Content acquisition policy

Publications in OpenAIRE

- Open Access content
- EC funded content
- National or other funded content

Data in OpenAIRE

- Dataset is linked with a publication in the OpenAIRE information space
- EC funded content
- National or other funded content
OpenAIRE Guidelines for data sources
Where we started...

• ‘oai_dc’ 15 elements based on DCMES schema
  • As established and ‘lowest common denominator’ OAI format
• Used in numerous interpretations -> interoperability issue
• Requires agreements on syntax (use of OAI-PMH & oai_dc) and semantics (controlled vocabularies)
  • DRIVER Guidelines v2 introduced in 2008
  • info:eu-repo Application Profile for encoding schemes
• Mainly used to describe textual resources
  • Continued in OpenAIRE Guidelines since 2009
In the meantime...

• OpenAIRE scope & requirements have expanded since then
  • By complementary types of data sources
    • For research data and research information
  • Non-textual content
  • Publication and data linking
  • Project contextualization
  • New global “authoritative” identifier based infrastructures
    • For authors/contributors
    • funders / projects
    • research outputs
Guidelines Timeline

- DRIVER: OA (2008)
- OpenAIRE: EC-FP7 (2010, 2012)
- OpenAIRE: EC + OA (2013)

Text Repositories

Text Repos. Aggregators

Text Rep. Aggr. Oai DC

Data-Rep. OAI_DATACITE

CRIS CERIF
Guidelines for Data Providers

1. Literature Repositories (and journal platforms) - Dublin Core (DRIVER)
2. Data Repositories (and archives/data centres) - Datacite
3. CRIS systems - CERIF-XML
Guidelines for Literature based Data Sources

• Continues the DRIVER v2 Guidelines
• Transfer protocol primarily but not limited: OAI-PMH
• Metadata format (so far): oai_dc
• Application Profile (so far): info:eu-repo
• Extended by properties to describe
  • Funder / Project information
  • Embargo Period / Embargo End Date
  • Open Access Status / Access Rights
  • Alternative Identifier
  • Referenced Dataset(s)
  • Referenced Publication(s)
OpenAIRE OAI-Set

- To group metadata relevant for OpenAIRE
- See [https://www.openaire.eu/content-acquisition-policy/content-acquisition-policy/content-acquisition-policy](https://www.openaire.eu/content-acquisition-policy/content-acquisition-policy/content-acquisition-policy)
- Metadata about Open Access Publications
- Metadata about Publications funded in EC-FP7 / H2020
- Metadata about Publications funded by other funders
  - OpenAIRE provides information about supported funding information

<table>
<thead>
<tr>
<th>setSpec*</th>
<th>setName</th>
</tr>
</thead>
<tbody>
<tr>
<td>openaire</td>
<td>OpenAIRE</td>
</tr>
</tbody>
</table>
### Element name | projectID
---|---
**DCMI definition** | dc:relation

**Usage** | Mandatory (if applicable)

**Usage instruction**
A vocabulary of projects is exposed by the OpenAIRE API: [http://api.openaire.eu/#cha_projects_http](http://api.openaire.eu/#cha_projects_http), and available for all repository managers. Values include funder, project name and projectID.

The projectID equals the Grant Agreement number, and is defined by the namespace: `info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectNumber/Jurisdiction/ProjectName/ProjectAcronym/`

**Example:**

```xml
```
<table>
<thead>
<tr>
<th><strong>Element name</strong></th>
<th>accessRights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DCMI definition</strong></td>
<td>dc:rights</td>
</tr>
<tr>
<td><strong>Usage</strong></td>
<td>Mandatory</td>
</tr>
<tr>
<td><strong>Usage instruction</strong></td>
<td>Use values from vocabulary Access Rights at <a href="http://purl.org/eu-repo/semantics/#info-eu-repo-AccessRights">http://purl.org/eu-repo/semantics/#info-eu-repo-AccessRights</a></td>
</tr>
<tr>
<td>- info:eu-repo/semantics/closedAccess</td>
<td></td>
</tr>
<tr>
<td>- info:eu-repo/semantics/embargoedAccess</td>
<td></td>
</tr>
<tr>
<td>- info:eu-repo/semantics/restrictedAccess</td>
<td></td>
</tr>
<tr>
<td>- info:eu-repo/semantics/openAccess</td>
<td></td>
</tr>
</tbody>
</table>

