

D3.3 - Roadmap to semantic and technical interoperability v2.0



APR. 2022

Scholarly Communication Services for EOSC users

D3.3 –Roadmap to semantic and technical interoperability v2.0

Version 1.0 – Draft
PUBLIC

This deliverable updates the initial roadmap detailed in D3.1 after 14 months of activity.

H2020-INFRAEOSC-2020-2
Grant Agreement 101017452

Document Description

D3.3 – Roadmap to semantic and technical interoperability v2.0

WP3 - Integration with the EOSC to provide Virtual Access

WP participating organizations: **CNR**, UNIBI, CERN, AthenRC, UGOE, ICM, UNIBO, CNRS, CITE

Contractual Delivery Date: 29/4/2022

Actual Delivery Date:

Nature: Report

Version: 1.0 (Draft)

Public

Preparation Slip

	Name	Organization	Date
From	Paolo Manghi	OpenAIRE AMKE	03/2022
Edited by	Paolo Manghi Claudio Atzori Raphael Tournoy Dimitris Pierrakos Silvio Peroni Andreas Czerniak Nicolas Liampotis Georgios Kakalettris	OpenAIRE AMKE CNR CNRS AthenaRC UNIBO UNIBI GRNET CITE	03/2022
Reviewed by	Alessia Bardi Paolo Manghi	CNR OpenAIRE AMKE	04/2022
Approved by	Paolo Manghi	OpenAIRE AMKE	24/04/2022
For delivery	Eleni Koulocheri	OpenAIRE AMKE	28/04/2022

Revision History

Issue	Item	Reason for Change	Author	Organization
V1.0	Draft version	drafting	Editors	

	05/04/2022			
V1.1	Draft version 10/04/2022	Input from partners	Editors	
V1.2	Revised version 19/04/2022	Reviewing	Alessia Bardi	CNR
V1.3	Reviewed version 20/04/2022	Corrections based on revisions	Claudio Atzori	CNR

Contents

1	<i>Introduction</i>	6
2	<i>Semantic integration with the EOSC</i>	8
2.1	EOSC Resources	9
2.1.1	PUBLISH	10
2.1.1.1	Zenodo.org	11
2.1.1.2	Episciences.org	11
2.1.1.3	Argos	11
2.1.1.4	MONITOR	11
2.1.1.5	UsageCounts	11
2.1.1.6	Scholexplorer	11
2.1.2	DISCOVER	11
2.1.2.1	PROVIDE	12
2.1.2.2	EXPLORE and CONNECT	12
2.2	PIDs: ensuring PID are pervasively and correctly used by Nexus services	12
2.2.1	PUBLISH	12
2.2.1.1	Zenodo.org	13
2.2.1.2	Episciences.org	13
2.2.1.3	Argos	14
2.2.2	MONITOR	14
2.2.2.1	UsageCounts	14
2.2.2.2	OpenCitations	15
2.2.2.3	OpenAPC	16
2.2.3	DISCOVER	17
2.2.3.1	EXPLORE and CONNECT	17
3	<i>Technical integration with the EOSC</i>	18
3.1	EOSC onboarding	19
3.2	EOSC AAI	19
3.3	EOSC Accounting and monitoring for services	21
3.4	EOSC Accounting for research products	21
3.5	RDA Frameworks	21
3.5.1	Argos	22
3.5.2	UsageCounts	22

Table of Figures

Figure 1 - OpenAIRE-Nexus service portfolio.....	7
Figure 2 - EOSC Resource Data Model	10

Disclaimer

This document contains description of the OpenAIRE-Nexus project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the OpenAIRE-Nexus consortium and can in no way be taken to reflect the views of the European Union.

OpenAIRE-Nexus is a project funded by the European Union (Grant Agreement No 101017452).



Publishable Summary

PLANNING THE INTEGRATION WITH THE EOSC – M1-M14

The aim of this deliverable is to provide an update on the lines of action that enhanced, during the first 14 months of the project, the OpenAIRE-Nexus services to adhere to the EOSC Interoperability Framework (IF) recommendations, i.e., protocols, APIs, data model, exchange formats.

1 INTRODUCTION

As requested by the Work Programme, OpenAIRE-Nexus services will be “...flexible in order to take into account all the relevant governance and business models, rules for participation, operational requirements, standards, etc. in accordance with topic INFRAEOSC-03-2020.” [cit. Work Programme]. The aim of this deliverable is to draw the lines of action that will, in the first 14 months of the project, enhance the OpenAIRE-Nexus services to adhere to the EOSC Interoperability Framework (IF) recommendations, i.e., protocols, APIs, data model, exchange formats. At the time of writing the EOSC is still “in the making” and such recommendations are only partly established and in most cases under discussion by the community. Recommendations have been provided by several **EOSC working groups**, operating independently or jointly supported by the eInfraCentral, EOSC-Secretariat, EOSC-Hub, EOSC-Enhance, and OpenAIRE-Advance H2020 projects. In particular, the EOSC IF recommendations¹ define two main categories of Interoperability Frameworks (IFs):

- **EOSC Core IFs:** required to integrate EOSC resources into a common ecosystem enabled by the EOSC Core services;
- **EOSC Exchange IFs:** collected and maintained in a registry of research IFs, used to facilitate the uptake and consolidate known IFs across communities, enable the definition of crosswalks/mediators between them, to favour convergence on common standards; moreover, EOSC resources will refer to the IFs they comply with, enable resource discovery and clustering by IF compliance.

