Authority of assertion in repository contributions to the PID graph

George Macgregor[1]
Barbara S. Lancho-Barrantes[2]
Diane Rasmussen Pennington[3]

[1]*Scholarly Publications & Research Data,*
University of Strathclyde, Glasgow G4 0NS
http://orcid.org/0000-0002-8482-3973

[2]*School of Architecture, Technology and Engineering,*
University of Brighton, Brighton BN2 4GJ
https://orcid.org/0000-0001-9994-8886

[3]*Social Informatics Research Group,*
Edinburgh Napier University, Edinburgh EH10 5DT
https://orcid.org/0000-0003-1275-7054

April 19, 2023

Abstract

The principles surrounding Linked Open Data (LOD) and their implementation within digital libraries are well understood. Such LOD implementations may remain challenging, but successes are now well documented and continue to demonstrate the benefits of disseminating and enriching existing metadata with improved semantics and relational associations [1]. Often facilitated in machine-readability enhancements to metadata by harnessing serializations of the Resource Description Framework (RDF) and its reliance of URIs, these LOD approaches have ensured digital libraries – and similar GLAMR initiatives elsewhere – contribute to the growing knowledge graphs associated with the wider semantic web by declaring statements of fact about web entities. Within open scholarly ecosystems a growing use of persistent identifiers (PIDs) to define and link scholarly entities has emerged, e.g. DOIs, ORCIDs, etc.. The requirement for greater URI persistence has been motivated by several developments within the scholarly space; suffice to state that, when combined with appropriate structured

data, PIDs can support improvements to resource discovery, as well as facilitate contributions to the 'PID graph' – a scholarly data graph describing and declaring associative relations between scholarly entities [2].

While the increased adoption of PIDs has the potential to transform scholarship, ensuring that these PIDs are used appropriately, encoded correctly within metadata, and that all relevant relational associations between scholarly entities are declared presents challenges. This is especially true within open scholarly repositories, from where many contributions to the PID graph will be made but – unlike many LOD contexts – from where the authority to assert specific relations may not always exist. Such declarations need to demonstrate reliability and provenance and are central to the interlinking of heterogeneous textual objects, datasets, software, research instruments, equipment, and the related PIDs these items may generate, such as for people, organizations, or other abstract entities.

This paper will explore the issues that arise when levels of authority to assert are lacking or are uncertain, and review results from a related study exploring the 'PID literacy' of scholars [3]. If the PID graph is to demonstrate reliable growth and adequate relational depth, it will be necessary for scholars to interact meaningfully with PID centric systems and to demonstrate a level of 'PID literacy' in their (re)use and creation of PIDs, thereby supporting wider repository metadata initiatives designed to improve research discovery and any relational declarations to the PID graph. In other words, the creation of repository metadata – and scholarly object metadata more generally – is increasingly participatory, requiring scholars to declare relational associations so that repository systems in turn enjoy authority to assert relations. Our work suggests that digital scholarship deficiencies among scholars, including a lack of PID literacy, has the potential to undermine these metadata initiatives, with poor levels of understanding among scholars of why PID referencing is necessary, even for ubiquitous types such as DOIs and ORCIDs [3]. Low levels of scholars' PID literacy is part of wider concern surrounding scholars' open research practices, particularly surrounding transparency and reproducibility [4]. We consider some of the disciplinary differences to be observed between scholars and explore possible solutions.

References

[1] G. Candela, P. Escobar, R. C. Carrasco, and M. Marco-Such, 'Evaluating the quality of linked open data in digital libraries', *J. Inf. Sci.*, vol. 48, no. 1, pp. 21–43, Feb. 2022, available: https://doi.org/10.1177/0165551520930951; https://hdl.handle.net/10045/109459

[2] H. Cousijn et al., 'Connected Research: The Potential of the PID Graph', *Patterns*, vol. 2, no. 1, p. 100180, Jan. 2021, available: https://doi.org/10.1016/j.patter.2020.100180.

[3] G. Macgregor, B. S. Lancho-Barrantes, and D. R. Pennington, 'Measuring the concept of PID literacy: user perceptions and understanding of PIDs in support of open scholarly infrastructure', *Open Inf. Sci.*, vol. 7, no. 1, Jan. 2023, available: https://doi.org/10.1515/opis-2022-0142.

[4] H. M. Sandy et al., 'Making a case for open research: Implications for reproducibility and transparency', *Proc. Assoc. Inf. Sci. Technol.*, vol. 54, no. 1, pp. 583–586, 2017, available: https://doi.org/10.1002/pra2.2017.14505401079.

Biography

Dr George Macgregor (presenting) 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. George has a PhD from the University of Strathclyde, supervised by Prof. Diane M. Rasmussen Pennington.

Dr Barbara S. Lancho-Barrantes is a senior lecturer at the University of Brighton. She has a PhD in Bibliometrics from the University of Extremadura (Spain), developed within SCImago research group, and has led the Bibliometrics Service at the University of Leeds. She has been a postdoctoral researcher in Data Analytics and Scientometrics at the Tecnologico de Monterrey (Mexico). Her research interests mainly focuses on factors influencing research productivity, citation flows and mappings on Sustainable Development Goals (SDGs).

Prof. Diane M. Rasmussen Pennington is Research Lead for the Applied Informatics Subject Group, and Director of the Social Informatics Research Group in the iSchool at Edinburgh Napier University. Her research centres around social informatics, which broadly examines the relationships among individuals, society, information, and technology. Her current areas of focus are ethical metadata, non-textual metadata, new research methods for social media analysis, and online health information provision. Diane earned an MS and a PhD in information science from the School of Library and Information Sciences (now the College of Information) at the University of North Texas.