**Examples:**

```xml
<dc:rights>info:eu-repo/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>Mandatory (if applicable)</td>
</tr>
<tr>
<td>Usage instruction</td>
<td>Recommended when accessRights = info:eu-repo/semantics/embargoedAccess. The date type is controlled by the namespace info:eu-repo/date/embargoEnd/, see <a href="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</a>. Encoding of this date should be in the form YYYY-MM-DD (conform ISO 8601).</td>
</tr>
</tbody>
</table>

Examples:

```
<dc:date>info:eu-repo/date/embargoEnd/2011-05-12</dc:date>
```
## Alternative Identifier

<table>
<thead>
<tr>
<th>Element name</th>
<th>Alternative Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCMI definition</td>
<td>dc:relation</td>
</tr>
<tr>
<td>Usage</td>
<td>Recommended</td>
</tr>
<tr>
<td>Usage instruction</td>
<td>List alternative identifiers for this publication that are not the primary identifier (repository splash page), e.g., the DOI of publisher’s version, the PubMed/arXiv ID. The term is defined by info:eu-repo/semantics/altIdentifier info:eu-repo/semantics/altIdentifier/&lt;scheme&gt;/&lt;identifier&gt; where &lt;scheme&gt; must be one of the following: ark,arxiv, doi, hdl, isbn, purl...</td>
</tr>
<tr>
<td>Example</td>
<td>&lt;dc:relation&gt;info:eu-repo/semantics/altIdentifier/doi/10.1234/789.1 &lt;/dc:relation&gt;</td>
</tr>
</tbody>
</table>
## Referenced Dataset

<table>
<thead>
<tr>
<th>Element name</th>
<th>Referenced Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCMI definition</td>
<td>dc:relation</td>
</tr>
<tr>
<td>Usage</td>
<td>Recommended</td>
</tr>
<tr>
<td>Usage instruction</td>
<td>Encodes links to research datasets connected with this publication. The syntax of info:eu-repo/semantics/dataset is: info:eu-repo/semantics/dataset/ &lt;scheme&gt;/&lt;identifier&gt; where &lt;scheme&gt; must be one of the following: ark, arxiv, doi, hdl, isbn, purl...</td>
</tr>
</tbody>
</table>
## DCMI definition

**Usage instruction**

Encode links to publications referenced by this publication. The syntax of info:eu-repo/semantics/reference is: info:eu-repo/semantics/reference/<scheme>/<identifier> where <scheme> must be one of the following: ark, arxiv, doi, hdl, isbn...

### Examples:

