

‘We are here not because we are metadata-breakers; we are here in our efforts to become metadata-makers’

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

Metadata Manager, The London School of Economics and Political Science

Received: 8 August 2025 | Published: 22 September 2025

ABSTRACT

This article outlines a Wikidata project undertaken by LSE Library’s Metadata team to enhance the discoverability of a unique set of oral history interviews about the British Suffrage Movement. The paper outlines each stage of creating linked open data to represent interviews, interviewees and associated entities, and reflects on challenges less familiar to traditional cataloguing practices in RDA and MARC21, including data protection impact assessments, data modelling, URL stability, and creation and enrichment of contextual linked data, including non-LCSH identifiers. The project explores new ways to interrogate and visualise metadata through SPARQL queries, and highlights opportunities to contribute to Wikipedia and Wikimedia Commons. Reflections on conflict of interest, community engagement, and the evolving role of metadata professionals are shared, alongside early insights into project impact. The article demonstrates how Wikidata can extend traditional metadata practice, offering new opportunities for collaboration, discovery, and representation.

KEYWORDS Wikidata; open linked data; discovery

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

Introduction

One hundred and seventeen years on I wonder what Emmeline Pankhurst would make of my appropriation of her famous words ([Pankhurst, 1908](#)) for an article about ‘breaking out’ of MARC metadata silos. As we venture beyond the traditional remit of library metadata teams and embrace new ways to make suffrage voices discoverable, I like to think she would cast her vote of confidence in our direction!

Many *Catalogue & Index* readers will already be familiar with LSE’s Wikidata thesis project¹ which successfully increased the reach and engagement of this content while also developing the skills of the Metadata team in contributing to the broader linked open data ecosystem ([Williams, 2022](#)).

¹ https://www.wikidata.org/wiki/Wikidata:WikiProject_LSEThesisProject

Building on that success, and with the aim of further expanding the team's Wikidata expertise, we explored further opportunities to contribute unique content to Wikimedia platforms and selected a set of oral history interviews about the British Suffrage Movement. The 205 interviews (with 183 individuals) were conducted by the historian Brian Harrison between 1974 and 1981, and were subsequently deposited with The Women's Library². The interviews (with surviving suffrage campaigners and their relatives and employees) also explore broader themes around women's organisations, employment and family life, the birth control movement, politics, peace activism, and trade unionism ([LSE Library, no date](#)).

Wikidata metadata creation and enhancement

Having not previously worked with oral history interviews, in Wikidata or any other context, I initially explored the Wikidata pilot pages of the Program for Cooperative Cataloging ([Wikidata, no date](#)) to find other universities working with similar materials. These examples informed my understanding of appropriate data models to support rich contextual linked data and resulted in the creation of local data models for collection and interview level records, and for entities represented in them.

Following the principle that producing 'something' 'good enough' is more valuable than aiming for perfection and potentially achieving nothing, I outlined a basic iterative and experimental project approach:

- Complete Data Protection Impact Assessment (DPIA) and create Wikidata project page.
- Automate creation of basic QIDs for names not yet represented in Wikidata.
- Enrich new and existing QIDs using local authorities and interview summaries.
- Create QIDs for interviews, using linked data to connect interviewees and named entities.
- Edit relevant Wikipedia pages for individuals/organisations.
- Explore creation of Wikipedia pages for notable individuals currently missing from Wikipedia.

Our project page³ ([Williams, no date](#)) demonstrates that we largely followed this structure, working out the detail as the project progressed, and adding a couple of extra steps as required.

Before beginning we explored data protection considerations and completed a DPIA. Interviews which do not have consent to be made public were excluded from the project. Interviews and related metadata for the rest of the collection are already publicly available, but the DPIA highlights the way in which the project will bring

² <https://www.lse.ac.uk/library/collection-highlights/the-womens-library>

³ https://www.wikidata.org/wiki/Wikidata:WikiProject_The_Women%E2%80%99s_Library_LSESuffrageInterviewsProject#Tasks

metadata together in new ways to increase discoverability in the semantic web environment, and assesses potential risks accordingly.

