

How to make your bepress Digital Commons IR more compatible with OpenAIRE 3.0

Prepared by Mita Williams (Leddy Library, University of Windsor) for the CARL Open Repositories Working Group's Task Group on OpenAIRE

October 2019

How to make your bepress Digital Commons IR more compatible with OpenAIRE 3.0

CARL Open Repositories Working Group (ORWG) OpenAIRE Task Group (2021-03)

This document is not currently available in French. If you are at a CARL institution and would benefit from French translation, please contact CARL at info@carl-abrc.ca.

Ce document n'est pas actuellement disponible en français. Si vous travaillez dans un établissement membre de l'ABRC et qu'une version française de ce document vous serait utile, veuillez SVP contacter l'ABRC par courriel au info@carl-abrc.ca.

Introduction.....	1
Guidelines.....	1
Examples	2
An example OpenAIRE 3.0 record.....	3
A process towards OpenAIRE 3.0 Compliance	3
The OpenAIRE Compatibility Test	4
Understanding the Validation Results.....	6
Mandatory Fields	6
Recommended Fields	14
Conclusion.....	17

Introduction

Not only does making your IR OpenAIRE 3.0 compatible extends the reach of the work in your IR, access to the OpenAIRE 3.0 Validation Tool allows you to identify and correct problems and inconsistencies in your IR's metadata.

Guidelines

The OpenAIRE Guidelines for Literature Repositories are found here:
<https://guidelines.openaire.eu/en/latest/literature/index.html>

As of June 2019, Digital Commons cannot be made compliant with OpenAire 4.0 because it is based on OAI-PMH and bepress presently can only work with [OAI-DC](#)

The guidelines for OpenAIRE 3.0 can be found here:

https://guidelines.openaire.eu/en/latest/literature/index_guidelines-lit_v3.html

The metadata options for Digital Commons can be found here:

https://www.bepress.com/reference_guide_dc/metadata-options-digital-commons/

Examples

You will need to work with your bepress client rep to make your site compatible with OpenAIRE.

You might want to point to these known examples of OpenAIRE compliant Digital Commons instances:

- Epublications @ RCSI
 - <https://epubs.rcsi.ie/>
 - https://explore.openaire.eu/search/dataprovider?datasourceId=openoair____::9701a1c165dd9420816bfec5edd6c2b1

- Arrow @ TU Dublin
 - <https://arrow.dit.ie/>
 - https://explore.openaire.eu/search/dataprovider?datasourceId=openoair____::39e4973ba3321b80f37d9b55f63ed8b8

- Scholarship @ UWindsor
 - <https://scholar.uwindsor.ca/>
 - https://explore.openaire.eu/search/dataprovider?datasourceId=openoair____::327708dd10d68b1361ad3addbaca01f2

An example OpenAIRE 3.0 record

If you would like to look an example record, please go to:

https://scholar.uwindsor.ca/do/oai/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:scholar.uwindsor.ca:etd-8722 and hit Control-U to View Source

