

D3.1 - Roadmap to semantic and technical interoperability v1.0



FEB. 2021

Scholarly Communication Services for EOSC users

D3.1 –Roadmap to semantic and technical interoperability v1.0

Version 3.0 –Final
PUBLIC

This deliverable details the initial roadmap to achieve semantic and technical interoperability of *OpenAIRE-Nexus* services, by taking as input the state of decisions and results of the EOSC-ENHANCE project, the EOSC-Secretarial WG (EAWG), EOSC-Hub, and OpenAIRE-Advance.

H2020-INFRAEOSC-2020-2
Grant Agreement 101017452

Document Description

D3.1 – Roadmap to semantic and technical interoperability v1.0

WP3 - Integration with the EOSC to provide Virtual Access

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

Contractual Delivery Date: 28/2/2021

Actual Delivery Date: 10/03/2021

Nature: Report

Version: 3.0 (Final)

Public

Preparation Slip

	Name	Organisation	Date
From	Paolo Manghi	OpenAIRE AMKE	3/3/2021
Edited by	Paolo Manghi Raphael Tournoy Dimitris Pierrakos Silvio Peroni Jochen Shirrswagen Nicolas Liampotis	OpenAIRE AMKE CNRS AthenaRC UNIBO UNIBI GRNET	3/3/2021
Reviewed by	Alessia Bardi	CNR	3/3/2021
Approved by	Paolo Manghi	OpenAIRE AMKE	4/3/2021
For delivery	Eleni Koulocheri	OpenAIRE AMKE	5/3/2021

Revision History

Issue	Item	Reason for Change	Author	Organization
V1.0	Draft version 13/01/2021	drafting	Editors	

V1.1	Draft version 2/02/2021	Input from partners	Editors	
V2.0	Draft version 07/03/2021	Input from revisions	Alessi Bardi	CNR
V2.1	Draft version 9/03/2021	Corrections based on revisions	Editors	
V2.2	Final version	Final version	Paolo Manghi	OpenAIRE AMKE
V3.0	Final formatted version	Applied format fixes	Paolo Manghi, Alessia Bardi	OpenAIRE AMKE, CNR

Contents

1	<i>Introduction</i>	7
2	<i>Semantic integration with the EOSC</i>	9
2.1	EOSC Resources	10
2.1.1	PUBLISH.....	11
2.1.2	DISCOVER.....	14
2.2	PIDs: ensuring PID are pervasively and correctly used by Nexus services	14
2.2.1	PUBLISH.....	15
2.2.2	MONITOR.....	16
2.2.3	DISCOVER.....	19
3	<i>Technical integration with the EOSC</i>	20
3.1	EOSC onboarding	21
3.2	EOSC AAI	21
3.3	Accounting and monitoring	22
3.4	Usage statistics	22
3.5	RDA Frameworks	23
3.5.1	Argos.....	23
3.5.2	UsageCounts.....	24

Table of Figures

Figure 1 - OpenAIRE-Nexus service portfolio.....	8
Figure 2 - EOSC Resource Data Model	11

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-M13

The aim of this deliverable is to draw the lines of action that will, in the first 13 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. The technical roadmap in OpenAIRE-Nexus will focus on the EOSC Core Interoperability Framework, 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.

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 13 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 to, 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 13 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 INFRAEOSC-03 awarded project. 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 to 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 minimize waste