The technical roadmap in OpenAIRE-Nexus will focus on the EOSC Core IFs, making a distinction between *semantic integration*, i.e., adapting services and back-ends to recommended data models, concepts, vocabularies, and *technical integration*, i.e., adapting services to recommended APIs and data exchange formats.

The next sections will describe the specific IFs to be implemented in the first 14 months of the project. As mentioned above, only a few **mandatory** IFs are today defined and endorsed by the EOSC Core-related working groups, while others are **expected** and will be identified and endorsed by the EOSC-Future project (GA 101017536). OpenAIRE-Nexus will ensure the mandatory IFs are implemented across all services (shown in Figure 1). As to the expected IFs, the project will select those for which known and consolidated standards exist to ensure the services comply with them, thereby anticipating the choices that will be made by the EOSC Core working groups or facilitating the integration to any other similar standard. Such choices will be made trying to

¹ EOSC IF recommendations: <https://dx.doi.org/10.2777/620649>

minimize waste of resources and ensuring an overall benefit for the services involved.

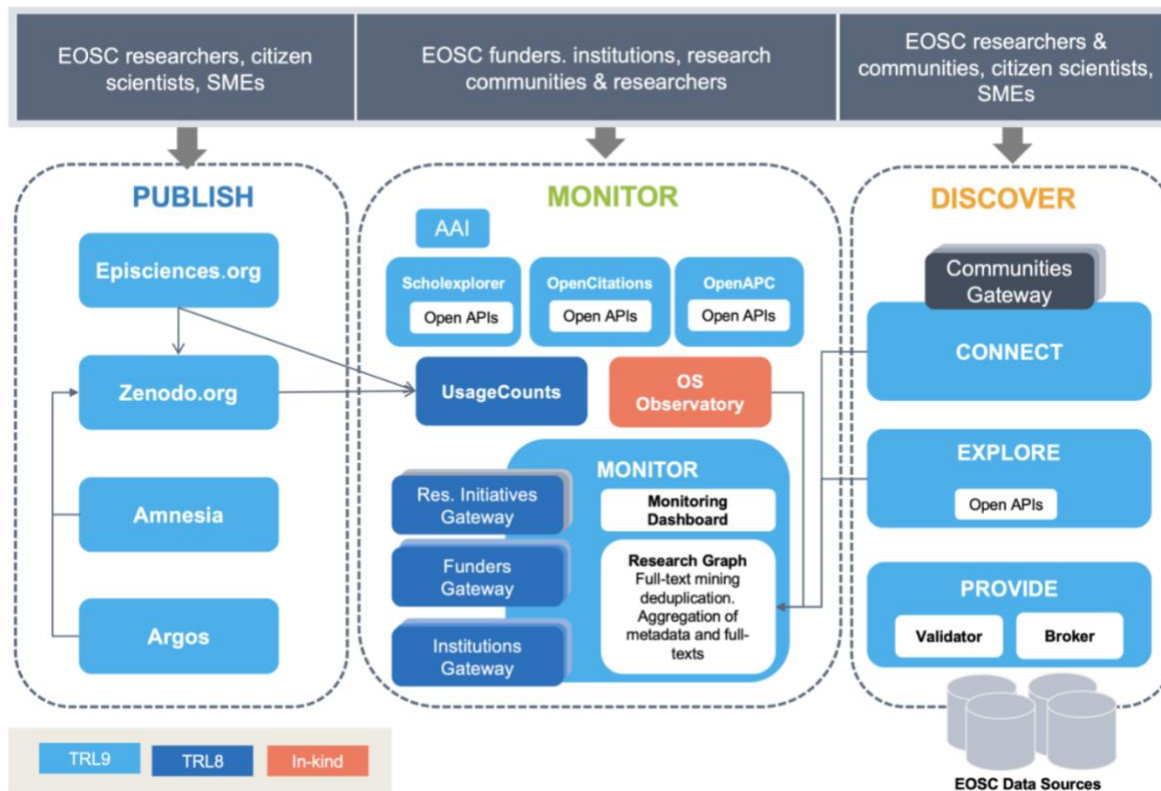


Figure 1 - OpenAIRE-Nexus service portfolio

2 SEMANTIC INTEGRATION WITH THE EOSC

Table 1 provides an updated view of the semantic frameworks adopted by OpenAIRE-Nexus services at M14, specifying how these are addressed by the specific portfolios and services:

- **The EOSC resource description framework**, which establishes the metadata profiles of EOSC resources, intended as *services* and *research products*;
- **Persistent identifiers (PIDs)**, to ensure uniform, validated, and up-to-date references across research entities in the EOSC.

Columns highlighted in orange indicates the activity for supporting the specific PIDs is still ongoing.

