

# Rioxx 3: A Modernised Metadata Profile

Paul Walk, Antleaf Ltd, paul@paulwalk.net

George Macgregor, University of Strathclyde, george.macgregor@strath.ac.uk

Petr Knoth, Open University, petr.knoth@open.ac.uk

## Abstract

Rioxx (formerly RIOXX) is a metadata application profile which was originally developed to facilitate reporting to funders in the UK. Since then, over the last 7 years it has proved useful also to aggregator services harvesting metadata records from repositories, and feedback from those services has indicated a number of ways in which Rioxx could be improved.

This presentation will explain how Rioxx has a new governance group which has been working since 2019 to prepare a new version, one which is designed to meet a broader range of use-cases. We will focus on the changes we have made, including:

- far greater use of persistent identifiers (PIDs);
- a greater focus on the Web as the overarching context (i.e. use of HTTP(S) URIs)
- greater support for expressing important "events" in the lifecycle of scholarly publications, in a response to requirements from open-access funders

We will also compare and contrast Rioxx with OpenAIRE.

Finally, we explain the "radically open" approach we have taken to development, involving community feedback at each stage.

Rioxx is no longer a UK-specific profile, and we believe that Rioxx v3 has potential value for repositories in the global context.

## Keywords

Rioxx; metadata; harvesting;

## Audience

This presentation is aimed at those repository managers who are interested in making their repository content more discoverable and interoperable, and at service providers who wish to process metadata from repositories, especially to support discovery and open-access reporting use-cases.

# Proposal

## Background

### Origin

The Rioxx (previously called "RIOXX") metadata application profile was developed for repositories to share metadata about the scholarly resources they contain. It has been deployed as a metadata application profile in approximately 70 institutional repositories in the UK, and has software implementations in DSpace and ePrints and Pure.

### The need for revision

While originally designed to meet the reporting requirements of Research Councils UK (RCUK), Rioxx had also proven to be generally useful as a standard for sharing metadata between repositories and network services such as large-scale metadata aggregators (e.g. Core<sup>1</sup>). Feedback from such service providers had resulted in a short list of proposed changes and improvements to the RIOXX profile.

### New governance - and a name-change

The profile's sources and documentation - including the main website - are maintained by Antleaf<sup>2</sup>. Before developing the profile further, Antleaf had been keen to find a suitable and representative organisation to provide a governance role and to take oversight of any future revisions. In 2019 the United Kingdom Council of Research Repositories (UKCORR) formed a Governance Group to take responsibility for oversight of revisions and development. To mark this change in circumstance, the name of the profile was changed from RIOXX to Rioxx.

## Presentation

In our presentation, after a very brief introduction and history - we intend to cover the following topics in more detail.

### 1. Why have we developed Rioxx v3 - what problems does it solve?

#### ***Greater use of Persistent Identifiers (PIDs)***

In the last few years there has been a significant growth in recognition of the importance of the use of PIDs to persistently and unambiguously identify scholarly resources, authors, funders etc. While earlier incarnations of Rioxx included some provision for the use of PIDs, Rioxx v3 is much more liberal in their use. The presentation will show where Rioxx has been enhanced to make more use of PIDs, and will explain some of the benefits that arise from this, and the types of use-cases which are made possible by this development.

#### ***Supporting the reporting of "lifecycle events" for open access resources***

The original purpose of serving the needs of the (now defunct) RCUK is no longer a concern for repositories. However, the interest from funders in setting and enforcing open-access requirements for papers and other resources published with their funding is growing (see for example PlanS and its recognition that repositories can host and report on open-access publications). Often, open-access policies from funders require information about important dates in the lifecycle of the publications - for example the date that the metadata record for a resource was made available in a repository. The

presentation will explain how Rioxx makes it easy to report on various life-cycle events through the greater use of more semantically precise dates.

### ***Supporting discovery***

Rioxx v3 introduces changes to better support the discovery of resources in repositories. The first of these is better use of controlled vocabularies. Recently, COAR published version 3 of its *Resource Types* controlled vocabulary. This allows more precise characterisation of the types of resources held in repositories, and Rioxx v3 includes the use of this vocabulary.

Another significant change is the use of properties to identify the relationships between related expressions of the same "work", better enabling the works arising from fluid publication lifecycles to be expressed and associated. This is a feature that will become increasingly important as users (both human and software) of repositories will want to discover, for example, alternative manuscripts ('works') related to published papers, as well as related datasets, software, research protocols, etc. Such associative relations, underpinned by recognised PID schemes, also presents opportunities for enhancing the relational potential of the global scholarly graph, especially through aggregation agents <sup>3</sup>.

The presentation will explore some of these enhancements to support discovery in more detail.

## **2. What is its relevance to the world outside the UK?**

Although its origins lie in a UK-specific context, Rioxx v3 has been developed with a global perspective. Unlike earlier versions, nothing in Rioxx v3 is UK-specific. The changes have mostly been introduced to support some general use-cases, such as facilitating automated discovery of resources described by the metadata records (through explicit linking and semantic typing of resources using schema.org). This is essential for making repositories more machine-friendly and thus supporting basic use cases of harvesters, aggregators and other down-stream services, and reporting on events and characteristics relating to the open-access status of resources, primarily of interest to funders. This may include inspection of associations between resources and their related or underlying research entities, such as grants and/or projects, both of which are semantically differentiated within Rioxx v3.

The presentation will detail some of the ways in which Rioxx v3 supports several use-cases.

## **3. How is this different to OpenAIRE? Why have another profile?**

Rioxx differs from OpenAIRE in a few key respects:

1. Rioxx is "lighter" than OpenAIRE, presenting a flatter expression of repository resources. This makes Rioxx metadata simpler to express, validate and consume by software processes;
2. Rioxx demonstrates greater use of URIs and PIDs to express Web resources, so better supports web-centric discovery, aggregation and relational use-cases. Rioxx v3 follows the principle that only information that is authoritatively created in the repository should be modelled in the repository record while the rest should be machine accessible through explicit links to other systems accessible through PIDs included in the metadata record.
3. Rioxx (arguably) satisfies a broader range of use cases, especially in relation to Open Access.

The presentation will go into a little more detail, addressing the question of when Rioxx might be more suitable than OpenAIRE, and when OpenAIRE might be a better choice (e.g. when contributing to OpenAIRE and satisfying EC funding requirements).

## A Few words on "radically open development"

Finally, having covered these three topics in more detail, we will end with a few words about the development process. We called this "radically open development" because, unlike some comparable metadata profile developments, the process used to develop Rioxx has been completely open to participation by any interested party, using open collaborative technologies such as GitHub Issues and blog comments.

## References

- <sup>1</sup> Knoth, Petr and Zdrahal, Zdenek (2012). CORE: three access levels to underpin open access. D-Lib Magazine, 18(11/12), article no. 4. DOI: <https://doi.org/10.1045/november2012-knoth>
- <sup>2</sup> Rioxx: The Research Outputs Metadata Schema Version 3.0. (2022). Available: <https://www.riox.net/profiles/>
- <sup>3</sup> Macgregor, George and Dowson, Nicola and Eadie, Mick and Jones, Bev and Knoth, Petr and Walk, Paul (2022) Enhancing discovery and enriching the scholarly graph with the Research Outputs Metadata Schema (Rioxx). In: The 3rd Workshop on Open Citations and Open Scholarly Metadata 2022 (WOOC2022), 2022-10-05 - 2022-10-05, Online. <https://doi.org/10.17868/strath.00082458>