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

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

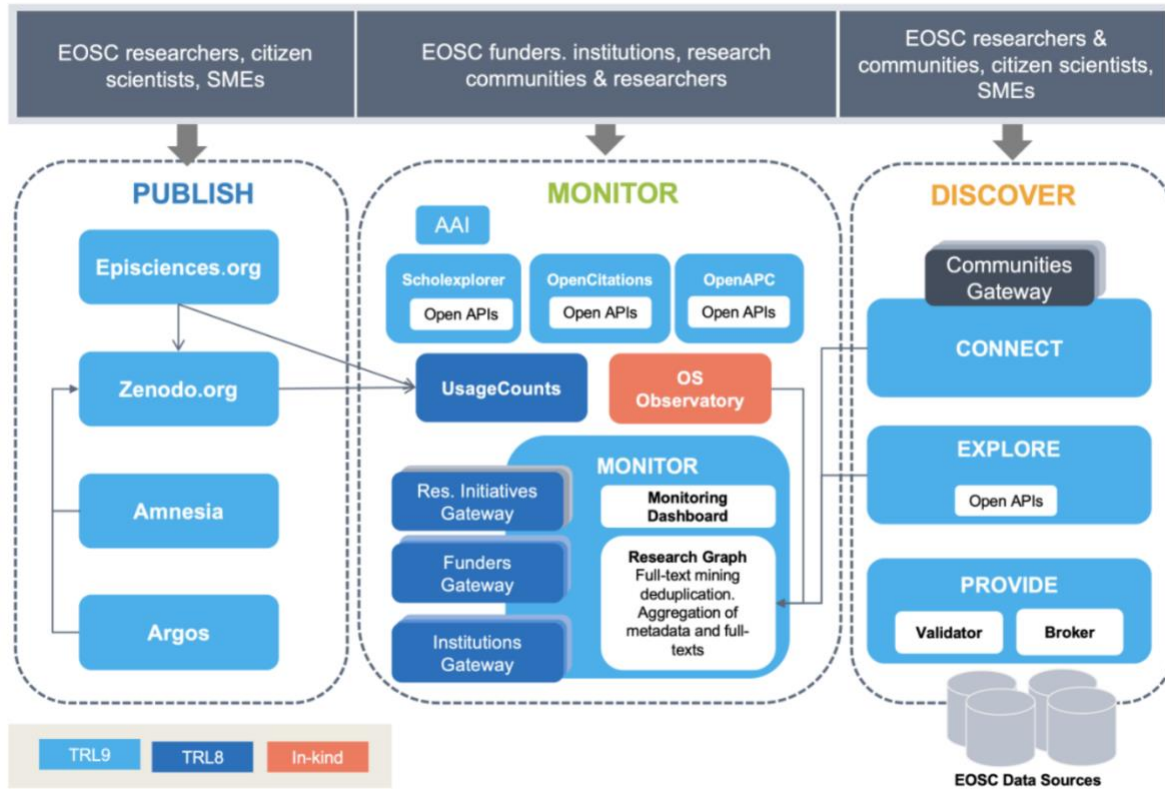


Figure 1 - OpenAIRE-Nexus service portfolio

2 SEMANTIC INTEGRATION WITH THE EOSC

Table 1 shows the semantic frameworks that will be adopted by OpenAIRE-Nexus services and how these will be 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.

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	Compliance to OpenAIRE guidelines for publishers	ORCID, DOIs - links to projects and links to other objects (datasets, software)
	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 (DOIs), organizations, ORCID
	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	DOIs, handles, URN, PubMed ID, ArXiv, orcid workid, Patent application number in EPODOC format, NII Citation ID, PubMed Central ID, European Patent Office application ID
	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 characterised by different EOSC resource types (text, dataset, software, etc.)	The types of PIDs that are/will be adopted by OpenCitations strictly depend on the particular collection in consideration. Several PID types will be added, as indicated by the table below.

	8	Schoexplorer	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	For publications: URL, DOI, PMID, PMCID, ISSN, ISBN For organisations (research performing institutions and publishers): GRID, ROR
	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 by the INFRAEOSC-03 funded project.