At this stage we also needed to address the issue of URL stability. The metadata records, with accompanying descriptions for each interview, are hosted on our archives catalogue, but an imminent system migration meant that existing URLs would change. The audio files are accessed via the LSE website but plans to re-digitise the collection meant that these URLs were also subject to change. Consequently, neither option was stable enough to provide reliable reference URLs on the hundreds of Wikidata entities we would create. Via Wikidata help pages⁴ ([Wikidata, no date](https://www.wikidata.org/wiki/Help:Sources)) I explored the option to use a *'stated in'* qualifier instead. This could be linked to the collection level QID⁵ for the project. If URLs changed then only one QID would need updating to deprecate the old URL and provide the current one.

With metadata for the interviews and related authority records extracted from our Archives system I was able to reconcile names with Wikidata via OpenRefine. Of 222 names 98 were already represented in Wikidata, and 124 were not. Information about these individuals was spread across interview metadata, authority records and interview summaries, so manual creation of full records would be time consuming. I prioritised a 'quick win' by focusing on missing names and uploading an easily assembled spreadsheet of basic information to OpenRefine, from where we could write a schema to batch upload to Wikidata. This included statements for label (entity name), description, alias, given name, family name, occupation, gender, dates of birth and death, and identifying that an oral history for the individual was held at LSE, all of which was standard metadata easily extracted from our Archives metadata exports. Had we been creating entities for living people I would not have presumed to assign such personal data as gender, but all names associated with the project were thought to be deceased, and the inclusion of Wikidata's sex or gender property was relevant to the historical context. The newly created QIDs were extracted along with existing ones, and a bulk update to add P5008 properties *'on focus list of Wikimedia project'*: *The_Women's_Library_LSESuffrageInterviewsProject*⁶ gathered them into one identifiable project set.

All names were then ready for manual enhancement, a step considered a worthwhile investment of time given the uniqueness of the content and the collection's importance as a key source for British women's history of the 20th century. The wider team used interview metadata and summaries to add further statements to the Wikidata entities, including country of citizenship, place of birth or death, employer, educated at, notable works, affiliation, spouse, and relatives. All enhancements were recorded on a spreadsheet which was later used to batch load all edits to Wikidata. Metadata for the interviews themselves required manual enhancements in three stages, each designed to create rich contextual linked data:

⁴ <https://www.wikidata.org/wiki/Help:Sources>

⁵ <https://www.wikidata.org/wiki/Q100380678>

⁶ <https://www.wikidata.org/wiki/Q117322976>

- **Significant people mentioned in interviews** (defined as those either already represented in Wikidata or identifiable through at least one verifiable attribute). For the latter, the team populated a 'missing people' spreadsheet with name, Wikidata properties to represent identifying information (such as employer or family relationship), and reference URLs to verify these statements. This was uploaded to Wikidata via OpenRefine, with a schema which also automated the addition of each name in a significant person statement, with the qualifier 'object of statement has role: subject' to the correct interview QID. Identifying significant names for each interview was a manual process that took place in 2023. Newly digitised audio files and accompanying transcriptions may now allow us to further enhance our linked data by using named entity recognition to generate fuller metadata and identify additional associated individuals.
- **Organisational names**, such as National Union of Women's Suffrage Societies (NUWSS), Women's Social and Political Union (WSPU), and others were added as subjects of interviews where summaries indicated significant discussion.
- **Significant places and events**, such as imprisonments, specific arson events, or named marches enhanced contextual links.

The work with significant people raised an unexpected conundrum - should all the people mentioned in the interviews be included on the project's P5008 focus list in the same way as the interviewees and the people who were their main subjects? For example, the interview titled *Adams, Mrs Dorothy re Anna Munro* is with Dorothy Adams, talking about her employer, Anna Munro. Although Munro herself was not interviewed, she is the focus of the interview and would be valuably included on the focus list of the project by adding a statement to her QID to indicate this. The interview mentions a number of other names as well, such as Munro's husband, her children, both of whom are also interviewed as part of the project, and other notable figures in the suffrage movement. Wikidata encourages the creation of meaningful connections, so in that spirit all these names were added as significant people on the QID for the interview, but their inclusion on the project's focus list was less straightforward. Readers may hold differing views on the 'correct' approach, but after internal discussion it was decided that in the interests of suffrage research it was valuable to record as many names as possible so that the project could potentially bring to light new connections or relationships, and widen research possibilities. Guidance was given to the team in assessing how substantive a 'mention' was in order to avoid making excessively time-consuming connections with misleading value.