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <OAI-PMH
3   xmlns="http://www.openarchives.org/OAI/2.0/"
4   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
6     http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
7   <responseDate>2019-10-28T19:11:46Z</responseDate>
8   <request verb="GetRecord" identifier="oai:scholar.uwindsor.ca:etd-8722" metadataPrefix="oai_dc" >https://scholar.uwindsor.ca/do/oai/<
9
10   <GetRecord>
11
12   <record>
13     <header>
14       <identifier>oai:scholar.uwindsor.ca:etd-8722</identifier>
15       <timestamp>2019-06-20T01:00:08Z</timestamp>
16       <setSpec>publication:compsci_etd</setSpec>
17       <setSpec>publication:theses-dissertations-major-papers</setSpec>
18       <setSpec>publication:compsci</setSpec>
19       <setSpec>publication:science</setSpec>
20       <setSpec>publication:etd</setSpec>
21     </header>
22     <metadata>
23       <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http:
24         <dc:title>Improving Document Representation Using Retrofitting</dc:title>
25         <dc:creator>Mansoor, Zeeshan</dc:creator>
26         <dc:description>Data-driven learning of document vectors that capture linkage between them is of immense importance in natural language pr
27         <dc:date>2019-01-01T08:00:00Z</dc:date>
28         <dc:type>info:eu-repo/semantics/masterThesis</dc:type>
29         <dc:format>application/pdf</dc:format>
30         <dc:identifier >https://scholar.uwindsor.ca/etd/7721</dc:identifier>
31         <dc:identifier>https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=8722&#38;context=etd</dc:identifier>
32         <dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
33         <dc:rights>http://creativecommons.org/licenses/by-nc-nd/4.0/</dc:rights>
34         <dc:source>Electronic Theses and Dissertations</dc:source>
35         <dc:publisher>Scholarship at UWindsor</dc:publisher>
36
37         <dc:subject>data mining</dc:subject>
38         <dc:subject>document embedding</dc:subject>
39         <dc:subject>multi-view learning</dc:subject>
40         <dc:subject>natural language processing</dc:subject>
41
42       </oai_dc:dc>
43     </metadata>
44   </record>
45
46   </GetRecord>
47 </OAI-PMH>
```

A process towards OpenAIRE 3.0 Compliance

One way going forward is to use the OpenAIRE 3.0 validator to assess your compliance as is, and then use that information to inform your next steps.

