

# Enhancing discovery and enriching the scholarly graph with the Research Outputs Metadata Schema (Rioxx)

George Macgregor<sup>1</sup>, Nicola Dowson<sup>2</sup>, Mick Eadie<sup>3</sup>, Bev Jones<sup>4</sup>,  
Petr Knoth<sup>5</sup> & Paul Walk<sup>6</sup>

**3rd Workshop on Open Citations and Open Scholarly Metadata 2022**  
**05 October 2022**



Blog Universities

# Open access in research: catch up on the debate

The UK recently unveiled its proposal to make all publicly funded research open access. We round up some of the main views on the controversial plans here



# Rioxx background...

- Originally OAI-PMH metadata application profile for UK open repositories
  - <https://www.riox.net/>
- Supporting first major RCUK (now UKRI) Open Access Policy circa 2014 (Jisc support)
- Version 2.0 supported by many UK repositories
- Has demonstrated harvesting & aggregation benefits [3]
- Rioxx as OpenAIRE participation mechanism [2]

Crosswalk from RIOXX 2.0 to OpenAIRE 3.0

developed by the RIOXX team at EDINA in close collaboration with technical personnel at OpenAIRE. We would like to thank, in particular, Jochen Schirrwagen for his...

For the exact representation of properties within the OpenAIRE 3.0 metadata profile, please consult the documentation [OpenAIRE Guidelines: For Literature Repositories](#)

JUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL used in the table below should be interpreted as described in REF.

OpenAIRE	Implementation
dc:rights One or more	This MUST be included, if available.
dc:rights One or more	This MUST be included, if available.
dc:coverage Zero or one	This SHOULD be included, if available.
dc:description zero or more	This MUST be included, if available.
dc:format	This SHOULD be included, if available. If provided in the RIOXX record, values will be in the form of MIME types.



The logo for UKCORR, featuring the text "UKCORR" in a bold, sans-serif font. The letter "O" is highlighted in orange, while the other letters are white. The logo is set against a dark gray rectangular background.

UKCORR

## Rioxx version 3.0

- 2019 schema maintenance & governance transferred to [UKCORR](#)
  - Rioxx Governance Group (RGG)
- Rioxx raison d'être:
  - **update + optimize + 'aggressively open'!**
- Preserve discovery potential -- maintaining evidenced harvesting and aggregation benefits [[3](#)]
- Improving capture & modelling of graph relations to other scholarly entities
- Keeping a lid on complexity...

## Rioxx: The Research Outputs Metadata Schema Version 3.0 Release Candidate 1

Note that this is a **draft** version of the application profile - [please check here for the current version](#)

## Terminology

**Resource** refers to the resource in the repository which identified by the `dc:identifier` property in the Rioxx metadata record.

**Version of record** refers to a version of the resource described in the Rioxx metadata record, which has been made available, electronically, by a publisher.

Terms **MUST**, **MUST NOT**, **REQUIRED**, **SHALL**, **SHALL NOT**, **SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **MAY**, and **OPTIONAL** used in the table below should be interpreted as described in

## Properties in this profile

[dc:license\\_ref](#) | [dc:coverage](#) | [dc:description](#) | [dc:format](#) | [dc:identifier](#) | [dc:language](#) | [dc:publisher](#) | [dc:relation](#) | [dc:source](#) | [dc:subject](#) | [dc:title](#) | [dcterms:dateAccepted](#) | [rioxterms:author](#) | [rioxterms:grant](#) | [rioxterms:project](#) | [rioxterms:publication\\_date](#) | [rioxterms:record\\_public\\_release\\_date](#) | [rioxterms:type](#) | [rioxterms:version](#) | [rioxterms:version\\_of\\_record](#)

## Property details and examples

Property	Cardinality	Description
<code>license_ref</code>	One or more	<p>This is defined in the <a href="#">NISO Open Access Metadata and Indicators</a>. This property <b>MUST</b> take an HTTP(S) URI for its value. This URI <b>MUST</b> point to a license terms specifying how <i>the resource</i> may be used.</p> <p>property content <b>MUST</b> be encoded according to the <a href="#">W3CDTF</a> (a profile of <a href="#">ISO 8601</a>) which typically follows the following format: YYYY-MM-DD</p> <p>This property <b>MUST</b> include the attribute:</p> <ul style="list-style-type: none"> <li><code>start_date</code></li> </ul> <p>This attribute indicates the date upon which this license takes effect. Multiple <code>license_ref</code> properties may be included. Where several are included with the <code>start_date</code> attribute indicating the most recent date takes precedence.</p> <p>Example:</p> <pre>&lt;ali:license_ref start_date="2020-11-17"&gt;   https://creativecommons.org/licenses/by/4.0 &lt;/ali:license_ref&gt;</pre> <p>This approach allows the expression of 'embargoes', where a particular license takes effect at a date in the subjective future.</p> <p>In the absence of any other license, the copyright holder reserves all rights automatically. As a convenience, Rioxx provides a default license in this state:</p> <ul style="list-style-type: none"> <li><a href="https://www.riox.net/licenses/all-rights-reserved">https://www.riox.net/licenses/all-rights-reserved</a></li> <li><a href="https://www.riox.net/licenses/under-embargo-all-rights-reserved">https://www.riox.net/licenses/under-embargo-all-rights-reserved</a></li> </ul>
<code>coverage</code>	Zero or more	<p>Coverage (<code>dc:coverage</code>) will typically include a temporal period (a period label, date, or date range) and a geographical area.</p> <p>In line with the <a href="#">Openaire Guidelines</a>, which recommends the inclusion of this property, <code>dc:coverage</code> is a <b>REQUIRED</b> property.</p>
<code>description</code>	Zero or more	<p>This field may be indexed and its contents presented to people conducting searches. The <code>description</code> property is <b>OPTIONAL</b>.</p>

