Research Product Publishing Framework WG

Research Product Publishing Framework WG

Authored by Paolo Manghi (OpenAIRE), Alessia Bardi (CNR-ISTI)
Reviewed and approved by all members

Abstract

Abstract (max.2000 characters)
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 goal of the proposed working group (WG) is to define a Research Publishing framework to simplify the adoption of that practice, by enabling the services of research infrastructures to seamlessly integrate repository deposition workflows in the context of the EOSC. Such an interoperability framework will consist in API and formats that will allow for the implementation of end-to-end research workflows with an on-demand (semi-)automatic step of publishing that ensures the FAIRness of research outputs obtained thanks to the RI. After the first six months of activity, the WG plans to open the protocol specification for public consultation and suggest its inclusion in the EOSC Interoperability Framework.

The WG will contribute to the areas of EOSC Exchange for the definition of the EOSC Interoperability Framework on data publishing and open data, User Experience - Resource Sharing and Discovery, and User Experience - Resource Composability. The activities align with the mission of EOSC Future and with the foreseen collaboration activities with INFRAEOSC-07 projects and Science Projects.

The WG is endorsed by key members of INFRAEOSC initiatives (DICE, EGI, EUDAT, EOSC-Future, NEANIAS, OpenAIRE-Nexus, RELIANCE) and scholarly communication services (episciences.org, EUDAT B2SHARE, OpenCitations, Zenodo). Additional research communities will be involved to elicit requirements with questionnaires or interviews, to ensure a wider audience of interested parties is involved in the process prior to the first open consultation of the framework. The existing networks of WG members will be exploited for this purpose (e.g. network of OpenAIRE-compliant repositories, communities served by Episciences, B2SHARE and Zenodo).

Dissemination level of the document

☒Public

Version History

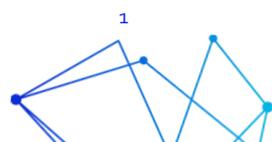| Version	Date	Authors	Description
V1	30/09/2021	Paolo Manghi (OpenAIRE), Alessia Bardi (CNR-ISTI)	Final version for proposal submission
V2	29/10/2021	Alessia Bardi (CNR-ISTI)	Addressed recommendations from TCB
V3	12/11/2021	All	Reviewed version

Copyright Notice

Table of Contents

EOSC FUTURE PROJECT WORKING GROUP ON Research Product Publishing Framework

1. Introduction

1.1. General Information

Working Group Name:	Research Product Publishing Framework
Brief Description of Objective:	Research Infrastructures (RIs) support their communities with tools and services, such as workflow engines, computational tools, and services for data analysis, to perform research activities. These are often not connected to scholarly communication services for publishing (e.g. thematic repositories). As a result, the publishing process is performed manually by the researchers, which is a significant hurdle in the research process and often delays or prevents the publication of results beyond the traditional articles or monographs (narrative descriptions of the research activity and its results), such as datasets, software, methods. Researchers are supported in the publishing process by data management teams and/or the libraries, most of which follow proprietary procedures. This WG aims to establish an EOSC repository deposition framework (API and formats) for bridging the gap between the places where research is conducted (RIs/clusters) and where research is published (thematic, institutional, and/or general-purpose repositories). By implementing the framework, supported by EOSC AAI delegation, RI/cluster services will be able to publish research products (data, software, literature, workflows, etc.) into repositories on behalf of researchers. Such workflows support the adherence to FAIR principles and quality and completeness of metadata (as often services already keep track of detailed metadata about provenance) that is usually not easy to achieve with a manual process, as the deposition activity is commonly intended by researchers as tedious and time-consuming work. The WG will identify real-case scenarios and integration patterns among the services of the members of the WG. Thanks to a landscape study additional partners and services (e.g. thematic repositories, clusters, RIs) can be identified and engaged with questionnaires or interviews for gathering additional requirements to ensure a wider audience of interested parties is involved. A first version of the protocol will be open for consultation and suggested for inclusion in the EOSC Interoperability Framework according to the processes defined by EOSC Future.
Intended Impact	The WG contributes to the technical roadmap of EOSC Future in the following areas: ● EOSC Exchange - EOSC Interoperability Framework on Data Publishing and Open Data for research product publishing (data, software, publications, workflows, etc): delivering API and formats to connect RI/cluster services to institutional/thematic/general-purpose repositories for the FAIR deposition of any type of research products in a timely manner ● User Experience - Resource Sharing and Discovery: the interoperability framework will automate the deposition process without revolutionizing researchers' practices, who will simply use the RI/cluster services as they are used to. Discovery will also improve sensibly, as deposited products will be accompanied by rich and consistent metadata, with links to other related products or services. Such graph-like metadata