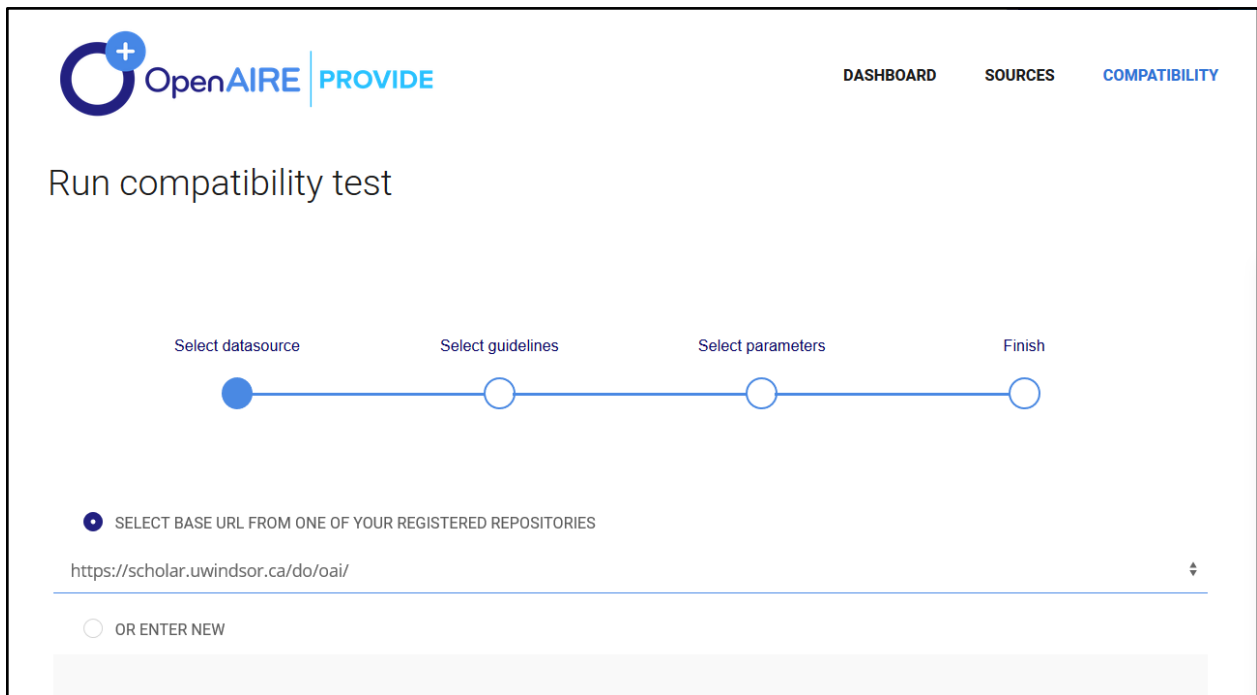
1. The first step is to [register your IR with OpenDOAR](#). The OpenAIRE validator pulls the repository information from this source. There are currently [50+ Digital Commons Institutional Repositories registered with OpenDOAR](#).
 - a. The form will ask you for a OAI-PMH URL. Your answer will end with **do/oai/**
e.g. <http://scholar.uwindsor.ca/do/oai/>
2. Go to the OpenAIRE website at <https://www.openaire.eu/> From the front page, you can travel to [the complete catalogue of Services](#) that OpenAIRE

provides or you can use this link to find [the services that are just related to Content Providers](#), either way, you want to navigate to the...

3. [The OpenAIRE Content Provider Dashboard](#). From this site you can find links to register your collection, validate said collection, enrich your collection with metadata or connections and to establish a service to measure use.
4. Before beginning, you must create an account:
<https://provide.openaire.eu/dashboard>
5. Select Validate from <https://provide.openaire.eu/compatibility/validate> and from the next page, select **Literature Repository**

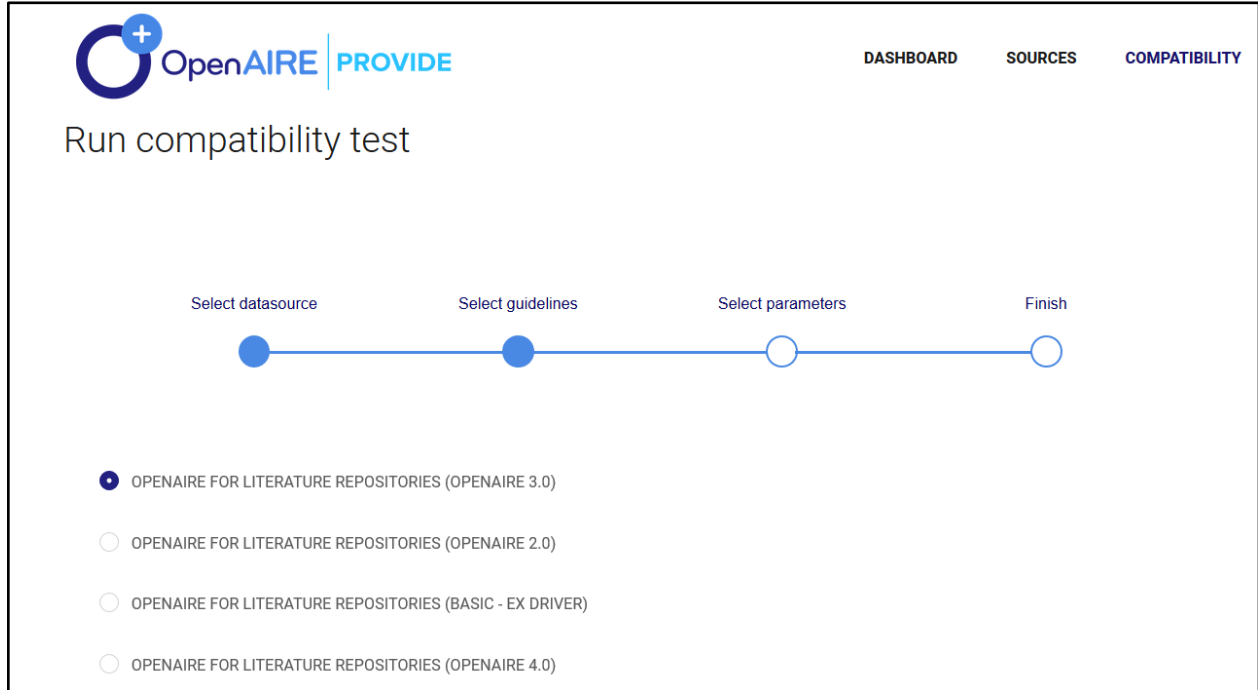
The OpenAIRE Compatibility Test

1. Select your IR



The screenshot shows the 'Run compatibility test' page in the OpenAIRE PROVIDE dashboard. The page features a progress bar with four steps: 'Select datasource' (active), 'Select guidelines', 'Select parameters', and 'Finish'. Below the progress bar, there is a radio button selected for 'SELECT BASE URL FROM ONE OF YOUR REGISTERED REPOSITORIES'. A text input field contains the URL 'https://scholar.uwindsor.ca/do/oai/'. Below this, there is another radio button for 'OR ENTER NEW' and a corresponding text input field.

2. Select OpenAIRE for Literature Repositories (OpenAIRE 3.0)



OpenAIRE PROVIDE

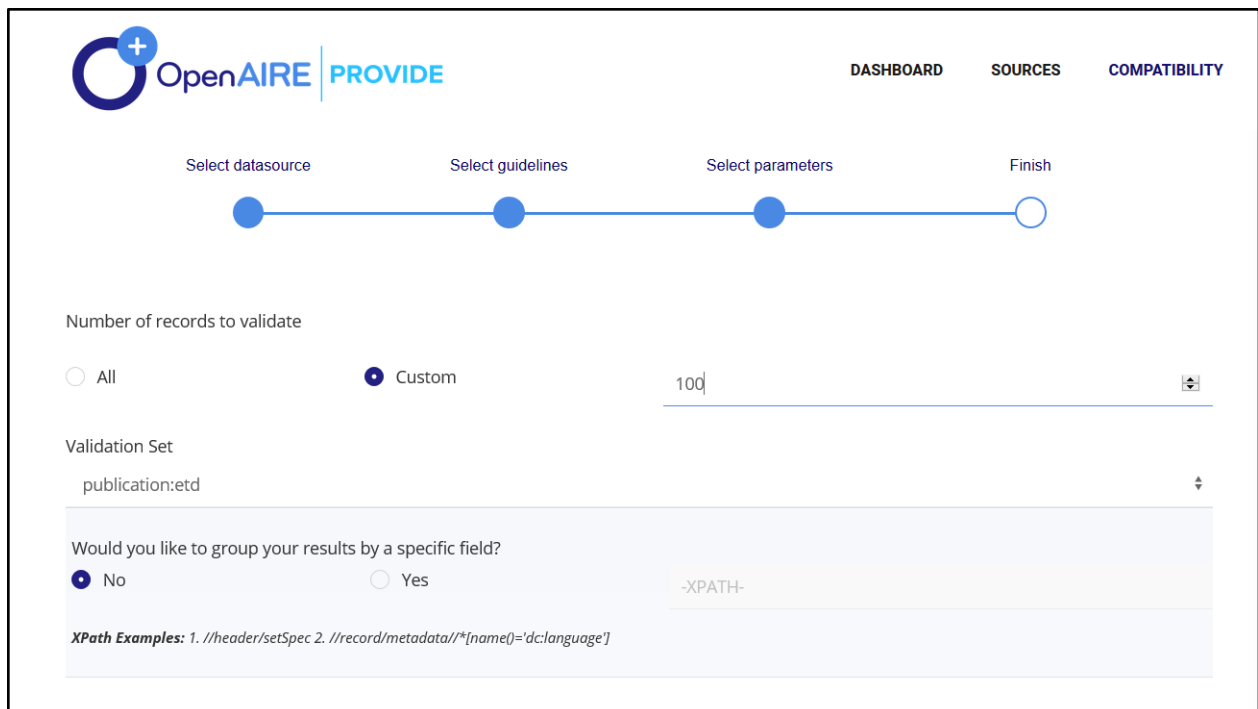
DASHBOARD SOURCES COMPATIBILITY

Run compatibility test

Select datasource Select guidelines Select parameters Finish

- OPENAIRE FOR LITERATURE REPOSITORIES (OPENAIRE 3.0)
- OPENAIRE FOR LITERATURE REPOSITORIES (OPENAIRE 2.0)
- OPENAIRE FOR LITERATURE REPOSITORIES (BASIC - EX DRIVER)
- OPENAIRE FOR LITERATURE REPOSITORIES (OPENAIRE 4.0)

3. From this page you can decide whether you want to test the entire set of records, which validation set you would like to test again, and whether you want to group results with a particular field.



