OpenAIRE Literature v.4

OpenAIRE Literature v.4 compliance for DSpace 5.x and DSpace 6.x

Issue https://github.com/DSpace/DSpace/issues/7492 and related Pull Requests at: https://github.com/DSpace/DSpace/pull/2859 and https://github.com/DSpace/DSpace/pull/2860

Full work

The original Pull Requests keep tracks the full work and discussion, you can found at:

- https://github.com/DSpace/DSpace/pull/2703
- https://github.com/DSpace/DSpace/pull/2709

Note that the decision take by developer and committers is to going on a modular approach and the configuration of the specific OpenAIRE "Profile" is demanded to the single institution.

4Science Team released two out-of-the-box branch that contains all code and configuration:

The complete code for version 5 is available here: https://github.com/4Science/DSpace/tree/openaire-lit-v4_dspace5

The complete code for version 6 is available here: https://github.com/4Science/DSpace/tree/openaire-lit-v4_dspace6

Summary

- OpenAIRE 4 Profile
- How to expose different metadata under dc.contributor.*
  - Step by step
- How to enable and map controlled vocabulary
- How to configure custom metadata / How to add a new metadata
- NOTE FOR JSPUI

OpenAIRE 4 Profile

The work of OpenAIRE v4 is a backport of the work done in DSpace 7.

Please note that in DSpace 7, as far as known at the current stage of development, an OpenAIRE oaire schema will be introduced, see: https://github.com/DSpace/DSpace/blob/master/dspace/config/crosswalks/oai/metadataFormats/oai_openaire.xsl and https://github.com/DSpace/DSpace/blob/master/dspace/config/registries/openaire4-types.xml

If you don't have a metadata field yet holding such information and you plan to add it, it could make sense to use the same schema (to use the same approach as in DSpace 7, where in the Entities project a new local “project” schema will be introduced – https://github.com/DSpace/DSpace/blob/main/dspace/config/registries/schema-project-types.xml).

You can found the new schema registry to use with DSpace 5.x and DSpace 6.x at:

- https://github.com/4Science/DSpace/blob/openaire-lit-v4_dspace6/dspace/config/registries/openaire4-types.xml
- https://github.com/4Science/DSpace/blob/openaire-lit-v4_dspace5/dspace/config/registries/openaire4-types.xml
- https://github.com/4Science/DSpace/blob/openaire-lit-v4_dspace6/dspace/config/registries/project-types.xml
- https://github.com/4Science/DSpace/blob/openaire-lit-v4_dspace5/dspace/config/registries/project-types.xml

Use the new schemas (“oaire” and “project”) during installation, then run the registry-loader” script to create it in the database.

Examples (from the [dspace-install]/bin folder):

```
> ./dspace registry-loader -metadata [dspace-install]/config/registries/project-types.xml

> ./dspace registry-loader -metadata [dspace-install]/config/registries/openaire4-types.xml
```

input-forms.xml

Note that if you collect data using submission also the input-forms.xml should be tune to accept the new metadata

To show the new format via OAI-PMH a new transformer should be configured in the OAI configuration.

The full metadata profile could be checked at: https://github.com/DSpace/DSpace/compare/dspace-6_x...4Science:openaire-lit-v4_dspace6 and https://github.com/DSpace/DSpace/compare/dspace-5_x...4Science:openaire-lit-v4_dspace5

How to expose different metadata under dc.contributor.*

To expose more metadata you need to modify the "**oai_openaire.xsl**" in (*[dspace-installation-folder]/config/crosswalks/oai/metadataFormats*).

Below an example of *</datacite:contributors>* managed in the mentioned file:

```
94            <!-- DC CONTRIBUTOR -->
95
96            <datacite:contributors>
97
98                <xsl:for-each
99                    select="doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name!='author' and @name!='editor']/doc:element">
100                   <datacite:contributor>
101                       <xsl:attribute name="contributorType">
102                           <xsl:text>Other</xsl:text>
103                       </xsl:attribute>
104                       <datacite:contributorName>
105                           <xsl:value-of select="./doc:field[@name='value']/text()"/>
106                       </datacite:contributorName>
107                   </datacite:contributor>
108               </xsl:for-each>
109
110               <xsl:for-each
111                   select="doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name='editor']/doc:element/doc:field[@name='value']">
112                   <datacite:contributor>
113                       <xsl:attribute name="contributorType">
114                           <xsl:text>Editor</xsl:text>
115                       </xsl:attribute>
116                       <datacite:contributorName>
117                           <xsl:value-of select="./text()" />
118                       </datacite:contributorName>
119                   </datacite:contributor>
120               </xsl:for-each>
121
122           </datacite:contributors>
```

Starting at line 98 you have a *for-each* statement that will provide select element to iterate the "dc.contributor.*" needed to build the *</datacite:contributors>* element.

