Adapting repositories to OpenAIRE Guidelines 4.0: University of Huelva repository, a case study

OPEN REPOSITORIES
2019 | HAMBURG

José Carlos Morillo-Moreno; University of Huelva, jcarlos.morillo@biblio.uhu.es

Emilio Lorenzo-Gil; Arvo Consultores, elorenzo@arvo.es

Eva Braña-Ferreiro; Arvo Consultores, ebrana@arvo.es

FACTS ABOUT OPENAIRE

• Indexes 26 million publications (25 million on open access) and 1 million examples of research data (870 K datasets)
• Aggregates 15K content providers (journals, repositories, aggregators etc.)
• OpenAIRE Guidelines apply to OA repositories and through implementation of the guidelines, repositories comply with European Commission Open Access requirements. These guidelines, in their version 3, are used by a whole set of harvesters, aggregation infrastructures (national, regional and thematic ones) and discovery tools across Europe and Latin America.
• The approval in late November 2018 of the OpenAIRE Guidelines for Literature Repository Managers 4.0 marked a step further in this evolution of requisites. The incorporation (and exposure) of ORCID identifiers, together with a new set of typology, rights and versioning metadata are the main changes of these guidelines.

REASONS FOR SWITCHING FROM OPENAIRE GUIDELINES 3.0 TO 4.0

• Authors need to link to their research output, easing name disambiguation
• Avoids overloading of oai_dc metadata
• Makes maintenance and mappings of controlled vocabularies easier with the help of identifiers
• Makes identification of resources easier
• Funders need to link funding information with research output
• To align with other scholarly communication infrastructures

OPENAIRE GUIDELINES 4.0 OBJECTIVES

• To build an application profile (new OAI-metadataPrefix) based on established and widely used metadata schemes (Dublin Core and DataCitev4.1) in repositories
• To allow for additional properties when needed. To obtain more granularity and flexibility regarding metadata coverage and extensions
• To support identifier schemes for authors, organizations, funders, projects resources
• To update controlled vocabularies (COAR Controlled Vocabularies) for resource types, subject headings & classifications, access rights, license information

CHALLENGES FOR REPOSITORIES

The need to maintain the compatibility of OAI-PMH in a variety of metadating schemas (OAI_DC, OpenAIRE 3, Drive etc.). The appearance of OpenAIRE Guidelines 4.0 does not imply that compatibility with OpenAIRE Guidelines 3 may be abandoned because the various harvesting infraestructures could be adapted slowly to the new standards.

To include identifiers in the repository: ORCID iDs, Funder iDs, ISNI, etc. for subsequent exposure in the OAI-PMH interface

To restrict over-metadating: multiple vocabularies for the correct exposure of descriptive typology, rights and versioning metadating. The need to use the new COAR vocabularies and keep using DRIVER vocabularies for compatibility with current harvesting engines.

Associating ORCID iDs with the author´s metadata fields

In the implementation made (some other alternatives exist), the DSpace authority framework is used, together with the queries made through the ORCID API and the recovery of names and identifiers etc. (functionality present since 2014 in DSpace v5).

Relevant aspects are the normalization of the author name, the assignment of an internal authority key (an authority iD different from the ORCID iD) and the creation of a new entry in the SOLR authority-core.

Example in the SOLR authority-core

```
{   "id": "f3bff5d0-d97a-47f3-aac6-b70a4ee5f043",
    "field": "dc_contributor_author",     "value": "Alvarez Haro, Delia Cecilia",     "deleted": false,
    "creation_date": "2017-10-31T15:36:33.088Z",     "last_modified_date": "2017-10-31T15:36:33.088Z",
    "authority_type": "orcid",     "first_name": "Delia Cecilia",     "last_name": "Alvarez Haro",
    "orcid_id": "0000-0003-4565-2072"    }
```

Collect Funder information

Information about financial support (funding) must be included in resources metadata. The fundingReference element and subproperties replace the info:eu-repo/grantAgreement (DRIVER) syntax. The new elements extend the funding information and define it more precisely, so plan for additional metadata fields.

In order to collect European project information, perform the integration with OpenAIRE services (http://api.openaire.eu) to obtain an authoritative list of projects. The retrieved values include the project name and project ID.

For national or regional projects, plan for integration with services from many other R&D agencies and Funding Agencies to obtain the relevant list of project information.

Plan for new descriptions and vocabularies

Resource types, access rights and resource versions are defined with the respective COAR controlled vocabularies

If a new repository is being built, prepare metadating directly in order to incorporate corresponding COAR vocabularies. Equally, if at the time there is no collection of specific object attributes (such as those needed to complete a citation or those that correspond to the version), the capture of these values should be considered.

If your repository is OpenAIRE 3 compatible and it requires a complete transformation to OpenAIRE 4, perhaps the best option would be to embark on a mass metadata conversion project, swapping in the obsolete DRIVER values for the new COAR vocabularies.