OpenAIRE PROVIDE

DASHBOARD SOURCES COMPATIBILITY

Select datasource Select guidelines Select parameters Finish

Number of records to validate

All Custom 100

Validation Set

publication:etd

Would you like to group your results by a specific field?

No Yes -XPATH-

XPath Examples: 1. //header/setSpec 2. //record/metadata//[name()='dc:language']*

- You will receive a message to wait for the results of your compatibility test.

DASHBOARD SOURCES COMPATIBILITY CONTENT

OpenAIRE PROVIDE

Run compatibility test

MITA WILLIAMS

Select datasource Select guidelines Select parameters Finish

Validation has begun

You will be notified by email with the results of the compatibility test.

- When the test is done, you will receive an email reading something like, *The compatibility test you have submitted has finished. You can retrieve the results by following this url:*
<https://provide.openaire.eu/compatibility/browseHistory/14249>

Understanding the Validation Results

Mandatory Fields

4. Field Access Level (M)

Field Access Level (M)	Use terms from the info:eu-repo-Access-Terms vocabulary. View guideline
------------------------	--

In order to pass this test, your records must have a field dedicated to Access that has one of these options marked for each record:

4. Access Level (M)

4.1. DC Field

`dc:rights`

4.2. Usage

Mandatory

4.3. Usage Instruction

Use terms from the [info:eu-repo-Access-Terms vocabulary](#) . The values are:

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

4.4. Example

1	<code><dc:rights>info:eu-repo/semantics/openAccess</dc:rights></code>
---	---

If you don't have this field, you should ask your bepress representative to create it for you and to populate the fields with the 4 mandatory terms.

Unfortunately, there is no way to dynamically express this field or to associate it with the embargo date field that likely already exists for your content. This means that on a regular basis, you will need to manually change this field on titles that have passed from embargoed status to published status. However, if you run the OpenAIRE validation, it will alert you to mismatches between publication status through the application of **Rule: Field Embargo End Date (MA)**

2. Creator (M)

Field Creator (M)	Use inverted name, so the syntax will be the following: "surname", "initials" ("first name") "prefix". View guideline
