OpenAIRE Guidelines 2.0

Guidelines for content providers of the OpenAIRE information space

October 2012
## Contents

- Introduction ................................................................. 3
- Aim ............................................................................. 3
- What's new.................................................................... 3
- Acknowledgements & Contributors ................................ 4
- Editors ........................................................................... 4
- Experts & Reviewers ..................................................... 4
- License .......................................................................... 5
- Version .......................................................................... 5
- Relation to DRIVER Guidelines ........................................ 6
- OpenAIRE Set ............................................................... 7
  - OpenAIRE Set naming.................................................. 7
  - OpenAIRE Set Content definitions................................ 7
- Compatibility of aggregators ........................................... 8
- OpenAIRE use of oai_dc .................................................. 8
  - OpenAIRE specific use of three elements ....................... 9
  - All elements: short description ..................................... 11
Introduction

Aim
The OpenAIRE Guidelines 2.0 will provide orientation for repository managers to define and implement their local data management policies in compliance with the Open Access demands of the European Commission. Furthermore they will comply with the technical requirements of the OpenAIRE infrastructure that is being established to support and monitor the implementation of the FP7 OA pilot.¹

By implementing these Guidelines repository managers are facilitating the authors who deposit their publications in the repository, in complying with the EC Open Access requirements.

For developers of repository platforms the Guidelines provide guidance to add supportive functionalities for authors of EC funded research in future versions.

What’s new
In comparison with previous versions of the Guidelines, this 2.0 version introduces two main changes:

1. Support for aggregators to become OpenAIRE compatible in order to expose their metadata to the OpenAIRE infrastructure.

2. Extended namespace for project identification. The extended namespace is intended to support a generic way of expressing project information, allowing its use not only for EC/FP projects, but ideally for any funder and project (national or international). Although the use of the extended namespace is recommended, repositories and journals that are already OpenAIRE compatible will remain compatible with no additional work, by using only the mandatory parts (which are the same as in the OpenAIRE Guidelines 1.1).

Regarding this second point, to express project information, in future versions of the Guidelines more sustainable solutions will be sought. This may include the introduction of vocabularies for funders and funding programmes, the presentation of this information in a linked data fashion and compatibility with the CERIF standard.

¹ For information on the FP7 OA pilot and OpenAIRE see http://www.openaire.eu.
Acknowledgements & Contributors

Editors

- Marnix van Berchum (SURF, the Netherlands)
- Eloy Rodrigues (Universidade do Minho, Portugal)

Experts & Reviewers

- Ana Alice Baptista (Universidade do Minho, Portugal)
- Magchiel Bijsterbosch (SURF, the Netherlands)
- Sheridan Brown (Key Perspectives, United Kingdom)
- Natasa Bulatovic (Max Planck Digital Library, Germany)
- Mikael Karstensen Elbæk (Danish Technical University (DTU), Denmark)
- Ilaria Fava (CASPUR, Italy)
- Marek Imialek (University of Bielefeld, Germany)
- Samuele Kaplun (CERN, Switzerland)
- Gerald Lawson (NERC, United Kingdom)
- Alicia López Medina (UNED, Spain)
- Paolo Manghi (CNR, Italy)
- Natalia Manola (University of Athens, Greece)
- Julien Rault (Sciences Po, France)
- Stefanie Rühle (University of Goettingen, Germany)
- Jochen Schirrwagen (University of Bielefeld, Germany)
- Birgit Schmidt (University of Goettingen, Germany)
- Georgi Simeonov (Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria)
- Peter Stanchev (Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria)
- Friedrich Summann (University of Bielefeld, Germany)
- Maurice Vanderfeesten (SURF, the Netherlands)
License

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

Version

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>July 2010</td>
<td>Initial document</td>
</tr>
<tr>
<td>1.1</td>
<td>November 2010</td>
<td>Correction of names and references; addition of license and version statement</td>
</tr>
<tr>
<td>2.0</td>
<td>October 2012</td>
<td>Compatibility for aggregators; extended Namespace for Project Identification</td>
</tr>
</tbody>
</table>
Relation to DRIVER Guidelines

The OpenAIRE guidelines are supplementary and built on top of the DRIVER Guidelines. For OpenAIRE compatibility purposes, all the aspects of the DRIVER Guidelines are valid, with the following exceptions:

1. **Textual resources**

   DRIVER Guidelines 2.0\(^2\), pp.17-18, Part A – Textual Resources: 
   *Textual resources are open access, available directly from the repository for any user worldwide without restrictions such as authorisation or payment*

   Not mandatory for OpenAIRE. Resources can be open access, be under embargo or restricted access.