Table 1 - OpenAIRE-Nexus Services and EOSC Semantic Integration

Portfolio	Inst. No.	Service	EOSC Resources	PIDs pervasive adoption
PUBLISH	1	Zenodo.org	Already compliant with OpenAIRE guidelines	ORCID ROR
	2	Episciences.org	Is compliant with OpenAIRE guidelines V4	ORCID SWHID links to projects links to other objects (datasets, software) ROR DOI ArXiv ID HAL ID ISSN
	3	Argos	Utilizing reserved vocabularies and APIs from OpenAIRE (to be extended). Partial alignment with EOSC resource concepts (work in progress). EOSC users may be referenced in several areas of Argos Data Model (contributors, authors, managers etc.).	When depositing DMP into Zenodo.org, include links to: Datasets, DMPs (DOIs) ORCID Organizations (OpenAIRE IDs)
	4	Amnesia	n/a (integrated with OpenAIRE guidelines for datasets via Zenodo.org)	n/a (the service is stateless)
MONITOR	5	MONITOR	Including service entities as EOSC resources in the OpenAIRE Research Graph	n/a (the OpenAIRE Graph is open to all kinds of PID schemes)
	6	UsageCounts	Compliance to OpenAIRE guidelines for usage statistics	UsageCounts support any type of PID available in the OpenAIRE research graph

	7	OpenCitations.net	The main kind of object handled is the citation (as a proper first-class data entity) and the related entities involved in it, i.e., the citing entity and the cited entity that may be characterized by different EOSC resource types (text, dataset, software, etc.)	DOI	
				PMID	
				ORCID	
				VIAF	
	8	ScholarXplorer	Including software entities to the existing publications and data	n/a (the service is open to all kinds of PID schemes)	
	9	OpenAPC	Alignment with OpenAIRE Research Graph concepts to link cost data to publications, organizations, funders and projects	publications	URL
					DOI
					PMID
				organizations	PMCID
					ISSN
					ISBN
	10	OS Observatory	n/a (the service relies on the OpenAIRE Research Graph to generate stats)	n/a (the service on the OpenAIRE Research Graph)	
	11	AAI	n/a	n/a	
DISCOVER	12	PROVIDE	EOSC Data Source profiles	n/a (the service is agnostic of PID schemes for data sources)	
	13	EXPLORE UI	EOSC Data Source profiles and Services	Integration of the ORCID Search and Link functionality	
	13	EXPLORE APIs	EOSC Data Source profiles and Services	n/a	
	14	CONNECT	EOSC Data Source profiles and Services	Integration of the ORCID Search and Link functionality	

2.1 EOSC Resources

The EOSC Core working groups agreed on a common EOSC resource data model (see Figure 2), which consists of *services* of various kinds and *research products*, intended as outcomes of science, namely publications, datasets, software. The data model is under definition and endorsement and will be completed and implemented by the EOSC Core services in the context of the EOSC-Future project.

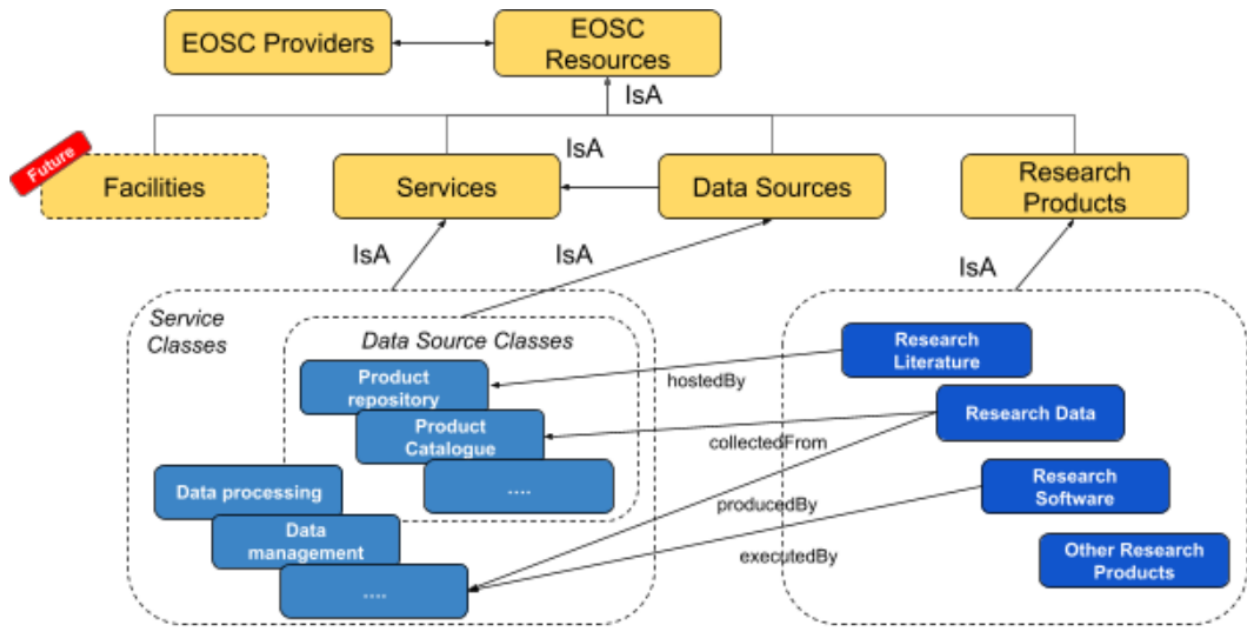


Figure 2 - EOSC Resource Data Model

Services: All OpenAIRE-Nexus services are registered to the OpenAIRE Service catalogue², which is compliant to the EOSC Resource Profiles. Due to the new release 4.0 of the profiles the OpenAIRE catalogues the synchronization with the EOSC resource catalogue is for the moment on hold.