-------------------	---

This should pass by default unless you have set up your IR differently.

14. Publication Date (M)

Field Date of Publication (M)	The date should be formatted according to the W3C encoding rules for dates and times. Only dates with year ≥ 1500 are considered as valid. View guideline
-------------------------------	---

This should pass by default unless you have set up your IR differently.

11 Description (Mandatory if Applicable)

Field Description (MA)	This element is used for a textual description of the content. When a resource consists of several separate physical object files, do not use dc:description to list the URLs of these files. View guideline
------------------------	---

This is likely already mapped to your Abstract field. If not, you will have to ask your bepress representative to make this so.

3. Project Identifier (MA)

Field Project Identifier (MA)	info:eu-repo/grantAgreement/Funder/FundingProgram/ /ProjectID /[[Jurisdiction]/[ProjectName] /[ProjectAcronym] Note: If any of the field values contains a forward slash (/), it needs to be escaped using URL encoding (%2F). For instance, My/Project would be represented as My%2FProject. View guideline
-------------------------------	--

Not applicable for theses and as such, not mandatory for this test set.

Work is being done to make adding Funding project information easier for IR managers: <http://www.carl-abrc.ca/news/carl-federal-granting-agencies-openaire-pilot/>

15. Field Publication Type (M)

Field Publication Type (M)	The first occurrence of the DC Element 'type' is mandatory and should be used for the type indication of the scientific output based on the info:eu-repo publication type vocabulary. View guideline
----------------------------	--

You will have to ask your bepress representative to add this field with these mandatory publication types to choose from:

15.4. Usage Instruction

Publication types (controlled):

The first occurrence of the DC Element `type` is mandatory and should be used for the type indication of the scientific output based on the `info:eu-repo` publication type vocabulary:

- `info:eu-repo/semantics/article`
- `info:eu-repo/semantics/bachelorThesis`
- `info:eu-repo/semantics/masterThesis`
- `info:eu-repo/semantics/doctoralThesis`
- `info:eu-repo/semantics/book`