If your repository is OpenAire 3 compatible and needs to maintain this compatibility for a time, the preferable option, chosen for the University of Huelva repository, is to delegate as far as possible the transformation of metadata in the OAI crosswalk. The rules of the crosswalk will transform the vocabularies used internally, DRIVER vocabularies, in the corresponding COAR terms. Clearly, this transformation faces the disadvantage that there is no bi-univocal correspondence, neither syntactical nor semantic, between both sets of vocabularies. This semantic loss should be regarded as a lesser evil as it does not subject all its digital resources to re-metadating.

New OAI-PMH schema for OpenAIRE v4

```
xoai.xml

..
<Format id="oai_openaire">
    <Prefix>oai_openaire</Prefix>
    <XSLT>metadataFormats/oaire.xsl</XSLT>
    <Namespace>http://namespace.openaire.eu/schema/oaire/</Namespace>
```

Exposing ORCID iDs

The application profile oaire uses the datacite schema to expose the author metadata, complementing it with the ORCID iD value, thus creating the <datacite: creator> element.

Note that other contributors will be mapped to <datacite:contributor> elements, following the recommendations for this application profile.

In the implementation performed for the University of Huelva repository, whose values come from the SOLR Authority-cache, if additional author properties exist, they are exposed in the <datacite: creator> element, such as institutional affiliation information.

```
<datacite:creator>
  <datacite:creatorName>Schirrwagen,
Jochen</datacite:creatorName>
    <datacite:nameIdentifier
nameIdentifierFor
nameIdentifierScheme="ORCID"
schemeURI="">http://orcid.org/0000-0002-0458-1004
        </datacite:nameIdentifier>
      <datacite:affiliation>Bielefeld
University</datacite:affiliation>
```

Exposing Funding information

In the oaire metadata schema, the project information is exposed through a fundingReference element. This transforms the value or values stored in the resources in the corresponding oaire elements. This may require reassembly of the values stored in the item, coming from a diversity of APIs, in the fundingReference subelements.

```
<oaire:fundingReference>
  <oaire:funderName>Ministerio de Economía, Industria y
Competitividad</dataite:funderName>
   <oaire:funderIdentifier funderIdentifierType="Crossref
Funder ID">  </oaire:funderIdentifier>
   <oaire:fundingStream>Plan Estatal de Investigación
Científica y Técnica y de      Innovación 2013-2016
</oaire:fundingStream>
     <datacite:awardNumber
awardURI="http://www.idi.mineco.gob.es/portal/site/MICIN
N/menuitem">12345
</oaire:awardNumber>
   <oaire:awardTitle>12345/C3-1</oaire:awardTitle>
</oaire:fundingReference>
```

Exposing COAR Controlled Vocabularies elements

If the metadating values of the resource already contain the appropriate values of COAR vocabularies, simply expose them on the OAI interface.

Otherwise, the best option we recommend is to adapt the OAI crosswalk through XSL transformations e.g. the following section is used to transform the "article" term into "journal article" COAR term.

```
<xsl:choose>
<xsl:when
test="doc:metadata/doc:element[@name='dc']/doc:elemen
t[@name='type']/doc:element/doc:field[@name='value']='in
fo:eu-repo/semantics/article'">m
 <oaire:resourceType resourceTypeGeneral="literature"
uri="http://purl.org/coar/resource_type/c_6501">journal
article</oaire:resourceType>  </xsl:when>
```

Example of oaire elements with COAR Vocabularies

```
<oaire:resourceType
uri="http://purl.org/coar/resource_type/c_6501"
resourceTypeGeneral="literature">journal
article</oaire:resourceType>

<datacite:rights
rightsURI="http://purl.org/coar/access_right/c_abf2">open
access</datacite:rights>

<oaire:version
uri="http://purl.org/coar/version/c_be7fb7dd8ff6fe43">NA</
oaire:version>
```

Completing other elements of the oaire profile

You should include some transformations in the OAI Crosswalk in order to achieve openAIRE 4 compatibility. A non-exhaustive list includes new fields, links to content files, dc fields that must be exposed as datacite fields and some others like citation related fields (recommended)

```
<datacite:subject>Neolítico</datacite:subject>

<datacite:date dateType="Issued">1997</datacite:date>

<oaire:licenseCondition
uri="http://creativecommons.org/licenses/by-nc-nd/3.0/es/"
>Atribución-NoComercial-SinDerivadas 3.0
España</oaire:licenseCondition>

<oaire:file  mimeType="application/pdf"
objectType="fulltext">http://rabida.uhu.es/dspace/bitstrea
m/10272/11446/2/fichero.pdf</oaire:file>
```

RESULTS

The Arias Montano repository, the institutional repository of the University of Huelva, took a first adaption step towards OpenAIRE 4 Guidelines in February 2018, three months after the first draft was made public, in the context of a Spanish Foundation for Science and Technology (FECYT) contract.