Data sources: The EOSC Resource profile has been extended to include [EOSC data source profiles](#), sub-class of EOSC services. Data sources are intended as services that store and provide access to research products (metadata and/or files). The new release 4.0 of the EOSC resource catalogue (expected for September 2022) will include data sources as specific kinds of services.

Research products: The EOSC Resource Profile has been extended to include EOSC Research Product profiles, namely the OpenAIRE guidelines for scholarly communication data sources³. Such a set of guidelines is being extended with community guidelines, starting with EOSC Life metadata guidelines based on bioschemas.org.

In the following we shall describe the effort that OpenAIRE-Nexus services will perform to integrate with the EOSC resource data model.

2.1.1 PUBLISH

² <http://catalogue.openaire.eu>

³ <http://guidelines.openaire.eu>

2.1.1.1 ZENODO.ORG

The service is already compliant with OpenAIRE Guidelines.

2.1.1.2 EPISCIENCES.ORG

The bibliographic metadata exposed by the Episciences service was made compliant with the OpenAIRE Guidelines for content providers 4.0. This makes sure its journals and publications (articles) are kept in-sync with the OpenAIRE Graph underlying MONITOR, hence be onboarded into the EOSC as research products and data sources.

2.1.1.3 ARGOS

In Argos semantic integration with EOSC includes acknowledgment and utilization of EOSC provided resources. On top of the status reported in D3.1 we can mention that the machine Actionable DMP support (RDA maDMP spec) is being continuously improved and covers new cases, as the data model of Argos and the one projected by maDMP have different foundations.

2.1.1.4 MONITOR

The update of the graph data model to support EOSC services as entities, together with links to other entities, has begun during the 1st half of the project. EOSC services will be added in the OpenAIRE research graph built on beta environment first, so that the different components of the OpenAIRE system can be adapted to support them.

2.1.1.5 USAGECOUNTS

The service adheres to the OpenAIRE guidelines for Usage Statistics in order to allow the collection of usage activity, either as raw usage activity via platforms' plugins or patches, or via COUNTER CoP compliant reports. The service is compliant with the COUNTER CoP⁴ for the major categories of e-resources (articles, books, etc.) and we are working towards the support for COUNTER CoP R5 and the new standard proposed by MakeDataCount, i.e., the COUNTER CoP for Research Data⁵.

2.1.1.6 SCHOLEXPLORER

The model of the service currently describes links (semantic relationships) between publications and datasets but will be extended to include links to EOSC service entities. The activities to fully align the data models of Scholexplorer and the OpenAIRE Research Graph were completed during the 1st half of the project and the next phase will finalize the integration of the EOSC service entity along with the links to the research products that acknowledge them.

2.1.2 DISCOVER

⁴ <https://www.projectcounter.org/release-5-code-practice/>

⁵ <https://www.projectcounter.org/code-of-practice-rd-sections/foreword/>

2.1.2.1 PROVIDE

PROVIDE manages data sources, referred to as *content providers* in OpenAIRE, and will adapt its data model to the EOSC data source profiles as defined by the EOSC. The action cannot be planned at the moment as data source profiles are being defined and refined at the time of writing. The new release 4.0 of the EOSC resource catalogue (expected for September 2022) will include data sources as specific kinds of services. PROVIDE will be then adapted to align with such a release.

2.1.2.2 EXPLORE AND CONNECT

At the time of writing, the integration of the EOSC service entities is being tested on the OpenAIRE beta environment. The plan assumes to integrate them as part of the graph and then to adapt the EXPLORE and CONNECT portals, so that the EOSC services can be used in the Search & link functionality.

2.2 PIDs: ensuring PID are pervasively and correctly used by Nexus services

PIDs are regarded as a key building block of the EOSC. OpenAIRE-Nexus will make sure that its services will pervasively adopt identifiers, by ensuring that, at different *quality degrees*:

- **QD1:** PIDs are part of the metadata
- **QD2:** PID schemas are part of the metadata (when more than one scheme is allowed, e.g., ROR and GRID for organisations)
- **QD3:** PID value in the metadata are validated by format, i.e., the template of the PID is respected
- **QD4:** PID value in the metadata are validated by resolution, i.e., at metadata ingestion time the PIDs are resolved to double-check they are correct

As an outcome, the project will also release an authoritative list of PID schemes, to be shared across the services and proposed as part of the EOSC Interoperability Framework. The OpenAIRE Graph will expose a PID resolution API, capable of returning the full metadata of any pair <PID, PID scheme> part of the Graph.

In the following, for each OpenAIRE-Nexus service, the list of PID schemes currently supported and the related *quality degree* will be provided, together with the intended roadmap for including further PID schemes and/or implementing higher quality degrees at M14 and M30 of the project.