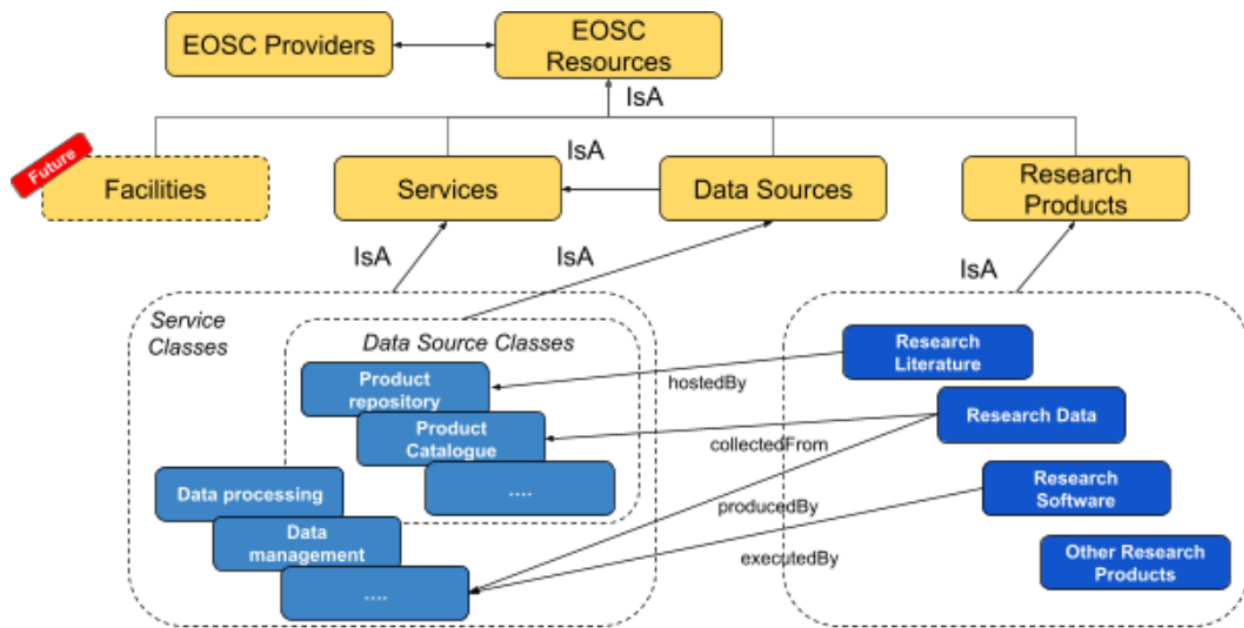


Figure 2 - EOSC Resource Data Model

Services: At the time of writing, only service profiles are endorsed and adopted by the community. The related specification, known as Service/Resource Description Template (SDT),² is the result of the eInfraCentral, EOSC-Hub, and EOSC-Enhance. Services must provide an SDT profile to onboard onto the EOSC, via the EOSC service catalogue. All OpenAIRE-Nexus services are registered to the OpenAIRE Service catalogue (<http://catalogue.openaire.eu>), which is compliant to the SDT and in-sync with the EOSC service catalogue.

Data sources: As a collaboration between the EOSC-Enhance and OpenAIRE-Advance, the SDT is being 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), as such they are related with them.

Research products: The OpenAIRE-Advance project has released, prior endorsement of research communities, a set of guidelines for scholarly communication data sources (<http://guidelines.openaire.eu>). Such guidelines are under refinement and endorsement via the EOSC-Enhance project, to become the EOSC research product profiles.

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 EOSC Service/Resource Description Template (SDT): <https://github.com/eInfraCentral/docs>

2.1.1.1 ZENODO.ORG

The service is already compliant with OpenAIRE Guidelines.

2.1.1.2 EPISCIENCES.ORG

The service will become compliant to the OpenAIRE Guidelines for content providers 4.0. This will make sure its journals and publications (articles) will be 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, more specifically:

- Entities provided by OpenAIRE are integrated in several areas of the current production deployment of Argos. Initial integration has been implemented in preliminary form (e.g. baseline filtering / ordering) and targets main entities, i.e. projects (grants), funders, organizations, datasets and licenses. In the course of the project, OpenAIRE metadata and vocabularies/taxonomies and protocols will be targeted.
- A mechanism to integrate external vocabularies and resources registries is provided. Services exposing those vocabularies under specific, yet simple to adopt, APIs can be utilized. In the course of the project, the mechanism will evolve to capture the emerging integration needs of new external vocabulary resources.
- Zenodo-provided DOIs are utilized for the linking of a DMP to its related research artifacts, however, this is not formalized, except for the publication of a DMP into Zenodo. The work will be to provide preliminary formalization for DOI support in other areas of DMPs.
- maDMP³ alignment with OpenAIRE Research Graph is an ongoing activity that will eventually provide into the research landscape modelled by OpenAIRE, in machine actionable form, the intentions and commitments of data producers and consumers.

2.1.1.4 AMNESIA

Amnesia generates anonymous datasets, which can be uploaded into Zenodo.org as datasets and therefore complies with the EOSC resource description framework.

2.1.1.5 MONITOR

The model of the Graph will be updated (in-kind) to include EOSC services as entity, together with links to other entities, such as research products. Mining algorithms to identify links between

³ RDA DMP Common Standard for machine-actionable Data Management Plans (maDMP) specification <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

services and research products (“publications referring/mentioning a service in the catalogue), will be integrated.

2.1.1.6 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 must also be compliant with both the COUNTER CoP⁴ for the major categories of e-resources (articles, books, etc) and the COUNTER CoP for Research Data⁵ in particular.

2.1.1.7 OPENCITATIONS

All the types of entities handled by OpenCitations are described in the OpenCitations Data Model (OCDM, <https://doi.org/10.6084/m9.figshare.3443876.v7>), which is implemented into the OpenCitations Ontology (OCO, <https://w3id.org/oc/ontology>) by aggregating terms from several existing ontologies including the Semantic Publishing And Referencing (SPAR) Ontologies (<http://www.sparontologies.net>), PROV-O (<https://www.w3.org/TR/prov-o/>) and Web Annotation Ontology (<https://www.w3.org/ns/oa>). The Data Model enables the description of published bibliographic resources, their related embodiments, bibliographic references, responsible agents and their roles with respect to bibliographic resources, citations, in-text reference pointers denoting bibliographic references that are embedded in the text of a document within the context (i.e. a discourse element) of a particular sentence, paragraph or section.

The list of types of bibliographic resources available in the current release of the OCDM is capable of describing the main part of the research products as envisioned in the EOSC data model, namely publications and datasets. Future development of the OCDM will consider to include software among the types of bibliographic resources.

