

The Metadata Team and institutional research visibility

Helen K. R. Williams  0000-0003-1259-7097

Metadata Manager, The London School of Economics and Political Science (LSE) Library

Received: 29 June 2026 | Published: 16 September 2026

ABSTRACT

In recent years metadata teams in many institutions have evolved to support the creation and management of metadata for institutional research outputs as well as for library collections. As organisational priorities around research visibility, impact and open scholarship grow there are new opportunities for metadata teams to contribute to related strategic goals.

This article explores how the Metadata team at LSE Library has extended its traditional metadata role to support LSE's Research for the World strategy. It outlines the creation of new metadata sources to strengthen open, structured metadata representing institutional infrastructure, and examines engagement with Wikimedia platforms to improve the discoverability and accessibility of researchers and their outputs in an increasingly complex digital ecosystem. The article also discusses the development of a combined Wikimedian in Residence and Research Visibility Champion role, highlighting the potential for metadata teams to support open science, and research dissemination and impact through collaboration, innovation and the strategic application of metadata expertise.

KEYWORDS research visibility; Wikimedia platforms; scholarly communication

CONTACT Helen K. R. Williams  h.k.williams@lse.ac.uk  The London School of Economics and Political Science

The Metadata team at LSE is committed to ensuring that its activities support the strategic priorities of the institution. Metadata is positioned as a foundational service which underpins the effective management of institutional knowledge. This alignment is made explicit in a "purpose tree" model which illustrates how metadata contributes to wider organisational goals. By applying metadata expertise beyond traditional library boundaries, and by engaging in collaborative partnerships, the team demonstrates its value in supporting LSE's *Research for the World strategy* ([London School of Economics and Political Science, 2025](#)). The strategy defines the "timeless purpose of a university – to advance the frontiers of knowledge, deepen understanding of the world and share that knowledge for the public good." ([London School of Economics and Political Science, 2025](#), p. 4) It also sets out an intention to improve "the visibility and accessibility of our research to a wide range of audiences"

with the aim of achieving a "global impact" in tackling "the world's most urgent challenges" ([London School of Economics and Political Science, 2025](#), pp. 8, 12 and 14).

Since 2020 the Metadata team has worked with Wikidata to increase the discoverability of the Library's unique and distinctive collections. In 2025 this work was extended to explore how such an approach could be transferrable to supporting the strategic priorities of the university's research strategy, particularly in enhancing research visibility.

Collaborative work with colleagues at Montana State University (MSU) on the use of Wikidata and knowledge graphs to enhance the discoverability of library people, places and services in machine-readable contexts ([Clark, Williams and Rossman, 2022](#)), demonstrates that research impact in a complex, disparate and crowded landscape depends on strong metadata foundations. Institution-level metadata describing organisations, researchers and research outputs needs to be optimised for the Semantic Web and AI tools, supporting discovery across the online environments used by both human audiences and machine agents.

Metadata practice often focuses on the creation of metadata for research outputs themselves, but it is also important for search engines, web-based tools, LLMs, and the AI tools they power, to have a correct understanding of institutional structure, enabling them to interpret the metadata relating to researchers and outputs. This is captured in an LSE Impact blog post in which the authors write:

"Infrastructure ... reflects how we organise trust, authority and participation. In scholarly communication, infrastructure shapes what gets seen, who gets recognised and how research is understood and used. These decisions, often invisible, carry real consequences for equity, access and accountability." ([Chodacki et al., 2025](#))

In support of LSE's *Research for the World strategy*, this work aimed to explore how metadata could:

- increase the discoverability of researchers and their association with LSE,
- ensure research outputs are accurately linked to appropriate organisational units,
- surface previously unrecognised associations between people, institutions and concepts, and
- provide a structured foundation to support the potential for greater reach, impact and re-use of LSE scholarship by both human and machine audiences.

An initial review of Google search results for LSE departments and research centres indicated a strong web-based representation of institutional infrastructure. However, existing metadata sources did not appear to support a comparable understanding in AI tools, such as CoPilot, which were unable to provide a complete or accurate picture, despite experimenting with various prompts. This suggested that these tools lacked

access to the foundational underlying metadata that would appropriately contextualise information about people and research outputs.