	will be exploited by the target repositories and will contribute to the EOSC Resource Catalogue, thereby improving the overall visibility of products and underlying services ● User Experience - Resource Composability: repositories implementing the interoperability framework will be easily composable in end-to-end workflows of communities. Based on the above, we expect the WG to contribute also to the technical activities envisaged in the collaboration of EOSC Future with the INFRAEOSC-07 projects and the Science Projects.
Endorsed by	Chris Ariyo, EUDAT B2SHARE service owner Andreas Czerniak, Bielefeld University Library/OpenAIRE-Nexus Georgios Kakaletris, NEANIAS, Project Technical Manager Raul Palma, PSNC, RELIANCE coordinator Silvio Peroni, University of Bologna, Director of OpenCitations Mark van de Sanden, Technical Coordinator EUDAT, EOSC Future TCB member Diego Scardaci, EGI, Technical Solution Team Lead, EOSC Future TCB member Jochen Schirrwagen, project coordinator at Bielefeld University Library, OpenAIRE-Nexus Debora Testi, CINECA, DICE project coordinator Raphaël Tournoy, CNRS, Episciences Project Manager
Proposed Chair(s)	Alessia Bardi, CNR, OpenAIRE Nexus WP7 leader Jose Benito Gonzalez Lopez, CERN, Zenodo.org technical manager Paolo Manghi, OpenAIRE, OpenAIRE Nexus coordinator, EOSC Future TCB member

1.2. Working Group Aims:

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.

As a solution to these problems, some communities investigated and realized 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, 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 such approaches is however 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 repository would require implementing and maintaining two different workflows.

The goal of the proposed working group is to establish a Research Product Publishing interoperability framework for the EOSC. The framework will enable the services of RIs and clusters to seamlessly integrate repository deposition mechanisms that target any EOSC repository complying with the protocol specified by the framework. The protocol will expose actions to publish any kind of EOSC research products, together with semantic links between them (e.g. cites, supplementTo, versionOf, documentedBy), in order to maximize FAIRness of research outcomes.

Services will be able to discover EOSC repositories compliant with the framework on demand, possibly further filter by discipline or EOSC research product type, and prompt the user with a choice of deposition. On the side of repositories, by implementing the framework protocol (APIs and model) and registering to the EOSC, they

offer themselves as the target of any EOSC research performing services willing to automate the publishing workflow. End-to-end experimental workflows and double-deposition workflows are common examples of such added value automation.

The benefits are multiple:

- Unburdening the work of scientists, which can focus on their experiments and skip the hurdle of the manual publishing process;
- Ensure publishing of science in a structured, complete, FAIR, reproducible, and community-driven manner;
- Ensuring high-quality metadata and fully-fledged monitoring/accounting of science, by systematically interlinking research products between them and with the services they are related with;
- Increasing repository visibility and adoption.

Inputs to the WG will be previous experiences and prototypes of integration between RI services and repositories, previous work performed by the RDA Research Data Repository WG (see Background information), standards and initiatives on this topic (EOSC TF on Semantic Interoperability, PEER project, SWORD protocol, COAR Next Generation Repositories WG), and the technical documentation of existing deposition/publishing API of repository services operated by organizations endorsing this WG, such as Zenodo.org (CERN/OpenAIRE), B2SHARE (EUDAT), EGI Application DB (EGI).

The WG will align to other technical activities of EOSC Future, in particular to the collaborations among INFRAEOSC-07 projects, Science Projects and the activities related to EOSC Resource Composability and the EOSC Task Force on Semantic Interoperability.

To facilitate the alignment and a successful definition and adoption of the framework, the WG includes members of INFRAEOSC-07 projects and European Research Infrastructures, and initiatives mentioned above. Additional research communities will be involved to elicit requirements with questionnaires or interviews, to ensure a wider audience of interested parties is involved in the process prior to the first open consultation of the framework. The existing networks of WG members will be exploited for this purpose (e.g. network of OpenAIRE-compliant repositories, communities served by Episciences, B2SHARE and Zenodo).

1.3.   Expected Outputs/Results of the Working Group

The WG will identify real-case scenarios and integration patterns among the services of the members of the WG. Thanks to the collaboration with the members of the EOSC TF on Semantic Interoperability and a landscape study, additional partners and services (e.g. thematic reposities, clusters, RIs) can be identified and engaged with questionnaires or interviews for gathering additional requirements to ensure a wider audience of interested parties is involved. The existing networks of WG members will be exploited for the same purpose (e.g. network of OpenAIRE-compliant repositories, communities served by Episciences, B2SHARE and Zenodo)

After six months, a first protocol recommendation will be open for consultation and suggested for inclusion in the EOSC Interoperability Framework according to the processes defined by EOSC Future.

As summarised in the table below, the WG plans to deliver:

1. A landscape study on (semi-)automatic publishing workflows/integration between RI and repository services, based on previous experiences and prototypes involving the members of this WG, available in the literature, or obtained as input from other involved parties [M3]
2. A collection of use case scenarios and integration patterns: it will include use cases among the services of the WG members (some are already collaborating in this sense in the context of the INFRAEOSC-07 projects collaboration, e.g. Zenodo.org, B2SHARE, episciences.org) [M3]
3. An analysis of existing repository deposition frameworks from the functional and non-functional perspectives (identification of common patterns, common problems, etc) [M5]

4. A protocol recommendation (API and formats) for EOSC repository deposition, open for consultation and suggested for inclusion in the EOSC Interoperability Framework based on the policies and processes defined by EOSC Future [M6]

1.3.1. Duration

The initial plan for 6 months, with a mid-point review in month 3.