- `info:eu-repo/semantics/bookPart`
- `info:eu-repo/semantics/review`
- `info:eu-repo/semantics/conferenceObject`
- `info:eu-repo/semantics/lecture`
- `info:eu-repo/semantics/workingPaper`
- `info:eu-repo/semantics/preprint`
- `info:eu-repo/semantics/report`
- `info:eu-repo/semantics/annotation`
- `info:eu-repo/semantics/contributionToPeriodical`
- `info:eu-repo/semantics/patent`
- `info:eu-repo/semantics/other`

Please be aware that this set of types may not be enough to express the variety of material that you might have in your IR. The [OpenAIRE 4.0 11. Resource Type](#) expands this list greatly and you might want to borrow from its set of terms, if the OpenAIRE 3.0 list of publication types are insufficient.

Excerpt from OpenAIRE 4.0 11. Resource Type


http://purl.org/coar/resource_type/c_15cd	patent
http://purl.org/coar/resource_type/c_18co	conference poster not in proceedings
http://purl.org/coar/resource_type/c_18cp	conference paper not in proceedings
http://purl.org/coar/resource_type/c_6670	conference poster
http://purl.org/coar/resource_type/c_5794	conference paper
http://purl.org/coar/resource_type/c_c94f	conference object
http://purl.org/coar/resource_type/c_f744	conference proceedings
http://purl.org/coar/resource_type/c_7a1f	bachelor thesis
http://purl.org/coar/resource_type/c_bdcc	master thesis
http://purl.org/coar/resource_type/c_db06	doctoral thesis
http://purl.org/coar/resource_type/c_46ec	thesis
http://purl.org/coar/resource_type/c_0857	letter

12. Publisher (MA)

Field Publisher (MA)	<p>The (commercial or non-commercial) publisher of the resource; not the (sub)institution the author is affiliated with. Publisher is used only in the bibliographic / functional sense, not an organisational one. Use only the full name of the given (commercial) publisher, not the name of an organization or institute that is otherwise [in a broader sense] associated with the creator.</p> <p>View guideline</p>
----------------------	--

This needs to be set up by your bepress consultant.

It's good practice to check the metadata directly. I discovered that even though our form field reads, 'University of Windsor' for our ETDs (as we are the publisher), the expressed metadata was the name of our repository.

Publisher	University of Windsor
Rights	info:eu-repo/semantics/openAccess
Creative Commons License	 <p>This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.</p>

