



## **EOSC IF Interoperability Guideline: Access to content via PID**

Alessia Bardi, Paolo Manghi, Jose Benito Gonzalez Lopez, Chris Ariyo, Andreas Czerniak, Paul Gondim van Dongen, Georgios Kakalettris, Raul Palma, Silvio Peroni, Hans van Piggelen, et al.

### **► To cite this version:**

Alessia Bardi, Paolo Manghi, Jose Benito Gonzalez Lopez, Chris Ariyo, Andreas Czerniak, et al.. EOSC IF Interoperability Guideline: Access to content via PID. EOSC Future Working Group on "Research Product Publishing". 2023. <hal-04636955>

**HAL Id: hal-04636955**

**<https://hal.science/hal-04636955v1>**

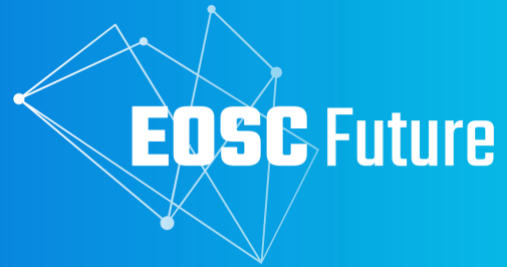
Submitted on 5 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



# **EOSC-IF**

## **Interoperability Guideline: Access to content via PID**

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



Version 1.0  
Month Year

# EOSC-IF / Interoperability Guideline: Access to content via PID

Lead by **CNR**

Authored by the EOSC Future Working Group on Research Product Publishing: Alessia Bardi (CNR), Paolo Manghi (OpenAIRE), Jose Benito Gonzalez Lopez (CERN), Chris Ariyo (EUDAT), Andreas Czerniak (Bielefeld University Library), Paul Gondim van Dongen (SURF), Georgios Kakalettris (NKUA), Raul Palma (PSNC), Silvio Peroni (University of Bologna), Hans van Piggelen (SURF), Mark van de Sanden (EUDAT), Diego Scardaci (EGI), Jochen Schirrwagen (Bielefeld University Library), Debora Testi (CINECA), Raphaël Tournoy (CNRS), Irena Vipavc (Social Science Data Archive, University of Ljubljana), Deborah Grbac (Library of Università Cattolica del Sacro Cuore di Milano), Carl-Fredrik Enell (EISCAT Scientific Association), Guido Aben (CS<sub>3</sub>MESH<sub>4</sub>EOSC), Ivan Heibi (University of Bologna), Jorik van Kemenade (SURF)

Reviewed by Michelle Williams (GÉANT)

## Dissemination Level of the Document

Public

## Abstract

An important aspect of Open Science is the possibility to re-use existing research products (e.g. research data), deposited in repositories and accessible via their persistent identifiers (e.g. handle, doi, ark). However, there is no standard way a service can access the actual content behind persistent identifiers, as these typically resolve to the landing pages of the research products.

The lack of standard for accessing the actual content identified by persistent identifiers makes the automatic consumption of research products hardly implementable and, when possible, limited to the persistent identifiers issued by a specific repository (e.g. the first prototype of the EGI Data Transfer Service integrated in the EOSC EXPLORE portal supported only DOIs from Zenodo).

The EOSC Future Working Group on Research Product Publishing proposes the adoption of the Publication Boundary Pattern of the SignPosting protocol and recommends it for inclusion as interoperability guideline in the EOSC IF.

## Version History

Version	Date	Authors/Contributors	Description
V0.1	08/05/2023	Alessia Bardi (CNR)	Initiation
V1.0	27/06/2023	Alessia Bardi (CNR)	Final version for submission

## Copyright Notice



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

## Table of Contents

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

## Glossary

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

## List of Abbreviations

Acronym	Definition
CRIS	Current Research Information System
EOSC	European Open Science Cloud
OJS	Open Journal System
PID	Persistent identifier
URL	Uniform Resource Locator
WG	Working Group

## 1 Intended Audience

- Data sources hosting payloads of research products of any type. Examples: data archives, data repositories, thematic repositories, institutional repositories, pre-print repositories, journal publishing platforms
- Services that would like to access the payloads of a research product given its persistent identifier

## 2 Description and main features

Persistent Identifiers (PIDs) are unique identifiers of resources that are usually resolvable URLs (e.g. DOI, Handle).

Given a PID URL, we can perform an HTTP request to get the resource. Typically, the returned content is the metadata of the resource (a description of it). Metadata can be returned in different format, depending on the server and on the specific HTTP request: an HTML landing page, json-ld, RDF are some examples. The feature is supported by the concept of “content-negotiation”, but there is no standardise mechanism to get directly to the actual resource, bypassing the landing page/metadata.

Therefore, software programs must implement specific strategies for specific servers, or crawl the landing pages to identify the URL from which the resource can be downloaded.

With this guideline, we suggest standard protocols that can be used to support software programs at consuming the resources identified by PIDs in a consistent way across servers and regardless the specific type of PID.

## 3 Response to Community Need

A researcher uses a thematic service to run analysis on a dataset available on a repository. Instead of downloading the dataset files from the repository and uploading them to a storage resource of the e-infrastructure, the researcher gives the dataset’s PID as input to the thematic service, which can get the files and store them where they can be analysed.

## 4 Licensing Information

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

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

## 5 Related Standards

Table 1 Related Standards

Title	Short Description	relatedStandardIdentifier
<b>SignPosting Publication Boundary Pattern</b>	Landing pages support humans that interact with scholarly objects on the web, providing descriptive metadata and links to content. These pages are not optimized for use by machine agents that navigate the scholarly web. For example, how can a robot determine which links on the myriad of landing pages lead to content and which to metadata? Signposting caters to machine agents by providing this information, and more, in a standards-based way.	<a href="https://signposting.org/publication_boundary">https://signposting.org/publication_boundary</a>

## 6 Integration Options

The SignPosting Publication Boundary Pattern can be implemented in two ways:

- via Linkset
- via HTTP link headers

## 7 Interoperability Guidelines

The Interoperability Guidelines are defined by the [SignPosting protocol, Publication Boundary Pattern](#)

## 8 Examples of solutions implementing this specification

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 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 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"

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