#	Planned Activities	Outputs	Due Month (indicative)
1	Landscape study on (semi-)automatic publishing workflows/integration between RI and repository services	Report (Input for activities 2 and 3)	M3
2	Identification of real-case scenarios and integration patterns among the services involved in the WG	Report (Input for activity 4)	M3
3	Analysis of existing repository deposition frameworks from the functional and non-functional perspectives (identification of common patterns, common problems, etc)	Report (Input for activity 4)	M5
4	Protocol recommendation v1.0 open for consultation and refinement	Recommendation	M6

Background Information

Agreeing on a protocol for research product publishing to be adopted in the EOSC is the first step to address some of the barriers that hinder the adoption of Open Science practices or the achievement of its objectives[1][2]:

● Manual publishing cost: researchers often consider the time they need to fill in a deposition form as time "stolen" to more important research activities. As a consequence, they tend to provide the mandatory information and ignore the rest, leading to poor metadata and a low level of FAIRness of the deposited product. With a (semi-)automatic deposition step integrated in the RI, most metadata can be filled in automatically (e.g. links between services, software, data; provenance metadata) so that researchers can focus on the missing parts (which may vary depending on the specific RI service at hand).
● Article-centric publishing and reward practices: many researchers overlook Open Science practices because they take time and bring no direct impact to their careers. It will take some time for omni-comprehensive reward policies to be widely adopted: by simplifying the publishing process of any type of research output and integrating it into their usual research practice, the EOSC will remove the stick and leave them with a "future carrot".
● Integration costs: RIs willing to offer an integrated deposition step have to spend time and technical effort at implementing a different API for each repository they want to target. As a result, when they do, they integrate one or few repositories, limiting the choice of researchers.
● Losing context: when published out of the RI, research products often lose their context (links to data, software, used services, configuration, etc.). This context is fundamental to achieve research reproducibility and to ensure a high level of FAIRness of the published research products.

The activities of the WG will be based on previous work about services, data, and metadata interoperability carried out in the context of:

● EOSC TF on Semantic Interoperability
● RDA Research Data Repository WG
● COAR Next Generation Repositories WG
● OpenAIRE Interoperability Guidelines
● OpenAIRE, which experimented integration among its services and community/RI services (e.g. EPOS GEP, VIP for neuroinformatics) [1][3][4][5]

- D4Science integration with Zenodo/B2Share [6]
- [Zenodo integration with Dryad](), [GitHub](), [SWH](), and [HADDOCK]()
- [Episciences integration with Zenodo, HAL, arXiv]()
- Ongoing work on implementing [COAR Notify]() into Episciences and HAL
- [COAR Controlled Vocabularies]()
- OpenAPC integration in the OpenAIRE Research Graph
- [PEER project]()
- [SWORD protocol]()

[1] Baglioni M. et al. (2019) The OpenAIRE Research Community Dashboard: On Blending Scientific Workflows and Scientific Publishing. In: Doucet A., Isaac A., Golub K., Aalberg T., Jatowt A. (eds) Digital Libraries for Open Knowledge. TPDL 2019. Lecture Notes in Computer Science, vol 11799. Springer, Cham. https://dx.doi.org/10.1007/978-3-030-30760-8_5. Open Access version at https://doi.org/10.5281/zenodo.3467104

[2] Brauer PC. et al. (2014) Start Smart and Finish Wise: The Kiel Marine Science Provenance-Aware Data Management Approach. In: 6th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2014), USENIX Association, https://www.usenix.org/system/files/conference/tapp2014/tapp14_paper_brauer.pdf

[3] Manghi, Paolo, Manunta, Michele, Baglioni, Miriam, Bardi, Alessia, Casu, Francesco, De Luca, Claudio, & Kokogiannaki, Argiro. (2018, July 18). Enabling Open Science publishing for Research Infrastructures via OpenAIRE: The EPOS use-case. https://doi.org/10.5281/zenodo.1412509

[4] Bardi, Alessia, Manunta, Michele, Toth-Czifra, Erzsebet, Vergoulis, Thanasis, Manghi, Paolo, & Baglioni, Miriam. (2019). D7.3 – Interoperability with Research Infrastructures. Zenodo. https://doi.org/10.5281/zenodo.3701394

[5] Pop, Sorina, Bonnet, Axel, Maumet, Camille, Kain, Michael, Barillot, Christian, & Glatard, Tristan. (2018, November 30). OpenAIRE Research Community use case: Neuroinformatics. Digital Infrastructures for Research 2018 (DI4R2018), Lisbon, Portugal. Zenodo. https://doi.org/10.5281/zenodo.1744873

[6] Bardi, Alessia, Bruseker, George, Durco, Matej, Illmayer, Klaus, Broeder, Daan, Lorenzini, Matteo, Resch, Stefan, Sugimoto, Go, Theodoridou, Maria, & Assante, Massimiliano. (2019). PARTHENOS D6.4 Report on services and tools (final). Zenodo. https://doi.org/10.5281/zenodo.2607212