Contents

Introduction .................................................................................................................3
Aim ............................................................................................................................. 3
DataCite ..................................................................................................................... 3
What’s different .......................................................................................................... 3

Acknowledgements & Contributors .....................................................................5
Editors ....................................................................................................................... 5
Experts & Reviewers ................................................................................................. 5

License .......................................................................................................................6

Version .......................................................................................................................6

OpenAIRE application of DataCite Metadata Schema ........................................7
Access rights information .......................................................................................... 12
Funding information .................................................................................................. 12
Related publications and datasets information ....................................................... 12
Embargo date information ........................................................................................ 13

Use of OAI-PMH .....................................................................................................14
Metadata Format ....................................................................................................... 14
OpenAIRE OAI Set ..................................................................................................... 14
Set content ................................................................................................................ 14

Further instructions ................................................................................................15
Introduction

Aim

The OpenAIRE Guidelines for Data Archive Managers 1.0 will provide instruction for data archive managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure. Data archives should be included in the OpenAIRE information space when data are related to a document e.g. a dataset cited by an article.

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.

DataCite

OpenAIRE has adopted the DataCite Metadata Schema as the basis for harvesting and importing metadata about datasets from data archives.

The core mission of DataCite is to build and maintain a sustainable framework that makes it possible to cite data through the use of persistent identifiers, DOIs.

OpenAIRE shares the goal of the DataCite Metadata Schema - to provide a domain agnostic metadata schema and provide interoperability through a small number of properties - making interoperability possible in the simplest manner possible and as a result keep the technical barriers for implementation as low as possible.

We would like to thank DataCite for their kind support and help making these OpenAIRE guidelines for Data Archive Managers possible.

What’s different

In this document you will find the needed properties to make your data archive compatible with the OpenAIRE infrastructure. OpenAIRE has adopted the DataCite Metadata Schema version 2.2 with some minor adjustments.

- OpenAIRE accepts other persistent identifier schemes and not only a DOI in 1.1 identifierType.
- OpenAIRE recommends links to related publications and datasets i.e. 12. RelatedIdentifier property and attributes are mandatory when applicable.
• OpenAIRE recommends using the 7. Contributor property to relate a dataset to funding information.
• DataCite optional properties 8. Date and 17. Description are mandatory in OpenAIRE.
• OpenAIRE enforces encoding schemes on DataCite property 16. Rights.
Acknowledgements & Contributors

Editors

- Mikael K. Elbæk (Technical University of Denmark)
- Lars Holm Nielsen (CERN, Switzerland)

Experts & Reviewers

- Samuele Kaplun (CERN, Switzerland)
- Paolo Manghi (CNR, Italy)
- Natalia Manola (University of Athens, Greece)
- Jochen Schirrwagen (University of Bielefeld, Germany)
- Birgit Schmidt (University of Goettingen, Germany)
- Mathias Lösch (University of Bielefeld, Germany)
- Frauke Ziedorn (DataCite, Germany)
- Pedro Principe (University of Minho, Portugal)
- Eloy Rodrigues (University of Minho, Portugal)
- Najla Rettberg (University of Goettingen, Germany)
License

This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License.

Version

| 1.0 | December 2012 | Initial document |
OpenAIRE application of DataCite Metadata Schema

OpenAIRE builds on the DataCite Metadata Schema by making some of the otherwise optional DataCite properties mandatory, as well as enforcing more specific encoding schemes on the values of some DataCite properties. The table below details the differences from the DataCite Metadata Schema.

Within OpenAIRE the use of elements and attributes is either:

- **mandatory (M)** = the element must always be present in the metadata record. An empty element is not allowed.
- **mandatory when applicable (MA)** = when the element can be obtained it must be present in the metadata record
- **recommended (R)** = the use of the element is recommended
- **optional (O)** = it is not important whether the element is used or not

The recommended status is made primarily to encourage users to input certain elements when creating a metadata record to enhance services.

