

## OR2024 Presentation Submission

The 19th International Conference on Open Repositories, June 3-6th 2024, Göteborg, Sweden

### Exploring the concept of 'custodianship' in harvesting repository resources and graphing their relations: Rioxx version 3.0

George Macgregor, University of Glasgow, UK, [george.macgregor@glasgow.ac.uk](mailto:george.macgregor@glasgow.ac.uk);

Petr Knoth, CORE & The Open University, UK, [petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk);

Paul Walk, Antleaf Ltd, UK, [paul@paulwalk.net](mailto:paul@paulwalk.net);

Nicola Dowson, The Open University, UK, [nicola.dowson@open.ac.uk](mailto:nicola.dowson@open.ac.uk);

Michael Eadie, University of Glasgow, UK, [michael.eadie@glasgow.ac.uk](mailto:michael.eadie@glasgow.ac.uk);

Beverley Jones, University of Sheffield, UK, [b.m.jones@sheffield.ac.uk](mailto:b.m.jones@sheffield.ac.uk);

Agustina Martínez-García, University of Cambridge, UK, [am857@cam.ac.uk](mailto:am857@cam.ac.uk);

### Abstract

This submission addresses concepts associated with Rioxx version 3.0, the schema and specification for which was published in late 2023, following feedback gleaned during OR2023. 'Rioxx: The Research Output Schema' proposes a metadata profile to better ensure superior harvesting and ergo aggregation of scholarly content. It also promote greater semantic interoperability, as well as the graphing of essential research output relations. To assist with its metadata modelling, Rioxx version 3.0 introduces the concept of direct and external custodianship. This submission will explore this concept, establish how custodianship is reflected in the Rioxx schema, and demonstrate how such modelling benefits both repositories and external software agents (such as harvesters and aggregators). The submission will also demonstrate how Rioxx can be used to underpin aspects of open research policy monitoring.

### Keywords

*Metadata modelling, repository resource discovery, scholarly graph, repository harvesting and aggregation, persistent identifiers*

### Audience

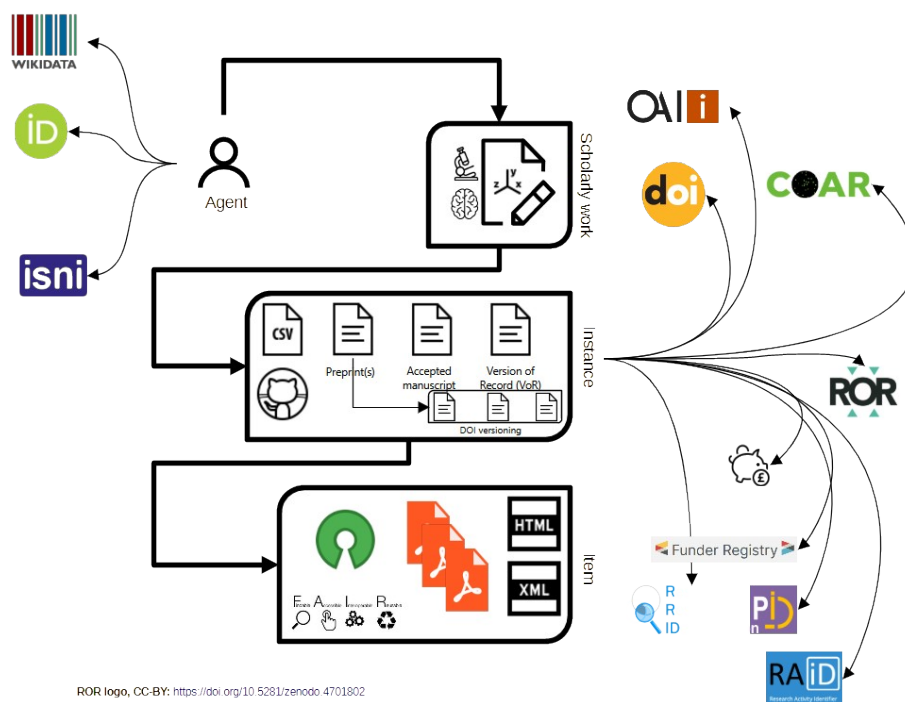
*This submission is aimed at repository managers, and repository software developers, who are interested in making open repository content more discoverable, interoperable, and easier to aggregate. It will also interest those who wish to demonstrate superior metadata modelling of research outputs in repositories and better model relational associations to other scholarly resources on the web.*

### Proposal (no longer than 3 pages)

### Introduction & proposal context

A need to better model and describe scholarly content remains an active issue in open repositories [1]. Issues of semantic interoperability continue to limit the efficacy of repository content discovery, harvesting, and aggregation [2]. Such interoperability limitations inhibit the opportunities for open research, particularly in federated research discovery and in content aggregation for the purposes of text and data mining (TDM) [3]. Many repositories continue to expose metadata inconsistently via OAI-PMH interfaces and with poor or inadequate semantics. Many more demonstrate inconsistent or non-existent encoding of where digital content is located within repositories, resulting in harvesting failures and excessive

computational overheads from aggregation agents [2]. These repository shortcomings have been noted for some time but remain inadequately addressed by the global repository community.