In response, structured organisational metadata was collated and added to Wikidata with the aim of improving this situation. Creating and editing Wikidata entities for institutional units, and their alignment with external identifiers, proved more complex than initially anticipated. Existing institutional identifiers were inconsistent and incomplete, underscoring the importance of actively managing organisational data in open, shared infrastructures such as Wikidata, from which other services may subsequently draw.

An audit revealed that 62% of departments and research institutes were either incorrect or missing in the Library of Congress Name Authority File (LCNAF). The Policy, Training, and Cooperative Programs Division of The Library of Congress responded to requests for correction and the creation of new authority records. Similarly, many International Standard Name Identifiers (ISNIs) contained errors, lacked key information, or did not exist. Amendments could be submitted via the ISNI online improvements tool, and new identifiers requested via the ISNI team at the British Library. These identifiers were then reconciled and linked across services to ensure consistency.

A review conducted six months after this work was completed showed that Copilot was subsequently able to provide a correct list of departments and research institutes, including URLs – functionality that had not been observed previously. While a direct causal relationship cannot be demonstrated, this outcome suggests that improvements to shared, structured organisational metadata may positively influence the quality of information returned by LLM-based tools. Subject to team capacity, this approach could be extended to all research centres affiliated with top-level organisational units. In addition, following the example of work at MSU ([Clark, Williams and Rossman, 2022](#)), Google Business Profiles could be claimed and curated for each unit in discussion with the university's communications division.

This foundational metadata work formed the basis for a research optimisation and visibility pilot exploring the role of Wikimedia platforms within these workflows. This was limited to four researchers due to the manual effort required to test the viability of sustained activity in this area. As with all Wikimedia-related work in the Library, a risk-managed approach was adopted. Creating metadata in external systems, beyond institutional control, is a departure from the tighter governance over metadata in LSE's internal systems. This was addressed using a Data Protection Impact Assessment and transparent communication about engagement with Wikimedia. This included articulating the benefits of contributing LSE-related metadata so that it can be enhanced by other editors, linked to external data, and re-used by web-based tools, researchers and machine agents.

Initially the pilot focused on the role of metadata in providing further value for our researchers and their outputs using Wikidata. In practice this involved a structured workflow:

1. Create or edit a Wikidata entity for the individual.
2. Use SPARQL queries to obtain publications on Wikidata already linked to the individual.
3. Use SPARQL queries to obtain a list of publications using the "author name string" property (indicating that links between the author and outputs are not yet established).
4. Use the Author Disambiguator tool to create missing links between the author and their outputs.
5. Export research outputs for the individual from the institutional CRIS and crosscheck with Wikidata.
6. Use the BHL to Wikidata tool to bulk add missing publications with DOIs, linking these to the individual via the Author Disambiguator where necessary.
7. Create a spreadsheet of metadata for outputs without DOIs and upload these to Wikidata in bulk using OpenRefine.

The bibliographic and person data contributed to Wikidata through this workflow connects to related entities already present in the platform. Researchers can then use tools such as Scholia to access visualisations of their outputs and associated data.

Wikidata has also been identified as a tool with "considerable potential...in social sciences in particular" (Arroyo-Machado, W., 2024, p. 4)¹, reinforcing the value of adding metadata for LSE's research outputs to Wikidata to enable connections and support wider discovery. Researchers participating in the pilot were provided with a short slide deck outlining the role of Wikidata, Wikipedia and Wikimedia Commons in research visibility, their relevance to the research process, and the potential benefits for individual researchers, including:

- enables Wikipedia editors to reference research outputs more easily in relevant articles,
- increases the visibility of research metadata to the LLMs underpinning generative AI tools,
- connects information about researchers and their work within a continually expanding network of related entities,
- supports visualisation through tools such as Scholia,
- enhances re-use by external systems, including library catalogues, that draw on Wikidata, and
- improves discoverability, with the potential to increase citations and research impact.

¹ Author's translation.