This repository was adapted to OpenAIRE 4 Guidelines in December 2018, one month after its publication.

A typical adaptation process will include completing the authorship information to achieve author disambiguation through ORCID iDs, the task of capturing new metadata required for the new oaire profile as well as the transformation of ·ocabularies used.

Example OpenAIRE 3

```
<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Nuevos datos sobre los patrones de poblamiento neolítico y calcolítico al aire libre en el piedemonte de las Sierras Subbéticas</dc:title>
<dc:creator>Gavilán Ceballos, Beatriz</dc:creator>
<dc:creator>Vera Rodríguez, Juan Carlos</dc:creator>
<dc:date>1997</dc:date>
<dc:type>info:eu-repo/semantics/article</dc:type>
<dc:identifier>Gavilán Ceballos, B., Vera Rodríguez, J.C.: "Nuevos datos sobre los patrones de poblamiento neolítico y calcolítico al aire libre
en el piedemonte de las Sierras Subbéticas". ANTIQVITAS. Nº 8, págs. 5-22, (1997). ISSN 1133-6609</dc:identifier>
<dc:identifier>1133-6609</dc:identifier>
<dc:identifier>http://hdl.handle.net/10272/11446</dc:identifier>
<dc:language>spa</dc:language>
<dc:rights>http://creativecommons.org/licenses/by-nc-nd/3.0/es/</dc:rights>
<dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
<dc:publisher>Museo Histórico Municipal de Priego de Córdoba</dc:publisher>
</oai_dc:dc>
```

http://rabida.uhu.es/oai/openaire?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:rabida.uhu.es:10272/11446

Example OpenAIRE 4

```
<oaire:resource xsi:schemaLocation="http://namespace.openaire.eu/schema/oaire/ https://www.openaire.eu/schema/repo-lit/4.0/openaire.xsd">
<datacite:titles>
<datacite:title>Nuevos datos sobre los patrones de poblamiento neolítico y calcolítico al aire libre en el piedemonte de las Sierras Subbéticas</datacite:title>
</datacite:titles> <datacite:creators>
<datacite:creator>
<datacite:creatorName>Gavilán Ceballos, Beatriz</datacite:creatorName>
<datacite:nameIdentifier schemeURI="https://orcid.org"   nameIdentifierScheme="ORCID">0000-    0001-7515-1186</datacite:nameIdentifier>
</datacite:creator> <datacite:creator>
<datacite:creatorName>Vera Rodríguez, Juan Carlos</datacite:creatorName>
<datacite:nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">0000-0002-5989-2694</datacite:nameIdentifier>
</datacite:creator></datacite:creators>
<datacite:subjects>
<datacite:subject>Neolítico</datacite:subject>
</datacite:subjects>
<datacite:dates> <datacite:date dateType="Issued">1997</datacite:date></datacite:dates>
<oaire:resourceType uri="http://purl.org/coar/resource_type/c_6501" resourceTypeGeneral="literature">journal article</oaire:resourceType>
<oaire:version uri="http://purl.org/coar/version/c_be7fb7dd8ff6fe43">NA</oaire:version>
<datacite:identifier identifierType="HANDLE">http://hdl.handle.net/10272/11446</datacite:identifier>
<datacite:alternateIdentifiers>
<datacite:alternateIdentifier alternateIdentifierType="ISSN">1133-6609</datacite:alternateIdentifier>
</datacite:alternateIdentifiers>
<dc:language>spa</dc:language>
<datacite:rights rightsURI="http://purl.org/coar/access_right/c_abf2">open access</datacite:rights>
<oaire:licenseCondition uri="http://creativecommons.org/licenses/by-nc-nd/3.0/es/">Atribución-NoComercial-SinDerivadas 3.0 España</oaire:licenseCondition>
<dc:publisher>Museo Histórico Municipal de Priego de Córdoba</dc:publisher>
<oaire:file objectType="fulltext">http://rabida.uhu.es/dspace/bitstream/10272/11446/2/Nuevos_datos_sobre_los_patrones_de_poblamiento.pdf</oaire:file>
</oaire:resource>
```

http://rabida.uhu.es/oai/openaire?verb=GetRecord&metadataPrefix=oai_openaire&identifier=oai:rabida.uhu.es:10272/11446

REFERENCES

COAR Working Group 2: Repository Interoperability , "COAR Roadmap Future Directions for Repository Interoperability," Feb. 2015
Loesch, Mathias, Rodrigues, Eloy, Principe, Pedro, and Schirrwagen, Jochen, "OpenAIRE Guidelines for Literature Repositories version 3.0", Zenodo, Apr. 2013.
Schirrwagen, Jochen and Baglioni, Miriam, "OpenAIRE Guidelines for institutional and thematic repository managers 4.0", Zenodo, Nov. 2018.