Finally, one of the formats used to export and provide citation data included in the OpenCitations collections is Scholix (<http://www.scholix.org>). Scholix is an interoperability framework for exchanging information about the links between scholarly literature and data maintained by the RDA/WDS Scholarly Link Exchange working group (<https://www.rd-alliance.org/groups/rdawds-scholarly-link-exchange-scholix-wg>).

2.1.1.8 SCHOLEXPLORER

The model of the service currently describes links (semantic relationships) between publications and datasets but will be extended to include links to EOSC software entities. The change will fully

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

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

align the data models of Schoexplorer and the OpenAIRE Research Graph, exposing resolution of PIDs via Schoexplorer for a wider pool of research products.

2.1.1.9 OPENAPC

The underlying schemas of the service allow the recording of cost information of Open Access publishing on the levels of publications (articles, books, publishers) and institutions (e.g. OA transformative agreements). Through integration with the OpenAIRE Research Graph, the analytical options are extended by complementary entities such as funders and projects related to publications.

2.1.2 DISCOVER

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. For the moment their implementation will be postponed to the release 2.0 of this deliverable.

2.1.2.2 EXPLORE AND CONNECT

EXPLORE and CONNECT offer UIs to search and browse the OpenAIRE Research Graph and will be adapted to include services as discovery and interlinking entities. Users will be able to search and browse, but also to “link” (claim functionality) services to other entities. If needed, the UI will also be enhanced to accommodate new information introduced by the adoption of the new EOSC data source profiles currently under definition.

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 M13 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 Datacite 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.

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, although it might be a part of metadata information, could 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 WorkID	X		X	X	X		X	X	X		X	X
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 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		
<i>Open Biomedical Citations in Context Corpus</i>												
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		
<i>OpenCitations Meta</i>												
DOI					X	X			X	X		
ORCID					X	X			X	X		
Crossref member ID					X	X			X	X		
ISSN					X	X			X	X		
ISBN					X	X			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 will integrate 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 will 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 to 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, we shall anticipate 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	To be done	To be done	Done	
	2	Episciences.org	Done	To be done	To be done	To be done	
	3	Argos	Done	Done	To be done	To be done	MaDMP
MONITOR	4	Amnesia	Done	n/a	To be done	Done	
	5	MONITOR	Done	Done	To be done	Done	
	6	UsageCounts	Done	Done	To be done	n/a	Counter code of practice for Research Data
	7	OpenCitations.net	Done	n/a	To be done	To be done	Scholix
	8	Scholexplorer	Done	n/a	Done	To be done	Scholix
	9	OpenAPC	Done	n/a	To be done	n/a	
	10	OS Observatory	Done	Done	To be done	n/a	
	11	AAI	Done	To be done	To be 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 AAI is already connected to the EOSC Portal AAI which is the Infrastructure Proxy service for the EOSC Core. Connection with other Infrastructure Proxy services (e.g. EGI, EUDAT, GEANT) will be established during the first 13 months of the project. 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](#).

In the first 13 month of the project, both Zenodo.org and episciences.org will adhere to the EOSC AAI user interfaces and integrate with the EOSC Core.

3.3 Accounting and monitoring

The EOSC has not yet provided recommendations regarding the modelling, sharing, and acquisition of indicators of usage, capacity, etc. as well as monitoring protocols. The EOSC working groups, to continue their work under the INFRAEOSC-03 funded project, are however proposing/drafting solutions based on standards and therefore easily interoperable with other standards in the field. On top of this, OpenAIRE Nexus (as well as other projects funded in INFRAEOSC-07) will have to establish an accounting and monitoring framework in order to yearly report to the Commission the units of access. Accordingly, the agreed methodology is to make sure all services in Nexus' portfolio will implement the same standard technologies to provide accounting information centrally to the OpenAIRE Catalogue. Such standards, based on Prometheus.io⁶, will facilitate interoperability with any standard the EOSC Core may identify in the future and therefore prompt reporting of accounting info to the Commission via the OpenAIRE catalogue.

3.4 Usage statistics

The EOSC will endorse common standards for collecting usage statistics (see section RDA Frameworks below). As OpenAIRE Nexus will operate OpenAIRE UsageCounts, which in turn will comply to such EOSC-endorsed standards and offer collection and aggregation of usage statistics, the following OpenAIRE-Nexus services will integrate with OpenAIRE UsageCounts and therefore

⁶ <https://prometheus.io/>

align to EOSC recommendations: episciences.org, Argos, 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, EXPLORE, CONNECT) while some are not generating such events.

To this aim, services will comply with both the 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. Accordingly, services will register to OpenAIRE UsageCounts and send information about usage activity by allowing for anonymization of IP-addresses to comply with privacy policies.

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.

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.