While these generalised benefits are persuasive and can be demonstrated to researchers through connected metadata and visualisations, measuring their impact remains challenging. Concrete evidence of value is difficult to obtain, particularly when seeking to justify the investment of staff time.

One of LSE's founding principles is that widely shared knowledge can contribute to a better society. In practice, however, research can remain hidden beyond academic audiences, presenting a barrier to the institutional aim of producing research which has an impact on society's greatest challenges. Metadata teams are well placed to address this, playing a pivotal role in improving the discoverability and accessibility of institutional research for non-specialist audiences. Such activity can encourage wider dissemination and understanding, support open science, and potentially increase opportunities for public engagement.

Wikipedia has extensive global reach and is used by both human readers and machines as a data source for other tools, including search engines, voice assistants and AI tools. Consideration was therefore given to whether engagement with Wikipedia could complement the emerging Wikidata activity in the team. This revealed that similar work was already underway elsewhere, indicating growing interest in this area. The UK Reproducibility Network (UKRN) primer on Wikimedia and Open Research advocates for the role of Wikimedia in maximising the reach of research ([Poulter, Sheppard and Mietchen, 2024](#)), and the National Institute of Health Research (NIHR) has reported success in using Wikipedia to disseminate health research ([Harangozó, 2024](#)).

Drawing on this work the research visibility pilot was expanded to include editing Wikipedia subject pages based on insights from the academic outputs of the four pilot researchers. These edits were verified with metadata citations to the relevant research publications. As with the Wikidata work, it was necessary to demonstrate the impact of this to justify the staff time required to translate specialist academic content into material suitable for a lay audience. A range of metrics was explored, including:

- Altmetric scores,
- article metrics, and
- Wikipedia page views.

Each of these showed some modest increases, but the measures were necessarily limited by the small scale of the pilot and the restricted set of research outputs. A larger dataset would be required to demonstrate robust evidence of impact and support firm conclusions.

Nevertheless, the initiative demonstrated clear potential, and there was an opportunity to apply for funding to support its expansion, with the appointment of a Wikimedian in Residence (WiR) identified as the most effective means of scaling activity. Developing a job description for this role highlighted the potential for

confusion in universities when engaging with Wikipedia, particularly where there is an expectation of creating and editing biographical Wikipedia pages for institutional faculty.

Readers familiar with Wikipedia will recognise that this would be incompatible with its principles and at odds with a number of policies, including *Conflict of Interest* ([Wikipedia, 2026a](#)), *Paid editing* ([Wikipedia, 2025](#)), and *Neutral point of view* ([Wikipedia, 2026b](#)), as well as raising potential issues in relation to *Verifiability* ([Wikipedia, 2026d](#)) and *Notability* ([Wikipedia, 2026c](#)). Conflict of Interest (COI) edits are likely to be removed, and editors who persist in such activity may be blocked from contributing. Despite this the COI policy frames such activity as "strongly discouraged" ([Wikipedia, 2026a](#)), which can lead to ambiguity in interpretation.

In this context the intention was not self-promotion, but to support the sharing of knowledge in ways that enhance the global discoverability and accessibility of research and thereby contribute to positive societal impact. While this may appear distinct from the marketing approach of commercial enterprises, the same risks arise, and consequently the same policies and procedures apply. It was important to establish that the focus of the WiR role would be on supporting open science and open knowledge, and enabling institutional subject experts to contribute to this through Wikimedia platforms. Page view data indicates that subject pages generally receive significantly higher levels of engagement than biographical pages. This helped to demonstrate the value of researchers contributing specialist subject expertise to support the wider accessibility of knowledge.

With the provision of funding the Library established a combined post of Wikimedian in Residence and Research Visibility Champion. The role is based in the Metadata team for three days a week focusing on Wikimedia activity and spends two days a week with the Open Research Services team contributing to bibliometrics and wider research visibility work. The post champions the use of tools to improve research visibility and develops, promotes and integrates Wikimedia in supporting open scholarship at LSE and in the wider social science community. Its remit includes advocacy, training, advisory support and workshop delivery, alongside serving as the institutional expert in Wikimedia and wider research visibility to foster knowledge exchange, impact and engagement. The role was developed in close collaboration with Wikimedia UK and continues to be delivered in partnership with them. At the time of writing the project is approximately six months into a two-year residency, during which its impact will be evaluated and key performance indicators reported.