*Figure 1: Illustration of emerging complexity in describing research outputs in repositories; defining semantics, identifying entities, and asserting associative relations.*

Addressing this shortcoming has been complicated by the growing fluidity of research outputs and in publication lifecycles more generally [Figure 1]. A typical research paper may exist as several different expressions (or instances) [4], the evolution of which may be significant to interpreting its contents. Similarly, a research paper increasingly demonstrates relational associations to other scholarly objects, some of which are key to interpreting the research findings contained therein, e.g. related research data, simulation software, research instruments, etc. Such relational associations, often via URI, are necessary to provide contextualization of the research paper, and deliver provenance, transparency, thereby forming a basis for reproducibility [4]. The challenge of better modelling these emerging changes to scholarly publication culture, and addressing the aforementioned metadata interoperability shortcomings of repositories, have converged and have increased the need for repositories to respond accordingly.

## Rioxx version 3: motivation

This submission addresses concepts associated with Rioxx version 3.0<sup>1</sup>, the schema and specification for which was published in late 2023. ‘Rioxx: The Research Output Schema’ is motivated by solving the issues described above and proposes a metadata profile to address repository metadata challenges, promoting superior harvesting and ergo aggregation of scholarly content. It also facilitates greater semantic interoperability, as well as the graphing of essential research output relations. To assist with its metadata modelling, Rioxx version 3.0 introduces the concept of ‘direct’ and ‘external custodianship’.

Originally an OAI-PMH metadata application profile for open repositories in the UK, Rioxx version 2.0 has been widely adopted by repositories in the UK since 2016. Such support has enabled superior discovery potential for repository content, owing to Rioxx’s evidenced harvesting and aggregation benefits [3].

1 Rioxx: The Research Output Schema: <https://rioxx.net/>

Building on feedback gleaned from OR2023 in relation to an earlier ‘candidate release’ [5], Rioxx version 3.0 has undergone significant revision, prior to its recent finalization. Version 3.0 adopts a global repository perspective to its schema and has introduced significant changes to the way in which research outputs are modelled and described. This includes superior capture of graph relations between other scholarly entities, harnessing greater use of persistent identifiers (PIDs) and Typed Links, and reusing semantics from prominent vocabularies while retaining OAI-PMH as the principal data harvesting mechanism [3].

## Direct and external custodianship

Custodianship within Rioxx refers to the concept that the location of a resource is significant to understanding the nature of the thing being described by the Rioxx record [Figure 2]. It is also highly relevant to those machines seeking to process, harvest, or aggregate Rioxx metadata. Resources under ‘direct custodianship’ of a repository are therefore those which are under direct management of a local repository or publication platform. These are resources that the local system controls and maintains. Resources outside this direct custodianship (e.g. resources at or hosted by third party services) demonstrate ‘external custodianship’, and form a wider scholarly graph around the primary resource being described by the Rioxx record.

