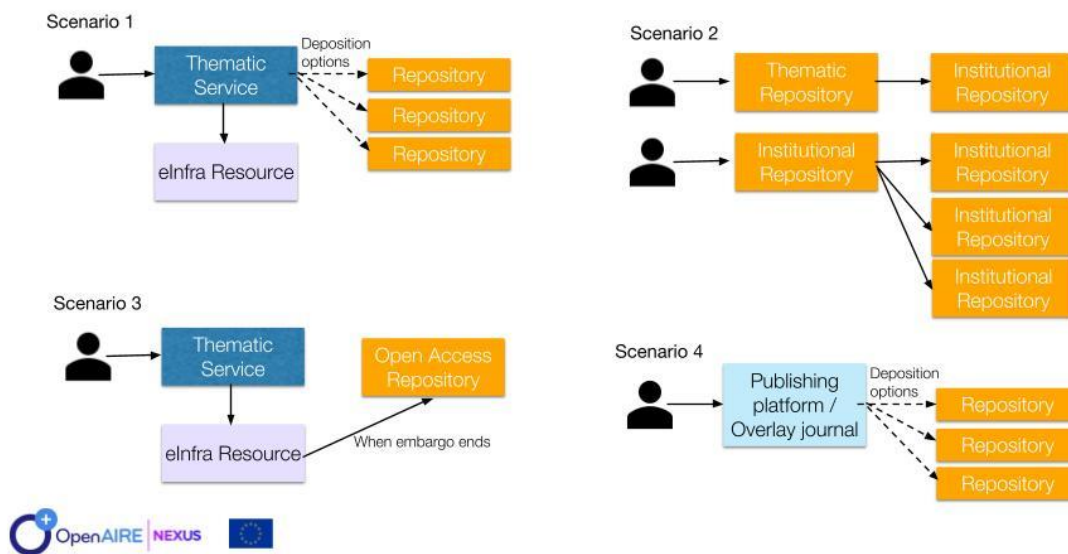


Figure 1 Scenarios that would benefit from the EOSC IF guidelines for research product deposition

3 Licensing Information

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

Signposting the Scholarly Web by <http://signposting.org> is licensed under a Creative Commons Attribution 4.0 International License.

4 Related Standards

Table 1: Related Standards

Title	Short Description	relatedIdentifier
Name or Title of the Related Standard	To describe the context in which the Interoperability Guideline uses or depends on these standards, protocols, APIs and guidelines, rather than providing a short description of the standards, etc, themselves.	Identifiers of related resources. These must be globally unique identifiers. ref <i>EIG.BAI.22</i> e.g. https://argo.eu.github.io/argo-monitoring/docs/Monitoring/guidelines/
SWORD v3	To implement (semi-)automatic deposition workflows in push mode	https://sword.cottagelabs.com/swordv3/
COAR Notify	To implement (semi-)automatic deposition workflows in pull mode	https://www.coar-repositories.org/notify/
Signposting		https://signposting.org/
OpenAIRE Guidelines v4.0	For the metadata exchange format	https://guidelines.openaire.eu/

5 Integration Options

These guidelines suggest two options:

- Push mode with SWORDv3
- Pull mode with COAR Notify + SignPosting

Both options can be used to send/receive payloads and metadata of research products or only metadata.

6 Interoperability Guidelines

Push mode: SWORD v3

[SWORD v3](#) comes with detailed documentation. In particular, the [Protocol Behaviours document](#) describes the requirements of clients and servers for a specific request.

A complete deposition succeeds when both metadata and files are successfully accepted and stored by the server.

According to the SWORD specification, servers must support Metadata Document in the DCMI format. This recommendation suggests the adoption of the format of the EOSC guidelines for research product onboarding, in particular the format of the [OpenAIRE guidelines version 4](#), which are based on the Datacite format.

The support must be declared by servers and used in clients' request as explained in [Section 19 Metadata Deposit](#), the identifier of the format (in form of URI) has to be decided.

Domain-specific communities may agree to accept additional metadata formats to support domain-specific use cases that cannot be addressed without domain-specific metadata formats.

Requests that are relevant in our context are:

- [Creating new objects](#)
- [Appending to objects](#) (to add metadata document and/or files)

Pull mode: COAR Notify + Signposting

[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 metadata and payload file(s) for the deposition.

1. Notify the target that a research product is available for ingest: [Request Ingest Pattern of COAR Notify](#). The id of the object must be a URL based on which it is possible to know where to find the metadata and the payload file(s)
2. The target fetches the payload file(s) via the object url element of the request
3. The target extracts the id and uses Signposting to know where to fetch the metadata record with the [Bibliographic Metadata Pattern](#). To specify the metadata format, a dedicated mime type (as done by Crossref and others) or a dedicated profile must be defined as described in the [Signposting conventions page](#).

Recommended metadata exchange formats

1. Latest versions of the [OpenAIRE guidelines](#) (EOSC guidelines for onboarding of research products)
2. Other community specific format may be adopted in addition to them

7 Examples of solutions implementing this specification

- HAL and Episciences 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 by Summer 2023.
- OpenAIRE is including the implementation of the framework in its roadmap to improve the metadata integration with Zenodo and other repositories.
- Episciences and Software Heritage are working on the implementation of COAR Notify in the context of the FAIRCORE4EOSC EC project (first version by Dec 2023).
- SWORDv3 is implemented by SoftwareHeritage and was used for the implementation of deposition workflow with HAL, Zenodo, and episciences.org

References:

- <https://www.inria.fr/en/hal-opening-up-to-software>
- <https://www.softwareheritage.org/2020/10/27/connecting-scholarly-repositories-with-the-software-heritage-archive/>
- <https://github.com/inveniosoftware/invenio-swh>

More information about the work of the EOSC Future Working Group on Research product Publishing can be found in the [wiki page](#) of the WG.