2.2.1 PUBLISH

2.2.1.1 ZENODO.ORG

Zenodo supports a variety of PID types, which are by default structurally validated (QD3), but are not validated on their resolution (QD4).

In terms of also applying resolution validation for personal (ORCID) and organizational (ROR) identifiers, users will be able to select such values from a database regularly updated from the official resolution source of the PID type in order to ensure QD4 compliance.

The table below pictures the roadmap of Zenodo in terms of PID scope, the letter “P” stays for “prototype”, to indicate an intermediate state of implementation.

Table 2 - Zenodo.org: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
ROR					P	P	P	P	X	X	X	X
ORCID	X	X	X		X	X	X	P	X	X	X	X
DOI	X	X	X		X	X	X		X	X	X	
BibCode	X	X	X		X	X	X		X	X	X	
SWHID	X	X	X		X	X	X		X	X	X	
PURL	X	X	X		X	X	X		X	X	X	
PubMedID	X	X	X		X	X	X		X	X	X	
arXiv	X	X	X		X	X	X		X	X	X	
ARK	X	X	X		X	X	X		X	X	X	
ISNI	X	X	X		X	X	X		X	X	X	
ISBN	X	X	X		X	X	X		X	X	X	
HAL	X	X	X		X	X	X		X	X	X	

2.2.1.2 EPISCIENCES.ORG

Episciences supports DOI, HAL and arXiv PIDs to handle the communication between journals and the open repositories supported by the platform. We plan to allow authors to add SWHID to their submissions in addition to links to datasets with DOI. Another priority is to allow authors to add their ORCID, and later their affiliations linked to the ROR. The identifiers will be added to export formats such as OpenAIRE metadata format and Crossref DOI registrations. It will also be used when submitting metadata content to the DOAJ. By design, most of the PIDs will be resolved because we need to fetch the metadata behind them.

Table 3 – episciences.org: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
ORCID					X	X	X		X	X	X	X
ROR									X	X	X	X
DOI	X		X	X	X	X	X	X	X	X	X	X
SWHID					X	X	X	X	X	X	X	X
HAL	X		X	X	X	X	X	X	X	X	X	X
ArXiv ID	X		X	X	X	X	X	X	X	X	X	X
ISSN	X		X		X	X	X		X	X	X	

2.2.1.3 ARGOS

Argos currently natively supports directly two types of PIDs. In particular, Argos may link researchers / contributors to its data sets as well as to the Data Management Plan of those datasets. This linking is performed programmatically, and researchers' identifiers are directly validated against ORCID API, used for the selection and existence check of the ORCID entity (users can opt for skipping the validation). Argos also supports DOIs for the publication of its DMPs. Those DOIs are issued directly by Zenodo as a PID issuer and are always validated.

In addition, Argos allows the users to define custom data types that can be used for introducing DOIs, ORCIDs or PIDs of any other scheme in custom metadata fields of a DMP (via the Argos DMP template editor). Those currently are not checked for validity in any way.

Argos now utilizes Zenodo PIDs for obtaining dataset metadata from the relevant API and prefilling its dataset description forms with data. Furthermore, Argos now implements and engages a URL validator for artifacts declared by URL identifiers in its Data Management Plans.

Table 4 – Argos: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
ORCID	X	P	P	P	X	P	P	P	X	P	P	P
DOI	X	P	P	P	X	P	P	P	X	P	P	P

2.2.2 MONITOR

2.2.2.1 USAGECOUNTS

UsageCounts service supports collection of usage data from institutional repositories using PIDs like DOIs, Handles, URNs, PubMed IDs and arXiv which are mapped to OpenAIRE ids via the deduplication process. Support of other PIDs like ORCIDs, will be part of a future development of the service.

Table 5 – UsageCounts: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
ORCID			X	X			X	X	X		X	X
WorkID												
DOIs	X		X	X	X		X	X	X		X	X
Handles	X		X	X	X		X	X	X		X	X
URN	X		X	X	X		X	X	X		X	X
PubMed ID	X		X	X	X		X	X	X		X	X
arXiv	X		X	X	X		X	X	X		X	X

2.2.2.2 OPENCITATIONS

Several PIDs are stored and handled in the collections made available by OpenCitations. However, the types of PIDs and the way they are ingested may vary according to the particular collection in consideration. The table presented in this section, thus, describe the status of all the PID types stored according to four specific collections:

- OpenCitations Indexes (<http://opencitations.net/index>): it is the main collection, actively developed, which stores more than 759 million citation links between more than 60 million entities. Only metadata about citations (seen as first-class data entity) are stored, while there is no information available in the collection about the citing and cited entities characterising citations except the PIDs identifying them.
- OpenCitations Corpus (<http://opencitations.net/corpus>): it is the first original collection released by OpenCitations, and currently contains around 14 million citation links to over 7.5 million cited resources. The collection includes information about bibliographic resources, resource embodiments, bibliographic references, responsible agents, agents' roles, and identifiers. Currently, it is now updated and it is used as a 'sandbox'.
- Open Biomedical Citations in Context Corpus (<http://opencitations.net/ccs>): this new collection has been built reusing and largely extending the software developed to create the OpenCitations Corpus. In addition to all the kinds of entities already described in the OpenCitations Corpus, the Open Biomedical Citations in Context Corpus also includes detailed information about in-text references (e.g., "(Daquino et al. 2020)"), groupings of in-text references, discourse elements (including sentences, paragraphs, footnotes, captions, tables, sections), and citation annotations. Currently, it contains 5 million citations and metadata related to 2.5 millions of bibliographic resources (articles, issues, volumes, journals, etc.) in the biomedical domain.
- OpenCitations Meta: this collection, planned to be released in 2022 (originally, it was planned to be released at the end of 2021), will contain basic bibliographic metadata (but not citation data) about all the citing and cited entities described in the OpenCitations Indexes.