Historically, metadata teams have sometimes been positioned as 'back room' functions, with their contributions to strategic priorities not always fully recognised or communicated. The work outlined here demonstrates how experience and expertise in creating and managing research metadata, combined with alignment to institutional priorities, can extend the function of metadata teams into new areas that enhance institutional research visibility and make a measurable contribution to

organisational priorities. Many years ago, LSE's metadata team was described, by a kind blogger, as "some kind of indispensable repository metadata ninja unit" ([Playforth, 2012](#)), a depiction that continues to inform its work. As the metadata landscape continues to evolve, the strategic contribution of metadata teams within institutions is an area in which metadata managers increasingly seek to demonstrate value. Engagement with Wikimedia represents one way in which metadata teams can remain central and indispensable in new ways within a changing environment.

AI statement

I used Microsoft CoPilot to proof-read my final draft of this article, incorporating some of its suggestions to improve clarity.

References

- Arroyo-Machado, W. (2024) 'Explorando Wikidata para la investigación en Ciencias Sociales', *Infonomy*, 2(2). Available at: <https://doi.org/10.3145/infonomy.24.034> [Accessed: 14 May 2026]
- Chodacki, J., Buttrick, A., Alperin, J. P., Dean, C. and Mentis, D. (2025) 'Infrastructure is a collective responsibility – lessons from scholarly publishing and metadata', *LSE Impact of Social Sciences blog*, 3 June. Available at: <https://blogs.lse.ac.uk/impactofsocialsciences/2025/06/03/infrastructure-is-a-collective-responsibility-lessons-from-scholarly-publishing-and-metadata/> [Accessed: 14 May 2026]
- Clark, J. A., Williams, H. K. R. and Rossman, D. (2022) 'Wikidata and knowledge graphs in practice: using semantic SEO to create discoverable, accessible, machine-readable definitions of the people, places, and services in libraries and archives', *Information Services & Use*, 42(3–4), pp. 377–390. Available at <https://doi.org/10.3233/ISU-220171> [Accessed: 11 August 2026]
- Harangozó, A. (2024) How Wikipedia can help to disseminate research: an innovative NIHR project, *NIHR Blogs*, 5 February. Available at: <https://www.nihr.ac.uk/blog/how-wikipedia-can-help-disseminate-research-innovative-nihr-project> [Accessed 14 May 2026]
- London School of Economics and Political Science (2025) *Research for the World strategy*. The London School of Economics and Political Science. Available at <https://www.lse.ac.uk/Research/Assets/RftW-Strategy-FINAL.pdf> [Accessed: 14 May 2026]
- Playforth, S. (2012) 'The value of cataloguing: CILIP Cataloguing and Indexing Group conference 2012', *The toast in the machine*, 12 September. Available at: <https://playforth.wordpress.com/2012/09/12/the-value-of-cataloguing-cilip-cataloguing-and-indexing-group-conference-2012/> [Accessed 14 May 2026]
- Poulter, M. L., Sheppard, N. and Mietchen, D. (2024) *Wikimedia and open research: a primer from UKRN*. Available at: <https://doi.org/10.31219/osf.io/au48t> [Accessed 14 May 2026]

- Wikipedia (2025) *Paid editing (guideline)*. Available at: https://en.wikipedia.org/wiki/Wikipedia:Paid_editing_guideline [Accessed: 11 August 2026]
- Wikipedia (2026a) *Conflict of interest*. Available at: https://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest [Accessed: 11 August 2026]
- Wikipedia (2026b) *Neutral point of view*. Available at: https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view [Accessed: 11 August 2026]
- Wikipedia (2026c) *Notability*. Available at: <https://en.wikipedia.org/wiki/Wikipedia:Notability> [Accessed: 11 August 2026]
- Wikipedia (2026d) *Verifiability*. Available at: <https://en.wikipedia.org/wiki/Wikipedia:Verifiability> [Accessed: 11 August 2026]