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

George Macgregor[1], Nicola Dowson[2], Mick Eadie[3], Bev Jones[4], Petr Knoth[5] & Paul Walk[6]

[1]Scholarly Publications & Research Data, University of Strathclyde, Glasgow (UK) (0000-0002-8482-3973)
[2]Learner & Discovery Services, Open University, Milton Keynes (UK) (0000-0001-7900-5098)
[3]University of Glasgow Library, Glasgow (UK) (0000-0003-3295-6915)
[4]Scholarly Communications, University of Sheffield (UK) (0000-0002-3900-9374)
[5]Knowledge Media Institute (KMi), The Open University, Milton Keynes (UK) (0000-0003-1161-7359)
[6]AntLeaf Ltd (UK) (0000-0003-1541-5631)

This contribution considers recent updates to Rioxx, the Research Outputs Metadata Schema [1]. Originally an OAI-PMH metadata application profile for open repositories in the UK, Rioxx version 2.0 has been broadly supported by institutional repositories in the UK since 2016 and has also enabled a level of OpenAIRE participation from institutions for whom compliance was always problematic [2]. More recently Rioxx has entered candidate release status for version 3.0 in which a more holistic and less UK centric approach to the schema has been adopted. This approach has sought to continue the schema's previously evidenced harvesting and aggregation benefits [3] while also improving the capture of graph relations to other scholarly entities.

Rioxx 3.0 introduces superior modelling of associative relationships between scholarly entities, harnesses greater use of persistent identifiers (PIDs), and reuses semantics from schema.org [4] while retaining OAI-PMH as the principal data harvesting mechanism. The schema delivers a useful contribution to burgeoning open scholarly graphs since it can better model relational associations between scholarly entities, enabling research organizations to capture these relations at deposit source and contribute them to the graph via aggregation services, such as CORE [5,6]. Such data once aggregated, and merged with existing PID data, has the potential to greatly enrich the formal PID graph but also enhance user discovery of related research entities, most notably research publications, datasets, software, projects, and grants. Explicit encoding of resource deposit and exposure dates further presents opportunities for better tracking open access growth and monitoring the evolution of this growth at scale.

This contribution will explore examples of features from the Rioxx candidate release, exploring the significance of the changes between version 2.0 and 3.0, and demonstrating some of its relational potential with respect to the scholarly graph. We will also consider the deployment of schema.org vocabularies as a useful, semantically interoperable mechanism for better communicating output type semantics. Future developments, such as the creation of a new Rioxx validator within the CORE repository dashboard, will be briefly summarized.

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/

Macgregor, G., Dowson, N., Eadie, M., Jones, B., Knoth, P., & Walk, P. (2022). Enhancing discovery and enriching the scholarly graph with the Research Outputs Metadata Schema (Rioxx). Paper presented *at The 3rd Workshop on Open Citations and Open Scholarly Metadata 2022*. https://doi.org/10.17868/strath.00082458

[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: http://oro.open.ac.uk/35755/