Recommendations for the EOSC Interoperability Framework from the EOSC Future WP3 Working Group on Research Product Publishing

Edited by
Alessia Bardi, CNR-ISTI, OpenAIRE Nexus WP7 Leader

WG chairs
Alessia Bardi, CNR-ISTI, OpenAIRE Nexus WP7 Leader
Jose Benito Gonzalez Lopez, CERN, Zenodo.org technical manager
Paolo Manghi, OpenAIRE, OpenAIRE Nexus coordinator, EOSC Future TCB member

WG members

- Chris Ariyo, EUDAT B2SHARE service owner
- Andreas Czerniak, Bielefeld University Library/OpenAIRE-Nexus
- Paul Gondim van Dongen, SURF
- Georgios Kakaletris, NEANIAS, Project Technical Manager
- Raul Palma, PSNC, RELIANCE coordinator
- Silvio Peroni, University of Bologna, Director of OpenCitations
- Hans van Piggelen, SURF
- 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
- Debora Testi, CINECA, DICE coordinator
- Irena Vipavc, Social Science Data Archive, University of Ljubljana
- Grbac Deborah, Università Cattolica del Sacro Cuore di Milano
- Carl-Fredrik Enell, EISCAT Scientific Association
- Guido Aben, CS3MESH4EOSC
- Ivan Heibi, OpenCitations
- Jorik van Kemenade, SURF, Manager of the SURF data repository

Recommended protocols for research product deposition

Push mode: SWORD v3

SWORD v3 comes with detailed documentation. In particular, the Protocol Behaviours document describes the requirements of clients and servers for a specific request.
A complete deposition succeeds when both metadata and files are successfully accepted and stored by the server.
According to the SWORD specification, servers must support Metadata Document in the DCMI format. This recommendation suggests the adoption of the format of the EOSC guidelines for research product onboarding, in particular the format of the OpenAIRE guidelines version 4, which are based on the Datacite format.
The support must be declared by servers and used in clients' request as explained in Section 19 Metadata Deposit, the identifier of the format (in form of URI) has to be decided.
Domain-specific communities may agree to accept additional metadata formats to support domain-specific use cases that cannot be addressed without domain-specific metadata formats.

Requests that are relevant in our context are:

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

Pull mode: COAR Notify + Signposting

COAR Notify + Signposting: the service can inform the repository that something new is available at a given accessible location (with COAR Notify) and the repository can then use the Signposting protocol (implemented by the service) to know where to get the metadata and payload file(s) for the deposition.

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

Recommended metadata exchange formats

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

Recommended protocols for accessing the content behind a PID

SignPosting (publication boundary)
No additional indications are needed.

Known implementation

The page https://signposting.org/adopters/ lists the known adopters. However, not all adopters implement the Publication Boundary Pattern that is relevant for our context. Those that do implement the pattern are listed below.

Open Journal System (OJS)

Open journal system is a platform for the management of research journals.
We count about 1K journals using OJS contributing to the OpenAIRE Research Graph.

Type of implementation: via linkset

Example:
1. curl -I https://doi.org/10.4401/ag-7507 to get the Location
2. curl -I http://www.annalsofgeophysics.eu/index.php/annals/article/view/7507 to get the Location again
3. curl -I https://www.annalsofgeophysics.eu/index.php/annals/article/view/7507 to know the linkset
HTTP/1.1 200 OK
Date: Fri, 10 Mar 2023 16:35:07 GMT
Server: Apache
Link: <https://www.annalsofgeophysics.eu/index.php/annals/sp-linkset/article/7507>; rel="linkset"; type="application/linkset+json"
. . .

5. curl https://www.annalsofgeophysics.eu/index.php/annals/sp-linkset/article/7507 to get the json file. It contains

"**item**": [
{
"href": "https://www.annalsofgeophysics.eu/index.php/annals/article/download/7507/6808",
"type": "application/pdf"
}
]

Pangaea

Pangaea is a data repository hosting more than 400K research data. It is a repository registered in the EOSC Marketplace and its research products are available in EOSC EXPLORE.

Type of implementation: via HTTP link headers

Example:

1. curl -I https://doi.org/10.1594/PANGAEA.954506 to get the Location
2. curl -I https://doi.pangaea.de/10.1594/PANGAEA.954506 to get the 'link' information. Find the entry with 'rel="item":

HTTP/2 200
. . .
<https://doi.pangaea.de/10.1594/PANGAEA.954506?format=metadata_dif>;rel="describedby";type="application/vnd.nasa.dif-metadata+xml",
<https://doi.pangaea.de/10.1594/PANGAEA.954506?format=zip>;rel="item";type="application/zip",
<https://orcid.org/0000-0002-2078-0361>;rel="author",
<https://orcid.org/0000-0001-7313-100X>;rel="author",
. . .

DSpace CRIS

According to https://signposting.org/adopters/, starting with version 5.8.2, the open source DSpace-CRIS system has built-in support for the Publication Boundary pattern.
About 20 CRIS systems are currently contributing to the OpenAIRE Research Graph, but some of them might not support the Publication Boundary (it depends on the specific platform they use and, for those that use DSpace CRIS, the specific version).

Type of implementation: via HTTP link headers

Example from the TUHH Open Research (TORE) platform:

1. curl -I http://hdl.handle.net/11420/14772 to get the Location
2. curl -I curl -I https://tore.tuhh.de/handle/11420/14772 to get the Link with rel=”item”

   HTTP/1.1 200 200
   . . .
   Link: https://doi.org/10.15480/336.4925; rel="cite-as"
   Link: https://tore.tuhh.de/bitstream/11420/14772/4/Readme.pdf; rel="item"; type="application/pdf"
   Link: https://tore.tuhh.de/bitstream/11420/14772/5/Recorded_data.zip; rel="item"; type="application/zip"
   Link: http://orcid.org/0000-0003-3288-7892; rel="author"
   Link: http://orcid.org/0000-0003-0505-4836; rel="author"
   Link: http://orcid.org/0000-0002-3336-4383; rel="author"
   Content-Type: text/html;charset=UTF-8
   Content-Language: de
   Vary: Accept-Encoding

Example from IZTECH GCRIS:

1. curl -I https://hdl.handle.net/11147/13225 to get the Location
2. curl -I https://gcris.iyte.edu.tr/handle/11147/13225 to get the Link with rel=”item”

   HTTP/1.1 200
   . . .
   Link: https://doi.org/10.1080/15567036.2023.2171512; rel="cite-as"
   Link: https://gcris.iyte.edu.tr/bitstream/11147/13225/1/Air%20density%20calculation%20at%20high%20altitude.pdf; rel="item"; type="application/pdf"
   Link: http://orcid.org/0000-0002-8071-3814; rel="author"