```xml
```
Meet H2020 OA Guidelines


<table>
<thead>
<tr>
<th>Property</th>
<th>DC Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU funding acknowledgment</td>
<td>dc:contributor</td>
<td>“controlled” terms : &quot;[European Union (EU) and &quot;Horizon 2020&quot;],[&quot;Euratom&quot; and &quot;Euratom research and training programme 2014-2018&quot;]&quot;</td>
</tr>
<tr>
<td>Peer reviewed</td>
<td>dc:type</td>
<td>info:eu-repo/semantics/publishedVersion</td>
</tr>
<tr>
<td>Embargo period</td>
<td>dc:date</td>
<td>• info:eu-repo/date/embargoEnd/&lt;YYYY-MM-DD&gt;</td>
</tr>
<tr>
<td></td>
<td>dc:rights</td>
<td>• &lt;YYYY-MM-DD&gt; (as publication date)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• info:eu-repo/semantics/embargoedAccess</td>
</tr>
<tr>
<td>Project information</td>
<td>dc:relation</td>
<td>info:eu-repo/grantAgreement/EC/H2020/[ProjectID]/[Jurisdiction]/[ProjectName]/[ProjectAcronym]/</td>
</tr>
<tr>
<td>Persistent identifier</td>
<td>dc:identifier or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dc:relation</td>
<td></td>
</tr>
<tr>
<td>License</td>
<td>dc:rights</td>
<td>URL of license condition</td>
</tr>
<tr>
<td>Persistent IDs for authors and</td>
<td>dc:creator</td>
<td>&lt;Lastname, Firstname; id_orcid 0000-0000-0000-0000&gt;</td>
</tr>
<tr>
<td>contributors</td>
<td>dc:contributor</td>
<td></td>
</tr>
<tr>
<td>Reference to related research</td>
<td>dc:relation</td>
<td>info:eu-repo/semantics/dataset/&lt;scheme&gt;/&lt;id&gt;</td>
</tr>
<tr>
<td>outcome</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Guidelines for Research Data based Data Sources

• Based on DataCite metadata scheme
• Proven standard for heterogeneous data sources and cross-disciplinary archives
• Maintained by a trusted and sustainable organization
• Support of data citation
  • Metadata Format: oai_datacite
  • OAI-Set: openaire_data
• Zenodo as reference implementation for OpenAIRE Data Repository Guidelines
Guidelines for Research Data based Data Sources

• Adaptations for OpenAIRE:
  • Incl. additional **PID schemes** to DOI, e.g. URN, ARC ...
  • Recommends **links to related** publications & datasets
  • Recommends **Contributor** prop.
    to relate dataset to funding information
  • **Date and Description** prop. are mandatory
  • Enforces (info:eu-repo) encoding scheme on **Rights** property
Guidelines for CRIS Managers based on CERIF-XML

- CERIF as de facto standard for CRIS data model
  - Entities for organization, person, project, output, data ..
- CERIF-XML as data exchange standard (in the CRIS world)
- Maintained by euroCRIS as sustainable organization
Relevant CERIF Entities for OpenAIRE
Guidelines for CRIS Managers based on CERIF-XML

• Adaptations for OpenAIRE:
  • Subset of CERIF entities relevant for OpenAIRE

• Semantic Interoperability
  • Reuse of CERIF Semantics vocabularies for classification and definition of relationships between entities
  • Consistency with the other OpenAIRE Guidelines
    • E.g. resource types, access rights
Guidelines for CRIS Managers based on CERIF-XML

• **Syntactic Interoperability**
  - Distinct namespace for OpenAIRE CERIF schema
  - Nesting only of multi-lingual attributes, federated identifiers, linked entities

• **System Interoperability:**
  - Transfer Protocol: OAI-PMH
  - OAI-PMH sets for
    - Each type of entity
    - The entire CERIF graph
OpenAIRE Compatibility

Status: Levels and OAI Sets

<table>
<thead>
<tr>
<th>OpenAIRE</th>
<th>OpenAIRE</th>
<th>OpenAIRE</th>
<th>OpenAIRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic</td>
<td>2.0</td>
<td>2.0 +</td>
<td>3.0</td>
</tr>
<tr>
<td>Only Open</td>
<td>EC funded</td>
<td>Open Access</td>
<td>Open Access and/or EC</td>
</tr>
<tr>
<td>Access content</td>
<td>content</td>
<td>and EC funded</td>
<td>funded and/or National/</td>
</tr>
<tr>
<td>via driver oai</td>
<td>via ec_fundedres</td>
<td>content</td>
<td>other funded content</td>
</tr>
<tr>
<td>set</td>
<td>sources oai set</td>
<td>via driver and</td>
<td>via openaire</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ec_fundedresour</td>
<td>oai set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ces oai set</td>
<td></td>
</tr>
</tbody>
</table>

Open Access and/or EC funded and/or National/other funded content
Future Directions

Next steps
Innovative Scholarly Services – Need of Improved Metadata Quality

• Features
  • Disambiguation by identifiers for authors / contributors, funders / project grants, research outputs
  • Explicit links to landing pages and fulltext files
  • Explicit links between publications and other research outputs
  • Detailed bibliographic citation information
  • Explicit and machine readable license information

• Future proof
  • Compatible with linked data principles
  • Extensibility towards Open Science properties

• Backward compatibility and low-barrier upgrade
  • Following DCMI Guide on Publishing Metadata (dcterms)
  • COAR IG Controlled Vocabularies
Innovative Scholarly Services – Need of Network Alignment

• Alignment & collaboration needed between major repository initiatives
  • LA Referencia, SHARE, RIOXX, OpenAIRE, ...
  • COAR-CASRAI OA Interoperability WG
    • Alignment on metadata elements and vocabularies
  • COAR IG Controlled Vocabularies
    • Community driven effort to supersede info:eu-repo application profile
    • Working on an up to date technological standard
    • Editorial Board to ensure expertise and long-term support
Find and discuss at: https://guidelines.openaire.eu
Contact us: guidelines@openaire.eu

July 2015
Pedro Principe, University of Minho
Jochen Schirrwagen, University of Bielefeld