Table 6 – OpenCitations: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
<i>OpenCitations Indexes</i>												
OCI	X		X	X	X		X	X	X		X	X
DOI	X		X	X	X		X	X	X	X	X	X
PubMed ID									X	X	X	X
<i>OpenCitations Corpus</i>												
DOI	X	X			X	X			X	X		

ORCID	X	X	X	X	X	X	X	X	X	X	X	X
PubMed ID	X	X			X	X			X	X		
PubMed Central ID	X	X			X	X			X	X		
Crossref member ID	X	X			X	X			X	X		
ISSN	X	X			X	X			X	X		
ISBN	X	X			X	X			X	X		
Open Biomedical Citations in Context Corpus												
OCI	X	X	X	X	X	X	X	X	X	X	X	X
InTrePID	X	X	X	X	X	X	X	X	X	X	X	X
DOI	X	X			X	X			X	X		
ORCID	X	X	X	X	X	X	X	X	X	X	X	X
PubMed ID	X	X			X	X			X	X		
PubMed Central ID	X	X			X	X			X	X		
Crossref member ID	X	X			X	X			X	X		
ISSN	X	X			X	X			X	X		
ISBN	X	X			X	X			X	X		
OpenCitations Meta												
DOI									X	X		
ORCID									X	X		
Crossref member ID									X	X		
ISSN									X	X		
ISBN									X	X		

2.2.2.3 OPENAPC

OpenAPC collects cost data on Open Access publishing. Depending on the type of cost data different data schemas (e.g., for articles and books) are supported. The data schema model defines the mandatory, backup and optional fields, <https://github.com/OpenAPC/openapc-de/wiki/schema>. Submitted cost datasets from institutions are normalized and enriched by additional PIDs in OpenAPC, <https://github.com/OpenAPC/openapc-de/wiki/Data-Submission-Handout>. As illustrated by the Table below, PID management will be enhanced for Cost data delivery institutions and Publishers.

Table 7 – OpenAPC: PIDs support

PID type	Current status				OpenAIRE-Nexus: M13				OpenAIRE-Nexus: M30			
	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4	QD1	QD2	QD3	QD4
APC dataset and Transformative Agreement dataset												
DOI	X		X	X	X		X	X	X		X	X
URL	X		X	X	X		X	X	X		X	X
ISSN	X		X		X		X		X		X	

PubMed ID	X				X				X			
PubMed Central ID	X				X				X			
WOS ID	X				X				X			
BPC dataset												
DOI	X		X	X	X		X	X	X		X	X
ISBN	X		X		X		X		X		X	
Cost data delivering institutions												
GRID ID	X		X		X		X	P	X		X	P
ROR ID	X		X		X		X	P	X		X	P
Publisher												
GRID ID					P		P	P	X		P	P
ROR ID					P		P	P	X		P	P

2.2.3 DISCOVER

2.2.3.1 EXPLORE AND CONNECT

EXPLORE and CONNECT web applications were integrated with the ORCID Search and Link wizard toolkit, in order to enable users logging in via ORCID to formally claim their scientific results and complete their ORCID record via OpenAIRE. This process helps to strengthen and validate the adoption of ORCID identifiers by data sources that are not currently included in the ORCID life-cycle, such as data repositories, research infrastructure databases, scientific databases, etc.

3 TECHNICAL INTEGRATION WITH THE EOSC

Table 8 shows the EOSC technical frameworks that will be adopted by OpenAIRE-Nexus services and how these will be addressed by the specific portfolios and services. EOSC Core technical integration activities consist in the implementation of APIs (protocols and exchange formats) in order to comply with the general interoperability frameworks identified by the EOSC. We shall consider the following as currently endorsed by the EOSC:

- **EOSC onboarding:** registration of services into the EOSC Catalogue;
- **EOSC AAI:** enabling access to service UIs and APIs via EOSC AAI.

Moreover, OpenAIRE Nexus anticipated some of the decisions to be taken by the EOSC, by adopting standards in the domain of activities indicated as relevant for the EOSC Core:

- **Accounting and monitoring of services,** adopting standards to make services uniform in terms of delivery of measures and indicators of usage and consumption;
- **Data usage statistics,** adopting standards to make services uniform in terms of usage indicators for research products (e.g., publications, data, software),
- **RDA frameworks,** outcomes of RDA IGs and WGs, becoming de facto best practices for the communities.