<table>
<thead>
<tr>
<th>ID</th>
<th>DataCite-property</th>
<th>Status</th>
<th>Encoding schemes (if different from DataCite)</th>
</tr>
</thead>
</table>
| 1   | Identifier        | M         | DOI (Digital Object Identifier) is preferred.  
The format should be “10.1234/foo”. |
| 1.1 | identifierType    | M         | OpenAIRE allows for others types of identifiers than DOIs compared to DataCite.  
*Controlled List:*
  Allowed values:  
  ARK  
  DOI  
  Handle  
  PURL  
  URI |

1 Only differences to the DataCite Metadata Schema v2.2 are noted here. Please refer to encoding schemes and allowed values in [http://schema.datacite.org/meta/kernel-](http://schema.datacite.org/meta/kernel-).
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Creator</td>
<td>M</td>
</tr>
<tr>
<td>2.1</td>
<td>creatorName</td>
<td>M</td>
</tr>
<tr>
<td>2.2</td>
<td>nameIdentifier</td>
<td>R</td>
</tr>
<tr>
<td>2.2.1</td>
<td>nameIdentifierScheme</td>
<td>R</td>
</tr>
<tr>
<td>3</td>
<td>Title</td>
<td>M</td>
</tr>
<tr>
<td>3.1</td>
<td>titleType</td>
<td>O</td>
</tr>
<tr>
<td>4</td>
<td>Publisher</td>
<td>M</td>
</tr>
<tr>
<td>5</td>
<td>PublicationYear</td>
<td>M</td>
</tr>
<tr>
<td>6</td>
<td>Subject</td>
<td>O</td>
</tr>
<tr>
<td>6.1</td>
<td>subjectScheme</td>
<td>O</td>
</tr>
<tr>
<td>7</td>
<td>Contributor</td>
<td>MA</td>
</tr>
<tr>
<td>7.1</td>
<td>contributorType</td>
<td>MA</td>
</tr>
<tr>
<td>7.2</td>
<td>contributorName</td>
<td>MA</td>
</tr>
<tr>
<td>7.3</td>
<td>nameIdentifier</td>
<td>MA</td>
</tr>
</tbody>
</table>
The projectID equals the Grant Agreement number, and is defined by the namespace\(^2\)

\[
\text{info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/[Jurisdiction]/[ProjectName]/[Project Acronym]}/
\]

For OpenAIRE compatibility, the first three elements are mandatory and the last three are optional. Repositories may choose to use only the first part of the namespace (as in previous versions of the Guidelines), or use the extended version with six parts, which is recommended.

When omitting fields in the extended version, the number of fields must nevertheless be preserved by using “/”. A correct example for omitting the the ProjectName field would therefore look like this:

EC/FP7/12345/EU//OpenAIREplus

| 7.3.1 | nameIdentifierScheme | MA | Applicable only when contributorType is “Funder”:
| Controlled List |
| Allowed values: |
| info |

| 8 | Date | M | Mandatory property in OpenAIRE instead of optional in DataCite. |

| 8.1 | dateType | M | Use “Issued” for the date the resource is published or distributed. To indicate the end of an embargo period, use “Available”. |

\(^2\) See [http://wiki.surffoundation.nl/display/standards/info-eu-repo/#info-eu-repo-GrantAgreementIdentifiers](http://wiki.surffoundation.nl/display/standards/info-eu-repo/#info-eu-repo-GrantAgreementIdentifiers)
To indicate the start of an embargo period, use “Accepted”.

Further dateTypes may be specified. Please refer to DataCite Metadata Schema v2.2 for details.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Language</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>ResourceType</td>
<td>R</td>
</tr>
<tr>
<td>10</td>
<td>resourceTypeGeneral</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>AlternateIdentifier</td>
<td>O</td>
</tr>
<tr>
<td>11.1</td>
<td>alternateIdentifierType</td>
<td>O</td>
</tr>
<tr>
<td>12.</td>
<td>RelatedIdentifier</td>
<td>MA</td>
</tr>
<tr>
<td>12.1</td>
<td>relatedIdentifierType</td>
<td>MA</td>
</tr>
<tr>
<td>13</td>
<td>Size</td>
<td>O</td>
</tr>
<tr>
<td>14</td>
<td>Format</td>
<td>O</td>
</tr>
<tr>
<td>15</td>
<td>Version</td>
<td>O</td>
</tr>
<tr>
<td>16</td>
<td>Rights</td>
<td>MA</td>
</tr>
</tbody>
</table>

Mandatory when applicable property in OpenAIRE instead of optional in DataCite.

