Modelling research output expressions : metadata schema modelling of publication lifecycles and scholarly entities

George Macgregor

[1]*Scholarly Publications & Research Data, University of Strathclyde, Glasgow G4 ONS*
[2]*Assistant Director - Digital Libraries, University of Glasgow, Glasgow G12 8QE*
https://purl.org/g3om4c
https://orcid.org/0000-0002-8482-3973

Abstract

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 [1]. Emerging from a version 3.0 candidate release, *Rioxx: The Research Output Metadata Schema* has recently been finalized [2]. Version 3.0 adopts a less UK centric approach to the 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 reusing semantics from prominent vocabularies while retaining OAI-PMH as the principal data harvesting mechanism [3]. Most notably, the schema borrows aspects of the Functional Requirements for Bibliographic Records (FRBR) conceptual entity–relationship model to understand relationships between scholarly entities, particularly group 1 entities pertaining to works and expressions [4]. This, combined with better encoding of these relational associations within metadata, means Rioxx can make a useful contribution to the evolution of open scholarly graphs, including the burgeoning 'PID graph'.

This paper explores the modelling of research output expressions within repository metadata schema. Starting with consideration of the Scholarly Works Application Profile (SWAP) [5] and using the recently finalized Rioxx v3.0 schema as case studies, we consider the need for, and the importance of, superior modelling of scholarly works, particularly within open repositories, and consider some of the social-technical impediments to delivering schema which are more adaptable, work agnostic, and PID-centric. We demonstrate how such repository

metadata can greatly enrich the formal PID graph thereby potentially unlocking new areas of scientific study; but also enhance user discovery of related research entities, most notably of research publication expressions, datasets, software, projects, and grants. The paper will also explore areas where FRBR thinking needs reactivating within repository contexts. Despite its general acceptance elsewhere, even with its limitations understood [6], experience within open repositories suggests that a commensurate conceptual shift remains elusive in repository metadata thinking and practice.

References

[1] P. Knoth and B. Notay, 'UKRI OA policy requirements for repositories and how to meet them', presented at the *Jisc Workshop*, 2021. Accessed: Apr. 14, 2023. [Online]. Available: https://www.slideshare.net/petrknoth/ukri-oa-policy-requirements-for-repositories-and-how-to-meet-them

[2] 'Rioxx: The Research Outputs Metadata Schema', 2023. Available: https://www.rioxx.net/

[3] G. Macgregor, N. Dowson, M. Eadie, B. Jones, P. Knoth, and P. Walk, '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)*, Online, Oct. 2022. Available: https://doi.org/10.17868/strath.00082458

[4] IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*. The Hague: International Federation of Library Associations and Institutions (IFLA), 2009. [Online]. Available: https://repository.ifla.org/handle/123456789/811

[5] J. Allinson, 'Describing Scholarly Works with Dublin Core: A Functional Approach', *Libr. Trends*, pp. 221–243, 2008. Available: https://eprints.whiterose.ac.uk/8522/

[6] E. O'Neill and M. Žumer, 'FRBR: Application of the Model to Textual Documents', *Libr. Resources Tech. Serv.*, vol. 62, no. 4, Art. no. 4, Oct. 2018. Available: https://doi.org/10.5860/lrts.62n4.176

Biography

Dr George Macgregor is a repository manager and technologist based at the University of Strathclyde, Glasgow (Scotland, UK). He works on repository and digital library developments, while also supporting institutional activities within open research, research discovery and digital scholarship. During his career George has worked in research and academic roles within universities.