Table 8 - OpenAIRE-Nexus Services and EOSC Technical Integration

Portfolio	Inst. No.	Service	EOSC onboarding	EOSC AAI	OpenAIRE Accounting (Virtual Access)	OpenAIRE Usage Stats	RDA Frameworks
PUBLISH	1	Zenodo.org	Done	Ongoing	Done	Done	
	2	Episciences.org	Done	Ongoing	Done	Done	
	3	Argos	Done	Done	Ongoing	Done	MaDMP
	4	Amnesia	Done	n/a	Done	Done	
MONITOR	5	MONITOR	Done	Done	Ongoing	Done	
	6	UsageCounts	Done	Done	Ongoing	n/a	Counter code of practice for Research Data
	7	OpenCitations.net	Done	n/a	Done	To be done	Scholix
	8	Schoexplorer	Done	n/a	Ongoing	To be done	Scholix
	9	OpenAPC	Done	n/a	Done	n/a	
	10	OS Observatory	Done	Done	Done	n/a	
	11	AAI	Done	To be done	Done	n/a	
DISCOVER	12	PROVIDE – Validator	Done	Done	Done	n/a	
	12	PROVIDE – Broker	Done	Done	Done	n/a	
	13	EXPLORE UI	Done	Done	Done	Done	

	13	EXPLORE APIs - DEVELOP	Done	Done	Done	Done	
	14	CONNECT	Done	Done	Done	Done	

3.1 EOSC onboarding

The minimal level of integration with the EOSC Core consists in services onboarding to the EOSC service catalogue. OpenAIRE-Nexus services are all onboarded by registering to the OpenAIRE service catalogue, one of the catalogues in-sync with the EOSC's.

3.2 EOSC AAI

According to the [EOSC AAI Federation Technical Interoperability Framework](#) entities acting as either Identity Providers or Relying Parties are required to implement and comply with the [AARC Interoperability Guidelines](#) approved by AEGIS. Relying parties are required to meet the following requirements:

- **Privacy policy:** Services need to document the data collected and processed. Services need to comply with the GEANT Data Protection [Code of Conduct](#) version 1 or any other code of conduct compatible with legislation and guidelines on data protection and privacy including GDPR (see [EOSC Rules of Participation](#)).
- **Acceptable Use Policy / Terms of Use:** Resource providers need to provide terms of use, including information about whether access requires authentication and authorisation, licencing, and any quotas or charges which may apply (see [EOSC Rules of Participation](#)). To reduce the burden on the users and increase the likelihood that they will read the terms as they access resources from multiple providers, it is recommended to adopt the [WISE Baseline AUP model](#).
- **Operational security and incident response policy:** Entities should meet the [Sirtfi](#) Security Framework requirements in order to facilitate coordinated response to security incidents across organisational boundaries. Specifically, the entity needs to perform the following tasks:
 - Passing a self-assessment of Sirtfi v1.0 or is known to be subject to a policy that encompasses all the requirements of the Sirtfi framework.
 - A security contact has been provided for the entity, or party providing Incident Response support on behalf of the entity.

Regarding technical requirements, the EOSC AAI currently supports two federation protocols, namely SAML2 and OpenID Connect:

1. Services implementing the SAML protocol are required to comply with the [SAML2int SAML 2.0 Interoperability Deployment Profile](#). The technical requirements are detailed in Section 6.3 of the [EOSC AAI Federation Technical Interoperability Framework](#).
2. Services implementing the OpenID Connect protocol need to support the technical requirements specified in Section 6.4 of the [EOSC AAI Federation Technical Interoperability Framework](#).

The OpenAIRE Login (AAI) service is already connected to the EOSC Portal AAI which is the Infrastructure Proxy service for the EOSC Core (EOSC Core Infrastructure Proxy). To alleviate the need for establishing trust with these services in a peer-to-peer manner (e.g., through the exchange of SAML metadata or static OIDC client registration mechanisms), the integration is expected to take place through the [EOSC AAI Federation](#). Specifically, OpenAIRE Login will be registered in the EOSC AAI Federation as follows:

1. As a SAML Service Provider (Infrastructure Proxy) to allow researchers from other Research Communities to access OpenAIRE services behind the OpenAIRE Login service
2. As a SAML Identity Provider (Community AAI) to allow OpenAIRE users to access services & resources offered by infrastructures participating in EOSC

The following activities are planned by M21 to meet the requirements of the EOSC AAI Federation:

- Update the OpenAIRE Login Acceptable Use Policy (AUP) to follow the [WISE Baseline AUP](#) and to include security contact information as required by the SIRTFI Security Framework
- Add support for user identifiers expressed as voPersonIDs [[AARC-G026](#)]. Specifically:
 - OpenAIRE Login will support consuming the user identifier released by Community AAI's through the `voPersonID` attribute
 - OpenAIRE Login will express the user identifier through:
 - the `voPersonID` attribute in the case of SAML services
 - the `voperson_id` Claim in the case of OIDC clients requesting either the `voperson_id` or `aarc` scope
- Add support for affiliation at home organisation information expressed as `voPersonExternalAffiliation` [[AARC-G025](#)]. Specifically:
 - OpenAIRE Login will support consuming affiliation at home organisation information released by Community AAI's through the `voPersonExternalAffiliation` attribute
 - OpenAIRE Login will express the affiliation at home organisation information through:
 - the `voPersonExternalAffiliation` attribute in the case of SAML services

- the `voperson_external_affiliation` claim in the case of OIDC clients requesting either the `voperson_external_affiliation` or `aarc scope`
- Add support for expressing assurance information as defined in REFEDS Assurance Framework (RAFv1.0) and [AARC-G021]

Argos already supports AAI integration with EOSC compliant technologies, which was been further extended during the last period.