2. **Use of dc:relation and dc:rights**

   DRIVER Guidelines, p.90: 
   *The DC:RELATION field can typically be used for describing relations to other expressions, or versions of the document.*

   For OpenAIRE compliancy purposes the use of dc:relation and dc:rights will be more specific than expressed in the DRIVER Guidelines 2.0. This will be addressed in the description of the new elements **accessRights**, **embargoEndDate** and **projectID** (see pages 9-10 of these Guidelines).

3. **Author names**

   DRIVER Guidelines 2.0, p.57, encoding scheme for the elements ‘Creator’ and ‘Contributor’: 
   *APA bibliographic writing style as in a reference list. Syntax: surname, initials (first name)*[http://en.wikipedia.org/wiki/Apa_style#Reference_list]

   Not mandatory for OpenAIRE. Full first name(s) may be provided if available.

4. **Use of MPEG-21 DIDL**

   DRIVER Guidelines 2.0, pp.37, DIDL document: 
   *The DRIVER community supports the implementation of the metadataPrefix ‘oai_dc’ and the metadataPrefix ‘didl’.*

   Use of metadataPrefix ‘didl’ is not supported by version 2.0 of the OpenAIRE Guidelines.

OpenAIRE Set
For harvesting of records relevant to OpenAIRE, the use of a specific set at the local repository is mandatory.

OpenAIRE Set naming
The OAI-PMH protocol states:
Repositories may organize items into sets. Set organization may be flat, i.e. a simple list, or hierarchical.  

The table below shows the preferred setName and setSpec that can be used to create the OpenAIRE set.

<table>
<thead>
<tr>
<th>setName</th>
<th>setSpec*</th>
</tr>
</thead>
<tbody>
<tr>
<td>The OpenAIRE set</td>
<td>EC_fundedresources set</td>
</tr>
</tbody>
</table>

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

OpenAIRE Set Content definitions
The specific content of the 'ec_fundedresources' set is to be determined at the local repository.

The content to be inserted in the OpenAIRE set must conform to the following:

- All the resources that will be harvested are outcomes from research projects funded by the EC.
- All the resources that will be harvested are peer-reviewed.

3 See: [http://www.openarchives.org/OAI/openarchivesprotocol.html#Set](http://www.openarchives.org/OAI/openarchivesprotocol.html#Set).
Compatibility of aggregators

Besides individual repositories and journals, also national aggregators can be made OpenAIRE compliant. Additional provenance information of the content providers harvested by the aggregator has to be presented to OpenAIRE.

Provenance information has to be provided in the `about` element of the record, according to the OAI-PMH guidelines: [http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm](http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm).

---

Example of provenance information.

### OpenAIRE use of oai_dc

The use of oai_dc is based on the usage by DRIVER, as expressed in the DRIVER Guidelines 2.0, pp. 52-82.

In three cases the use of the DC elements is specific to OpenAIRE, opposed to a more general use in DRIVER:

- **projectID** is needed to connect project information to the publication in the OpenAIRE information space. It equals to the Grant Agreement number as found in all documentation/correspondence between the EC and the researcher/coordinator. Further project information (project period, funding area etc.) will be ingested into the OpenAIRE information space by other means.
- **accessRights** will define the type of access to the publication
- When the value of accessRights is “embargoedAccess” **embargoEnd Date** will define the date of the embargo period.

---

4 The (generic) namespace can be used to express project information from other funders than the EC.
### OpenAIRE specific use of three elements

<table>
<thead>
<tr>
<th>Element name</th>
<th>projectID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DCMI definition</strong></td>
<td>dc:relation</td>
</tr>
<tr>
<td><strong>Usage</strong></td>
<td>Mandatory</td>
</tr>
<tr>
<td><strong>Usage instruction</strong></td>
<td>A vocabulary of projects will be exposed by OpenAIRE through OAI-MPH, and available for all repository managers. Values will include the project name and projectID. The projectID equals the Grant Agreement number, and is defined by the namespace(^5):&lt;br&gt;info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/Jurisdiction/ProjectName/ProjectAcronym/</td>
</tr>
<tr>
<td><strong>Do not confuse with</strong></td>
<td>--</td>
</tr>
</tbody>
</table>