# Rioxx version 3.0 RC

[Rioxx: The Research Outputs Metadata Schema v3.0 Release Candidate 1](#) published July 2022 [\[1\]](#)

<https://github.com/antleaf/riox>

# Rioxx 3.0 RC – discovery & contributing to the scholarly graph

- Housekeeping, 'PID-ification', and other animals... [More...](#)

Let's consider Rioxx wrt:

- **dc:identifier + dc:relation**
- **rioxxterms:project + rioxxterms:grant**



**dc:identifier**

HTTP(S) URI -- persistent identifier for the resource,  
i.e. DOI, Handle, URN;  
"splash-page" linking to related file(s)

**rioxterms:  
version\_of\_record**

HTTP(S) URI -- a persistent identifier for the  
published 'version of record' (VoR) of the resource

**dc:relation**

HTTP(S) URI(s) – communicating associative  
relationship to other scholarly entities

schema.org – Creative Work vocabulary  
definitions [4]



dc:identifier

rioxterms:  
version\_of\_record

dc:relation

```
<dc:relation type="https://schema.org/ScholarlyArticle"  
  deposit_date="2021-07-06" resource_exposed_date="2021-07-20">  
  https://www.repository.org/article_123456_preprint.pdf  
</dc:relation>  
  
<dc:relation type="https://schema.org/ScholarlyArticle"  
  deposit_date="2021-07-28" resource_exposed_date="2021-07-28">  
  https://www.repository.org/article_1234567.pdf  
</dc:relation>  
  
<dc:relation type="https://schema.org/ScholarlyArticle"  
  deposit_date="2022-03-14" resource_exposed_date="2022-03-14">  
  https://www.repository.org/article_1234567_JATS.xml  
</dc:relation>  
  
<dc:relation type="https://schema.org/Dataset"  
  deposit_date="2022-01-13" resource_exposed_date="2022-01-20">  
  https://doi.org/10.5281/zenodo.3538919  
</dc:relation>  
  
<dc:relation type="https://schema.org/SoftwareSourceCode"  
  deposit_date="2022-03-23" resource_exposed_date="2022-04-18">  
  https://github.com/covid19datahub/R  
</dc:relation>
```

Encoding deposit & exposure data

```
<rioxxterms:project>  
  https://handle.net/10378.1/1590366  
</rioxxterms:project>
```

Addressing confused semantics in scholarly graph, esp. grants and projects conflation – an issue in several schemas

Local identifier rendered as a persistent identifier or a RAiD handle

Superior communication and definition of grant entities via attributes and grant PID



rioxxterms:  
project

rioxxterms:  
grant

```
<rioxxterms:grant  
  funder_name="Wellcome Trust"  
  funder_id="https://isni.org/isni/0000000404277672">  
  https://doi.org/10.35802/218671  
</rioxxterms:grant>
```

```
<rioxxterms:grant  
  funder_name="Arts and Humanities Research Council"  
  funder_id="https://ror.org/0505m1554">  
  AH/W007622/1  
</rioxxterms:grant>
```