Controlled List

Allowed values (most common, please refer to DataCite Metadata Schema v2.2 for a complete list)

- IsCitedBy (indicates that B includes A in a citation)
- Cites (indicates that A includes B in a citation)
- HasPart (indicates A includes the part B)
- IsReferencedBy (indicates A is used as a source of information by B)

Use values from vocabulary Access Rights at [http://wiki.surf.nl/display/standards/info-](http://wiki.surf.nl/display/standards/info-)
eu-repo/info-eu-repo-AccessRights; values are:

info:eu-repo/semantics/closedAccess
info:eu-repo/semantics/embargoedAccess
info:eu-repo/semantics/restrictedAccess
info:eu-repo/semantics/openAccess

If the material is licensed under a Creative Commons license then you should provide links to applicable Creative Commons licenses, e.g.:

http://creativecommons.org/licenses/zero/1.0/
http://creativecommons.org/licenses/by/3.0/

<table>
<thead>
<tr>
<th>17</th>
<th>Description</th>
<th>MA</th>
<th>Mandatory when applicable property in OpenAIRE instead of optional in DataCite.</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.1</td>
<td>descriptionType</td>
<td>MA</td>
<td></td>
</tr>
</tbody>
</table>

The OpenAIRE Guidelines for Data Archive Managers are built on the DataCite Metadata Schema\(^3\). The following properties are described to make full integration to the OpenAIRE information space and allow OpenAIRE to make links between publications and datasets; datasets and funding information.

---

\(^3\) DataCite Metadata Schema version 2.2: [http://schema.datacite.org/meta/kernel-2.2/index.html](http://schema.datacite.org/meta/kernel-2.2/index.html)
Access rights information

OpenAIRE uses the access rights to enable a better user experience by declaring the access rights clear and explicit in the portal. Access rights are specified using the "Rights" property. Please see encoding scheme in the table above.

An example:

```
<rights>info:eu-repo/semantics/openAccess</rights>
```

Funding information

One of OpenAIRE's main goals is to link research output to (EC) research funding. The following application of the contributor property allows unique and persistent identification of the funder who has funded wholly or partly the dataset described.

The following properties are mandatory when applicable to provide funding information: 7. Contributor; 7.1 contributorType; 7.2 contributorName; 7.3 nameIdentifier; 7.3.1 nameIdentifierScheme:

An example for linking a research output to OpenAIREplus FP7 project:

```
<contributor contributorType="Funder"
     <contributorName>
      European Commission
    </contributorName>
    <nameIdentifier nameIdentifierScheme="info">
     info:eu-repo/grantAgreement/EC/FP7/282896
     </nameIdentifier>
</contributor>
```

Related publications and datasets information

For OpenAIRE datasets are only relevant when they are related to a publication, therefore it is mandatory when applicable for dataset records that OpenAIRE shall harvest to have information about the related publication. DataCite Metadata Schema allows the description of the resource being registered (A) typically a dataset but not limited to that and the related resource (B) in the case of OpenAIRE typically a publication.

Related publications/datasets must have the following properties: 12. RelatedIdentifier, 12.1 relatedIdentifierType, 12. relationType

An example:
Embargo date information

For OpenAIRE two main types of dates are relevant. When the data were made available, published or uploaded to a formal database, this is the date the data were *Issued*.

Sometimes data may be embargoed for a period; this information should be managed by the data provider and expressed by exporting an *Available* date to indicate the end of an embargo period and an *Accepted* date to indicate the start of an embargo period.

An example:

```xml
<dates>
  <date dateType="Issued">2011-12-01</date>
</dates>
```

An embargo example:

```xml
<dates>
  <date dateType="Accepted">2011-12-01</date>
  <date dateType="Available">2012-12-01</date>
</dates>
```
Use of OAI-PMH

OpenAIRE uses the OAI-PMH v2.0 protocol for harvesting dataset metadata.

Metadata Format

OpenAIRE expects metadata to be encoded in the DataCite metadata format (metadataPrefix oai_datacite). For information on how to use the individual DataCite elements, please refer to OpenAIRE application as documented in these guidelines and the DataCite Metadata Schema version 2.2.

OpenAIRE OAI Set

For harvesting the records relevant to OpenAIRE, the use of a specific OAI set at the local repository is mandatory. The set must have the following characteristics:

<table>
<thead>
<tr>
<th>setName</th>
<th>setSpec*</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenAIRE_data</td>
<td>openaire_data</td>
</tr>
</tbody>
</table>

*A harvester only uses the setSpec request to perform selective harvesting. The letters must be in small caps.

Set content

The specific content of the openaire_data set is to be determined at the local repository. Publications to be inserted in the OpenAIRE set must conform to the following criteria:

- All the resources that are related to peer-reviewed articles accepted for publishing.

---

Futher instructions

Best practices and cases of implementations will be available on the OpenAIRE support wiki TBA.