Step by step

Starting from the examples above if you want to expose "*dc.contributor.distributor*" add a new *for-each* statement and "*select*" the needed metadata: *doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name='distributor']/doc:field[@name='value']*

Modify also line matching the attribute "**contributorType**" that has to correspond the OpenAIRE Guidelines v4(https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_contributor.html#attribute-contributortype-m); in this case it should be "**Distributor**".

Now you need to exclude this metadata from "*Other*" category. Inside **datacite:contributors** element the first element **datacite:contributor** manages the "*contributorType=Other*". This statement contains a rule to prevent the writing of author or editor into the other category, the statement should be like the following rule: *[@name!='author' and @name!='editor'].* To manage also "Distributor" you have to modify this exception adding your metadata, adding this: *[@name!='author' and @name!='editor' and @name!='distributor'].* The complete example follows:

```
94            <!-- DC CONTRIBUTOR -->
95
96            <datacite:contributors>
97
98                <xsl:for-each
99                    select="doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name!='author' and @name!='editor' and @name!='distributor']/doc:element">
100                   <datacite:contributor>
101                       <xsl:attribute name="contributorType">
102                           <xsl:text>Other</xsl:text>
103                       </xsl:attribute>
104                       <datacite:contributorName>
105                           <xsl:value-of select="./doc:field[@name='value']/text()"/>
106                       </datacite:contributorName>
107                   </datacite:contributor>
108               </xsl:for-each>
109
110               <xsl:for-each
111                   select="doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name='editor']/doc:element/doc:field[@name='value']">
112                   <datacite:contributor>
113                       <xsl:attribute name="contributorType">
114                           <xsl:text>Editor</xsl:text>
115                       </xsl:attribute>
116                       <datacite:contributorName>
117                           <xsl:value-of select="./text()" />
118                       </datacite:contributorName>
119                   </datacite:contributor>
120               </xsl:for-each>
121
122               <xsl:for-each
123                   select="doc:metadata/doc:element[@name='dc']/doc:element[@name='contributor']/doc:element[@name='distributor']/doc:element/doc:field[@name='value']">
124                   <datacite:contributor>
125                       <xsl:attribute name="contributorType">
126                           <xsl:text>Distributor</xsl:text>
127                       </xsl:attribute>
128                       <datacite:contributorName>
129                           <xsl:value-of select="./text()" />
130                       </datacite:contributorName>
131                   </datacite:contributor>
132               </xsl:for-each>
133           </datacite:contributors>
```

How to enable and map controlled vocabulary

To expose custom metadata holding controlled subjects such as DDC you need to enable the controlled vocabulary. First, to configure you need to follow the DSpace documentation (https://wiki.lyrasis.org/display/DSDOC5x/Authority+Control+of+Metadata+Values#AuthorityControlofMetadataValues-HierarchicalTaxonomiesandControlledVocabularies)

The screenshot below is a piece of code in oai_openaire.xsl (*[dspace-installation-folder]/config/crosswalks/oai/metadataFormats*) and it shows the current situation.

```
309     <!-- datacite:subjects -->
310   <!-- https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_subject.html -->
311       <xsl:template
312     match="doc:element[@name='dc']/doc:element[@name='subject']"
313     mode="datacite">
314     <datacite:subjects>
315       <xsl:for-each select="./doc:element">
316         <xsl:apply-templates select="." mode="datacite" />
317       </xsl:for-each>
318     </datacite:subjects>
319   </xsl:template>
320
321   <!-- datacite:subject -->
322   <xsl:template
323     match="doc:element[@name='dc']/doc:element[@name='subject']/doc:element"
324     mode="datacite">
325     <xsl:for-each select="./doc:element/doc:field[@name='value']">
326       <datacite:subject>
327         <xsl:value-of select="./text()" />
328       </datacite:subject>
329     </xsl:for-each>
330   </xsl:template>
```

to expose the controlled vocabulary managed by “dc.subject.ddc” metadata as <datacite:subject> (as Guidelines want) you need to modify this code as follows:

```
309     <!-- datacite:subjects -->
310   <!-- https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_subject.html -->
311       <xsl:template
312     match="doc:element[@name='dc']/doc:element[@name='subject']"
313     mode="datacite">
314     <datacite:subjects>
315       <xsl:for-each select="./doc:element">
316         <xsl:apply-templates select="." mode="datacite" />
317       </xsl:for-each>
318     </datacite:subjects>
319   </xsl:template>
320
321   <!-- datacite:subject -->
322   <xsl:template
323     match="doc:element[@name='dc']/doc:element[@name='subject']/doc:element[@name!='ddc']"
324     mode="datacite">
325     <xsl:for-each select="./doc:element/doc:field[@name='value']">
326       <datacite:subject>
327         <xsl:value-of select="./text()" />
328       </datacite:subject>
329     </xsl:for-each>
330   </xsl:template>
331
332   <!-- dc.subject.ddc -->
333   <xsl:template
334     match="doc:element[@name='dc']/doc:element[@name='subject']/doc:element[@name='ddc']"
335     mode="datacite">
336     <xsl:for-each select="./doc:element/doc:field[@name='value']">
337       <datacite:subject>
338         <xsl:attribute name="subjectScheme">
339           <xsl:text>DDC</xsl:text>
340         </xsl:attribute>
341         <xsl:attribute name="schemeURI">
342           <xsl:text>http://dewey.info/</xsl:text>
343         </xsl:attribute>
344         <xsl:value-of select="./text()" />
345       </datacite:subject>
346     </xsl:for-each>
347   </xsl:template>
```