# Rioxx validation...

## Metadata quality

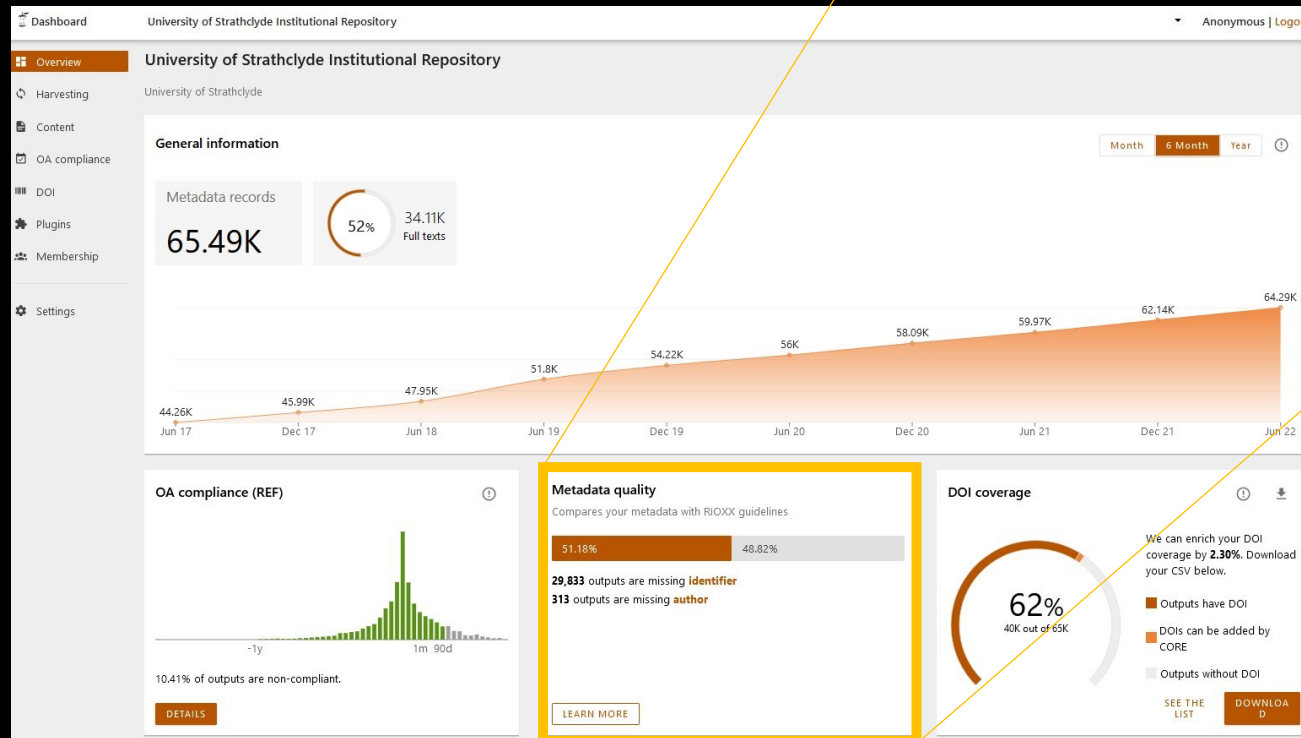
Compares your metadata with RIOXX guidelines

51.18%

48.82%

**29,833** outputs are missing **identifier**

**313** outputs are missing **author**



- Towards harvesting & aggregation using Rioxx 3.0
  - Greater harvesting efficacy
  - Greater contribution to content discovery
  - Contributing meaningful relational depth to the scholarly graph
- Towards Rioxx v 3.0 validation via [CORE Repository Dashboard \[5\]](#)
- Application within new [Plan S flavoured UKRI Open Access Policy...?](#)
- OpenAIRE mappings revisited

# Further information

- Rioxx: The Research Outputs Metadata Schema Version 3.0 Release Candidate 1: <https://www.rioxx.net/profiles/v3-0-rc-1/>
- Rioxx GitHub repo: <https://github.com/antleaf/rioxx>
  - Please comment, raise issues, help us formalise the release!

# References

- [1] Rioxx: The Research Outputs Metadata Schema Version 3.0 Release Candidate 1. (2022). Available: <https://www.rioxx.net/profiles/v3-0-rc-1/>
- [2] Rioxx. (2022). Crosswalk from RIOXX 2.0 to OpenAIRE 3.0. Available: [https://www.rioxx.net/mappings/crosswalk\\_rioxx\\_2\\_0\\_openaire\\_3\\_0/](https://www.rioxx.net/mappings/crosswalk_rioxx_2_0_openaire_3_0/)
- [3] Knoth, P. & Notay, B. (2021) UKRI OA Policy requirements for repositories and how to meet them. *Jisc Workshop*. Available: <https://www.slideshare.net/petrknoth/ukri-oa-policy-requirements-for-repositories-and-how-to-meet-them>
- [4] schema.org. (2022). Available: <https://schema.org/>
- [5] CORE. (2022). Available: <https://core.ac.uk/>
- [6] Knoth, P., & Zdrahal, Z. (2012). CORE: three access levels to underpin open access. *D-Lib Magazine*, 18(11/12). Available: <https://oro.open.ac.uk/35755/>

# Author affiliations

<sup>1</sup>Scholarly Publications & Research Data, University of Strathclyde, Glasgow (UK) ([0000-0002-8482-3973](#)) (Chair: [Rioxx Governance Group](#))

<sup>2</sup>Library Services, Open University, Milton Keynes (UK) ([0000-0001-7900-5098](#))

<sup>3</sup>University of Glasgow Library, Glasgow (UK) ([0000-0003-3295-6915](#))

<sup>4</sup>Scholarly Communications, University of Sheffield (UK) ([0000-0002-3900-9374](#))

<sup>5</sup>Knowledge Media Institute (KMi), The Open University, Milton Keynes (UK) ([0000-0003-1161-7359](#))

<sup>6</sup>AntLeaf Ltd (UK) ([0000-0003-1541-5631](#))