
OpenAIRE Guidelines

Release 3.0 alpha

OpenAIRE

Sep 16, 2021

Contents

1	Introduction	3
2	Application Profile Overview	5
3	What's new	7

These guidelines describe the application profile v3 for Data Archive managers to be compatible with OpenAIRE.

The OpenAIRE (*Open Access Infrastructure for Research in Europe*) Guidelines were established to support the [Open Access/Open Science strategy of the European Commission](#) and to meet requirements of the OpenAIRE infrastructure. This new version of the Guidelines, according to the expansion of the aims of the OpenAIRE initiative and its infrastructure, has a broader scope. In fact, these Guidelines are intended to guide repository manager to expose to the OpenAIRE infrastructure *open access and non-open access* publications together with funding information, where applicable.

1.1 Aim

The *OpenAIRE Guidelines for Data Archive Managers 3.0* provide orientation for software repository managers to define and implement their local data management policies in exposing metadata for data products according to the requirements of the OpenAIRE - Open Access Infrastructure for Research in Europe. These guidelines are intended to provide indications on how to make dataset products citable in order to make them first-level citizen of an Open Science, interlinked scholarly communication ecosystem. By adhering to the guidelines exposure, visibility, and re-use of repository content will be significantly increased.

By implementing the OpenAIRE Guidelines, data archive managers are facilitating the creation of enhanced publications and building the stepping-stones for a linked data infrastructure for research.

According to the Content Acquisition Policies (CAP, [10.5281/zenodo.1446408](#)) of the OpenAIRE infrastructure, metadata from data archives can be included and shown in the OpenAIRE Research Graph without any restrictions.

OpenAIRE is happy to assist in adherence to these guidelines.

1.2 Rationale

The goal of the OpenAIRE guidelines for dataset is to give immediate visibility of datasets as a “citable research product” based on the current state of the art in the scholarly communication, while indicating the way towards “good dataset citation practices”. Research dataset is currently available from the following kinds scholarly communication repositories:

- Institutional repositories: datasets descriptions are currently provided as Dublin Core metadata records
- Data repositories: dataset descriptions are currently provided as DataCite metadata records

The guidelines aim at making these repositories readily compliant so as to start exposing dataset entities to discovery and citation services. This means the guidelines should be endorsed by the community (e.g. include properties that reflect the need of dataset citation), do not impose high efforts to sources (e.g. mandatory citation metadata not available to sources), while recommending best practices (e.g. placing metadata recommended/optional for citation). Accordingly, the guidelines have been defined with a pragmatic approach, keeping mandatory properties to the minimum, focusing on properties for citation (attribution and access), disregarding discover-for-reuse properties, but keeping in mind that any property can be added in the future to reflect changes that should and hopefully will occur at the repositories side and in the behaviour of scientists who create, share, cite, and re-use research datasets.

The guidelines take inspiration from the following initiatives on datasets description and citation:

- [DataCite](#)

1.3 Acknowledgements & Contributors

1.3.1 Editors

- Andreas Czerniak (Bielefeld University Library, Germany, [orcid.org/](#))
- Aenne Loehden (Bielefeld University Library, Germany, [orcid.org/](#))

1.3.2 Experts & Reviewers

1.3.3 Versions

- 3.0-Draft June 2020, Updated to DataCite Metadata Schema v4.3 ([10.14454/f2wp-s162](#))
- [2.0 April 2014](#) , Updated to DataCite Metadata Schema v3.0
- 1.0 December 2012, Initial document

1.3.4 Citation

Application Profile Overview

The properties of the Application Profile for OpenAIRE Guidelines for Data Archives are listed in this section. The following requirement levels for the metadata properties are used:

Mandatory (M) The property must always be present in the metadata. An empty value for the property is not allowed.

Mandatory if Applicable (MA) When the property value can be obtained it must be present in the metadata

Recommended (R) The use of the property is recommended

Recommended than Mandatory (RtM) If the recommend element is used, then this is Mandatory.

Optional (O) It is not important whether the property is used or not, but if used it may provide complementary information about the resource

Optional than Mandatory (OtM) If the optional element is used, then this is Mandatory.

This documentation uses the following namespace abbreviations:

- datacite: <http://datacite.org/schema/kernel-4>

OpenAIRE-Field	Metadata Element	Refinement by Vocabulary
dc:title	datacite:title	title type
dc:creator	datacite:creator	name type
dc:contributor	datacite:contributor	name type contributor type
dc:datePublication	datacite:date	date type
dc:publicationYear	datacite:date	date type
dc:publisher	datacite:publisher	
dc:subject	datacite:subject	
dc:description	datacite:description	
dc:language	datacite:language	Allowed values are taken from IETF BCP 47, ISO 639-1 language codes. Examples: en, de, fr
dc:identifier	datacite:identifier	identifier type
dc:alternativeIdentifier	datacite:alternateIdentifier	alternateIdentifier type
dc:relatedIdentifier	datacite:relatedIdentifier	relatedIdentifier type relation type resourcetype general
dc:resourceType	datacite:resourceType	‘COAR Resource Type Vocabulary’_
dc:accessrights	datacite:rights	‘COAR Access Right Vocabulary’_
dc:size	datacite:size	
dc:version	datacite:version	‘COAR Version Vocabulary’_
dc:geolocation	datacite:geoLocation	
dc:fundingReference	datacite:fundingReference	funderIdentifier type

The application profile is implemented in XML Schema.

Not listed elements from DataCite schema v4.3 could be used as further optional (O) elements.

What's new

For the data guidelines version **3.0** at hand the major change is the harmonisation with the latest DataCite schema version 4.3 ([10.14454/r2wp-s162](https://doi.org/10.14454/r2wp-s162)).

This version updates the vocabulary for some elements in the application profile.

- Full alignment to the *OpenAIRE Content Acquisition Policies* (CAP, [10.5281/zenodo.1446408](https://zenodo.org/record/1446408)) published in Aug. 2018.
- The rights element has updated its vocabulary to [COAR Access Rights Vocabulary](#)
- Following DataCite, funding statements are moved from contributor to the dedicated field `fundingReference`.

3.1 How to contribute

Your feedback, esp. as a repository manager, is important for us. You can provide us feedback using the following channels:

- by commenting the Data Archive Feedback Document at Google: <https://...>
- by creating an issue in our guidelines github repository (you need to have a github account): <https://github.com/openaire/guidelines-data-archives/issues>
- by making annotations on the guidelines pages (you need to have an account for <https://web.hypothes.is>): just highlight some text and you can annotate the document
- by sending an e-mail to: *guidelines (at) openaire . eu*