In the above example we added the relative template (line 333 to 347) to expose the DDC. Note that we need to exclude the dc.subject.ddc from the standard <datacite:subject> mapping (see the difference between line 323 in both screenshot).

How to configure custom metadata / How to add a new metadata

Example: How to add a new metadata “dc.relation.ispartof” on OpenAIRE v.4 field “Citation Title”.

1) Check if the metadata exists in the metadata registry (table “metadatafieldregistry” or in the administrative section “<dspace-host>/dspace-admin /metadata-schema-registry”, see https://wiki.lyrasis.org/display/DSDOC5x/Metadata+and+Bitstream+Format+Registries).

2) To add this field in submission step you need to configure it in the input form file (“<dspace-install>/dspace/config/input-forms.xml”) like this:

```
109       <field>
110         <dc-schema>dc</dc-schema>
111         <dc-element>relation</dc-element>
112         <dc-qualifier>ispartof</dc-qualifier>
113         <repeatable>true</repeatable>
114         <label>Citation Title</label>
115         <input-type>series</input-type>
116         <hint>The title name of the container (e.g. journal, book, conference) this work is published in.
117             This property is considered to be part of the bibliographic citation.</hint>
118         <required>IF IS REQUIRED, WRITE A SENTENCEIF IT IS EMPTY</required>
119       </field>
```

Configure metadata in the input form as you like (please refers the official DSpace documentation https://wiki.lyrasis.org/display/DSPACE /Alter+submission+input+forms).

3) Currently the “citationTitle” element is mapped by the “dc.relation.ispartofseries” (see [dspace-installation]/config/crosswalks/oai/metadataFormats /oai_openaire.xsl). Please use “xsl” rules (https://www.w3.org/TR/xslt-10/) and the OpenAIRE Guidelines to write the correct implementation for the new “dc.relation.ispartof” metadata as you can see in the following example:

```
<!-- CITATION TITLE -->
<xsl:for-each
    select="doc:metadata/doc:element[@name='dc']/doc:element[@name='relation']/doc:element[@name='ispartof']/doc:element/doc:field[@name='value']">
    <oaire:citationTitle>
        <xsl:value-of select="." />
    </oaire:citationTitle>
</xsl:for-each>
```

Note that if you want Citation Title and Citation Volume elements as reported in the documentation:

Citation Title from the dc.relation.ispartofseries with instructions to configure a custom metadata holding the Journal Title if available
Citation Volume from the dc.relation.ispartofseries;

the field ”dc.relation.ispartofseries” into the [dspace-installation]/config/input-forms.xml has to be configured with “series” input-type.

Example: How to change the metadata used for the awards.

To switch from `dc.relation` to something else you need to customize your input-forms.xml and change the oaire crosswalk accordly, this is the line:

```
<xsl:variable name="funderid" select="doc:metadata/doc:element[@name='project']/doc:element[@name='funder']/doc:element[@name='identifier']/doc:element/doc:field[@name='value']"/>
<xsl:variable name="awardnumber" select="doc:metadata/doc:element[@name='oaire']/doc:element[@name='awardNumber']/doc:element/doc:field[@name='value']"/>
<xsl:variable name="awarduri" select="doc:metadata/doc:element[@name='oaire']/doc:element[@name='awardURI']/doc:element/doc:field[@name='value']"/>
<xsl:variable name="awardtitle" select="doc:metadata/doc:element[@name='dc']/doc:element[@name='relation']/doc:element/doc:field[@name='value']"/>

<xsl:variable name="check_fundingreference">
        <xsl:for-each select="$funder">
```

Pay attention to add an additional */doc:element[@name='<YOUR-QUALIFIER>']* in the xpath before the last *doc:element* if you use a qualified metadata.

NOTE FOR JSPUI

Installations with JSPUI, who want to adopt the ORCID patch to better exploit compliance with the OpenAIRE Guidelines, can find the patch available here:

https://wiki.lyrasis.org/display/DSPACE/ORCID+Integration+for+JSPUI