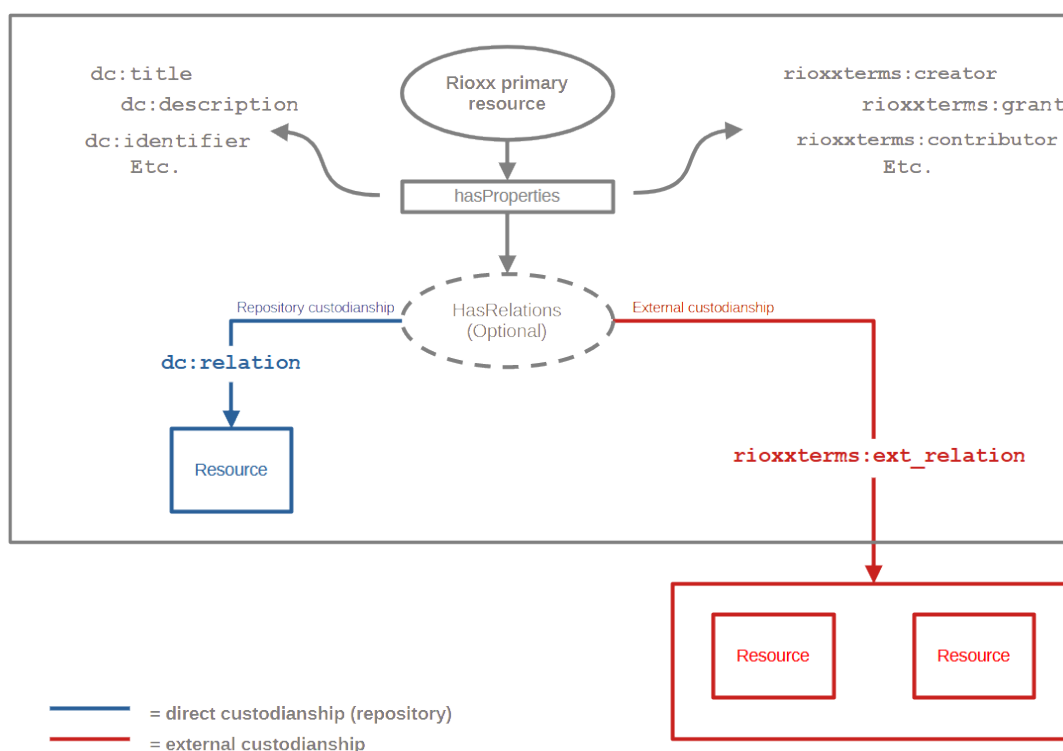


Figure 2: Conceptual diagram of Rioxx version 3.0 model and its approach to custodianship in describing research resources and graphing relations.

Version 3.0 uses <dc:relation> and introduces <rioxterms:ext\_relation>, combined with a range of attributes, to assist repositories in better communicating the location of repository file content, but also graphing relations to other scholarly entities. A series of additional, new properties and sub-properties have also been coined to promote greater specificity in resource description, particularly via URI vocabularies, and interoperability with other standards within the repository community and scholarly publishing more generally.

In addition to communicating the broader Rioxx approach to describing scholarly resources in open repositories (e.g. new properties, attributes, etc.), this submission will explore the underlying Rioxx concept of custodianship in detail. It will establish how it is encoded within the Rioxx schema, and demonstrate how such modelling can benefit both repositories and external software agents (such as harvesters and aggregators) by solving long standing repository discovery impediments and appropriately describing relations to disparate heterogeneous scholarly entities. It will also briefly seek to document the evolution of Rioxx, between OR2023 and OR2024.

The submission will conclude by demonstrating how Rioxx can be used to underpin aspects of open research policy monitoring and reporting, with particular reference to the CORE service<sup>2</sup>, and explain how the Rioxx Governance Group is working with research funders, repository developers, and adjacent initiatives such as Signposting and ResourceSync.

## References (if applicable)

- [1] Arlitsch, Kenning, and Carl Grant. 'Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape'. *Journal of Library Administration* 58, no. 3 (3 April 2018): 264–81. <https://doi.org/10.1080/01930826.2018.1436778>. Available: <https://scholarworks.montana.edu/xmlui/handle/1/14529>
- [2] Knoth, Petr, Matteo Cancellieri, and Martin Klein. 'Comparing the Performance of OAI-PMH with ResourceSync'. *14th International Conference on Open Repositories (OR2023)*. Hamburg, Germany, 2019. <https://www.slideshare.net/martinklein0815/comparing-the-performance-of-oaipmh-with-resourcesync>.
- [3] Pontika, Nancy, Petr Knoth, Matteo Cancellieri, and Samuel Pearce. 'Developing Infrastructure to Support Closer Collaboration of Aggregators with Open Repositories'. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 25, no. 4 (8 March 2016): 172–88. <https://doi.org/10.18352/lq.10138>.
- [4] Macgregor, George. 'Modelling Research Output Expressions : Metadata Schema Modelling of Publication Lifecycles and Scholarly Entities'. In *'Re-Discovery' : Metadata & Discovery Group (MDG) Conference 2023*. GBR, 2023. <https://doi.org/10.17868/strath.00085166>.
- [5] Walk, Paul, George Macgregor, and Petr Knoth. 'Rioxx 3: A Modernised Metadata Profile'. *18th International Conference on Open Repositories (OR2023)*. Stellenbosch: Zenodo, 2023. <https://doi.org/10.5281/zenodo.8091545>.

---

2 CORE: <https://core.ac.uk/>