At M14, for both Zenodo.org and episciences.org the activities are ongoing for adhering with the EOSC AAI user interfaces and integrate with the EOSC Core.

3.3 EOSC Accounting and monitoring for services

The EOSC-Future is currently developing a fully-fledged EOSC service onboarding workflow, which will include the options of monitoring and accounting for the services. Monitoring consists in checking the aliveness of the services via a list of service component entry points (URLs) organized into a topology. EOSC Accounting will instead provide APIs through which services will be able to push information about Virtual Access indicators and service-specific KPIs. OpenAIRE-Nexus service providers are contributing to such developments in strict collaboration with the EOSC-Future technical teams and will be among the early adopters of EOSC monitoring and accounting.

3.4 EOSC Accounting for research products

The EOSC accounting for research products is operated/powered by the OpenAIRE UsageCounts service, part of the OpenAIRE-Nexus portfolio. The service acquires usage statistics (views and downloads) about research products implementing the known standards COUNTER Code of Practice and the COUNTER Code of Practice for Research Data that provide standards-based usage statistics and enables comparability with statistics from other data sources.

OpenAIRE-Nexus services will integrate with OpenAIRE UsageCounts and therefore align to EOSC recommendations: OpenCitations, and Scholexplorer will send usage statistics regarding views of research products (access to the local “splash” pages for such products). As indicated by the table above, many OpenAIRE services are already compliant (e.g., Zenodo.org, Episciences, Argos, EXPLORE, CONNECT) while some are not generating such events.

3.5 RDA Frameworks

RDA WGs and IGs produced and will produce useful interoperability frameworks. In the first year of the project, we will consider the integration of maDMP, resulting from the RDA WG on

Actionable Data Management Plans, and the Counter Code of practices, resulting from the RDA WG on Data Usage statistics.

3.5.1 Argos

The RDA DMP Common Standards Working Group is paving the road towards machine-actionable Data Management Planning by developing a common information model and specifying the access mechanisms and protocols. With “machine-actionable” DMPs, a wide range of opportunities is opened, starting from the simple exchange of comprehensible information about a Data Management Plan, up to and even beyond declarations that will take all needed measures for verifying, depositing, describing, distributing, consuming and protecting datasets according to certain policies and intentions of their chain of actors. The common information model captures a variety of entities and concepts, such as funders, repositories, researchers, costs, manifestations and delivers candidate forms of those for actual implementation, the core one being the maDMP format. This format expressed in JSON is currently the main widely used carrier of DMP data, others being proprietary xml, json or document forms (OpenXML and PDF).

maDMP is of great interest to Open Science and the OpenAIRE infrastructure, as not only it promotes the adoption of FAIR principles, but it also helps to deliver machine readable associations of semantic links among data elements and actors around them, a feature of utter importance for the enrichment of the OpenAIRE Research Graph.

The team of Argos has onboarded the DMP Common Standards Working Group with several members of its design and implementation team and is pioneering the exposure and verification of the maDMP specification, exemplifying its first implementation in Summer 2020. Since this period there has been exchange of views and developments on the RDA WG, while the OpenAIRE Research Graph is also evolving. In this context, the collaboration between the OpenAIRE Nexus team and the RDA WG will be twofold. On the one hand, to follow the evolution of the maDMP model and contribute to its shaping and specification, until a common standard emerges. On the other hand, to exploit maDMP in the OpenAIRE Research Graph, as rich research data descriptive information container, mapping its elements to the OpenAIRE Research Graph core entities and attributes.

Furthermore, Argos team, being one of the leading DMP offerings, will seek to be represented in RDA Active DMPs IG and its product manager. Elli Papadopoulou is among the nominated new chairs to be elected by the group members in June 2022.

3.5.2 UsageCounts

Research data are different and more complex from other research products, like articles, when usage activity is considered. For example, multiple versions of data and multiple components might be available, and they are accessed and used in a variety of ways not only by humans but also by machines. The RDA WG on Data Usage Metrics worked towards the development of a

standard for Research Data that COUNTER formally endorsed as the official COUNTER recommendation for data. The WG is using this Code of Practice as a starting point for further recommendations.

Code of Practice for Research Data specifically targets research data usage. The recommendations are aligned as much as possible with the COUNTER Code of Practice Release 5 and mainly concern views and downloads – called investigations and requests in the Code of Practice. Many definitions, processing rules and reporting recommendations apply to research data in the same way as they apply to other resources. The Code of Practice for Research Data enables the reporting of usage statistics by different data repositories following common best practices, and thus is an essential step towards realizing usage statistics as a metric available to the community to better understand how publicly available datasets are being reused.

The UsageCounts service already exploits the *Code of Practice for Research Data* for research data usage statistics collected from Datacite and Institutional repositories and aggregates this information with the OpenAIRE Research Graph. The UsageCounts team will continue to participate in RDA Data Usage Metrics WG and work towards the improvement of the Code.