\(^5\) 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)
<table>
<thead>
<tr>
<th>Element name</th>
<th>accessRights</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCMI definition</td>
<td>dc:rights</td>
</tr>
<tr>
<td>Usage</td>
<td>Mandatory</td>
</tr>
<tr>
<td>Do not confuse with</td>
<td>--</td>
</tr>
</tbody>
</table>
| Examples          | <dc:rights>
|                   |   info:eurepo/semantics/openAccess
|                   | </dc:rights>                                     |

<table>
<thead>
<tr>
<th>Element name</th>
<th>embargoEndDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCMI definition</td>
<td>dc:date</td>
</tr>
<tr>
<td>Usage</td>
<td>Recommended</td>
</tr>
</tbody>
</table>
| Usage instruction | Recommended when accessRights = info:eurepo/semantics/embargoedAccess
|                   | The date type is controlled by the name space info:eurepo/date/embargoEnd/, see http://wiki.surffoundation.nl/display/standards/info-eu-repo/#info-eu-repo-DateTypesandvalue. Encoding of this date should be in the form YYYY-MM-DD (conform ISO 8601). |
| Do not confuse with | --                                               |
| Examples          | <dc:date>
|                   |   info:eurepo/date/embargoEnd/2011-05-12
|                   | </dc:date>                                       |
**All elements: short description**

For a detailed explanation of all the DRIVER elements see the DRIVER Guidelines, pp. 58-82.

**Legend**

- extended element compared to DRIVER Guidelines, see above
- **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

<table>
<thead>
<tr>
<th>Basic element</th>
<th>Status</th>
<th>Encoding schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>M</td>
<td>None, free text</td>
</tr>
<tr>
<td>Creator</td>
<td>M</td>
<td>None. Recommended practice is “last name, first name(s)”</td>
</tr>
<tr>
<td>Subject</td>
<td>MA</td>
<td>Choice of keywords and classifications can be free text (preferably in English) and defined by an URI scheme (preferably info:eu-repo/classification).</td>
</tr>
<tr>
<td>Description</td>
<td>MA</td>
<td>None, free text. Recommended practice is to include an abstract in English. “Abstract” is the default interpretation to the value for dc:description</td>
</tr>
<tr>
<td>Publisher</td>
<td>R</td>
<td>None</td>
</tr>
<tr>
<td>Contributor</td>
<td>O</td>
<td>None. Recommended practice is “last name, first name(s)”</td>
</tr>
<tr>
<td>Date</td>
<td>M</td>
<td>Date</td>
</tr>
<tr>
<td>Type</td>
<td>M</td>
<td>Publication type and Version type defined by an URI scheme (preferably info:eu-repo/semantics).</td>
</tr>
<tr>
<td>Format</td>
<td>R</td>
<td>IANA registered list of Internet Media Types (MIME types) [<a href="http://www.iana.org/assignments/media-types/">http://www.iana.org/assignments/media-types/</a>]</td>
</tr>
</tbody>
</table>

⁶ [http://wiki.surffoundation.nl/display/standards/info-eu-repo/#info-eu-repo-DateTypesandvalue](http://wiki.surffoundation.nl/display/standards/info-eu-repo/#info-eu-repo-DateTypesandvalue)
<table>
<thead>
<tr>
<th>Identifier</th>
<th>M</th>
<th>URI scheme, linking to persistent identifier (URN, handle, DOI), full text document or human start page. Identifiers should be unique.(^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>O</td>
<td>Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata [<a href="http://dublincore.org/documents/dc-citation-guidelines/">http://dublincore.org/documents/dc-citation-guidelines/</a>] as in dcterms:bibliographicCitation</td>
</tr>
<tr>
<td>Language</td>
<td>R</td>
<td>ISO 639-3</td>
</tr>
<tr>
<td>Relation</td>
<td>M</td>
<td>Project identification defined by namespace info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/Jurisdiction/ProjectName/ProjectAcronym/. Green fields mandatory; yellow fields recommended.</td>
</tr>
<tr>
<td>Rights</td>
<td>M</td>
<td>accessRights mandatory, from vocabulary &quot;AccessRights&quot;(^8)</td>
</tr>
<tr>
<td>Audience</td>
<td>O</td>
<td>None. “Education level” is the default value for dc:audience.</td>
</tr>
</tbody>
</table>

---

\(^7\) See [http://www.openarchives.org/OAI/openarchivesprotocol.html#UniqueIdentifier](http://www.openarchives.org/OAI/openarchivesprotocol.html#UniqueIdentifier)

\(^8\) [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)