```
<dc:identifier >https://scholar.uwindsor.ca/etd/7721</dc:identifier>
<dc:identifier>https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=8722&#38;amp;context=etd</dc:identifier>
<dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
<dc:rights>http://creativecommons.org/licenses/by-nc-nd/4.0/</dc:rights>
<dc:source>Electronic Theses and Dissertations</dc:source>
<dc:publisher>Scholarship at UWindsor</dc:publisher>
```

18. Field Resource Identifier (M)

Field Resource Identifier (M)	<p>Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system.</p> <p>View guideline</p>
-------------------------------	---

Bepress assigns each document a stable URL which is expressed as an identifier. Additional identifiers such as DOIs can be expressed as well.

```
<dc:identifier >https://scholar.uwindsor.ca/etd/7721</dc:identifier>
```

10. Field Subject (MA)

Field Subject (MA)	In the DC subject element two kinds of values are possible: encode either a keyword or a classification. When both are available use separate occurrences of this element. Use the first occurrence of the DC element 'subject' for a human readable keyword. View guideline
--------------------	---

Even though this is referred to as a 'Field Subject', for us it is mapped to 'keywords'

“Where terms are taken from a standard classification schema: encode each term in a separate element. Encode the complete subject descriptor according to the relevant scheme. Use the capitalisation and punctuation used in the original scheme.”

1. Title (M)

Field Title (M)	Preserve the original wording, order and spelling of the resource title. Only capitalize proper nouns. Punctuation need not reflect the usage of the original. Subtitles should be separated from the title by a colon. This instruction would result in Title:Subtitle (i.e. no space). If necessary, repeat this element for multiple titles. View guideline
-----------------	---

Originally, we kept the title case as expressed by the author of the work but in our work to become more OpenAIRE 3.0 Compliant, we are starting work to change our titles to Sentence Case.

Recommended Fields

7. [Field Alternative Identifier \(R\)](#)

Field Alternative Identifier (R)	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. View guideline
----------------------------------	--

These are best used for supplemental identifiers such as the DOI of the publisher's version of said object.

23. [Audience \(R\)](#)

Field Audience (R)	A class of entity may be determined by the creator or the publisher or by a third party. View guideline	2
--------------------	--	---

Usually non-applicable.

13. [Field Contributor \(R\)](#)

Field Contributor (R)	Examples of contributors are: a supervisor, editor, technician or data collector. The DC element "contributor" describes the scientist(s) that has/have made contributions to the given scientific output, not as a primary creator or (commercial) publisher. View guideline
-----------------------	--

Usually non-applicable.

[17. Format \(R\)](#)

Field Format (R)	Based on best practice, the IANA registered list of Internet Media Types (MIME types) is used to select a term from. View guideline
------------------	--

This is mostly expressed as `<dc:format>application/pdf</dc:format>`

This validation test is useful for identifying metadata that have a PDF missing for reasons unknown.

[20. Language \(R\)](#)

While OpenAIRE 3.0 and 4.0 would prefer `<dc:language>eng</dc:language>` it appears that it will accept `<dc:language>en</dc:language>`

[5. License Condition \(R\)](#)

Field License Condition (R)	It is preferred to refer to a rights service where the reuse rights are made clear to the end-user by using a URL. View guideline
-----------------------------	--

“Using Creative Commons right services, makes the usage rights much more clear to the end user...The URL provides the location where the license can be read. With creative common licenses the type of license can be recognized in the URL name itself. “

`<dc:rights>http://creativecommons.org/licenses/by-nc-nd/4.0/</dc:rights>`

16. Publication Version (R)

Field Publication Version (R)	Version (controlled): be used for the version of the scientific output based on the DRIVER-version info:eu-repo version terms. vocabulary. View guideline
-------------------------------	--

OpenAIRE 3.0 uses these terms

- info:eu-repo/semantics/draft
- info:eu-repo/semantics/submittedVersion
- info:eu-repo/semantics/acceptedVersion
- info:eu-repo/semantics/publishedVersion
- info:eu-repo/semantics/updatedVersion

9. Dataset Reference (R)

Field Referenced Dataset (R)	Encodes links to research datasets connected with this publication. View guideline	2
------------------------------	---	---

8. Field Reference Publication (R)

Field Referenced Publication (R)	Encode links to publications referenced by this publication. View guideline	2
----------------------------------	--	---

19. Field Source (R)

Field Source (R)	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system. View guideline
------------------	---

“Best practice: Use only when the described resource is the result of digitization of non-digital originals. Otherwise, use Relation. Optionally metadata about the current location and call number of the digitized publication can be added”

Conclusion

Working towards OpenAIRE 3.0 is good preparation for OpenAIRE 4.0.

Mita Williams
Scholarly Communications Librarian
Leddy Library, University of Windsor
October 2019