Interconnections and visualisations

One of the key advantages of working with Wikidata is the ability to explore and interrogate our data in ways that were not previously possible. In a MARC environment, metadata feeds into the library catalogue, where metadata specialists and many users are familiar with established search techniques to locate relevant

content. Wikidata, by contrast, offers a more dynamic and flexible approach. Using SPARQL queries we can break the data down into a whole variety of meaningful categories, offering researchers more granular insights. The categorisation of individuals for this project is supported by the data model used for interview QIDs, and further details can be seen in the *Names data* section on the project page. Importantly, by contributing our metadata to Wikidata, we can link it to existing entities and generate visualisations of those connections, something not possible when metadata remains siloed within an archives catalogue. For example, we can now explore questions such as:

- who else holds archives, or oral histories, of people connected with the interviews?
- What family relationships exist between individuals?
- What awards did they receive?
- What spread of occupations is represented?
- What significant events are mentioned across the interviews?

A full list of the SPARQL queries being used to interrogate the collection can be seen on the project page⁷.

Contributing to the wider Wikimedia community

While contributing metadata to Wikidata has been the cornerstone of this project, we also explored the broader suite of Wikimedia platforms to further enhance access and visibility. This is another step expanding the traditional roles of MARC21 or institutional repository metadata creators. As a new editor I found it to be an area where, despite my best attempts at interpreting Wikimedia policies, I made genuine mistakes, was corrected by other editors, and needed to revise my approach.

Those of us who manage metadata teams are often used to being the institutional experts in our domain, and less accustomed to others critiquing our metadata than we might once have been! Entering the world of Wikimedia and transparently seeking to make our content accessible to a wider global audience, in the right way, has been a steep learning curve, where I have needed to learn from the generosity of more experienced editors.

As an LSE employee I needed to declare this affiliation on my user talk page and avoid edits which could be construed as conflict-of-interest. This can perhaps be a particularly confusing area for university staff. Making unique content, with research potential, available for a global audience does not feel like an exercise in self-promotion in the same way as it might for a commercial entity, and yet the same policies and risks apply. In my case, I misunderstood guidance around external links and references to my employer in article text. Once alerted, I revised edits accordingly

⁷ https://www.wikidata.org/wiki/Wikidata:WikiProject_The_Women%E2%80%99s_Library_LSESuffrageInterviewsProject#SPARQL_queries

and engaged in discussion with fellow editors. I have also engaged in other areas of Wikipedia editing, rather than just those related to content from my institution, both to be, and demonstrate willingness to be, a 'good citizen' in this global knowledge space.

Perseverance and willingness to learn has enabled me to contribute to regularly viewed biographical and organisational Wikipedia pages and receive thanks from other editors. The Women's Institute article, for example, is viewed nearly 42,000 times a month ([Wikipedia, 2025](#)) at the time of writing, while the National Union of Women's Suffrage Societies sees over 14,000 visitors ([Wikipedia, 2025](#)). Biographical pages tend to attract lower page views than subject pages, though the actress Sybil Thorndike, one of the interviewees, receives nearly 51,000 monthly views ([Wikipedia, 2025](#)). Among other interviewees, the socialist politician Margaret Cole receives over 8,000 page views ([Wikipedia, 2025](#)), and the suffragette Leonora Cohen, over 4,500 a month ([Wikipedia, 2025](#)). We have also been able to add some images to Wikimedia Commons, which can now be re-used in Wikidata and Wikipedia, and in SPARQL generated image grids related to the project.

The global Wikimedia community enables collaboration beyond institutional boundaries, particularly on topics supported by established WikiProjects. The *Women In Red* project⁸ seeks to reduce systemic gender bias on Wikipedia by increasing the visibility of notable women who lack articles ([Wikipedia, no date](#)). The project maintains a list of women suffragists who do not yet have a Wikipedia article⁹ to which we have been able to contribute a number of names discovered through the Suffrage Interviews project. There may be future opportunities to create biographical pages for these under-represented women in collaboration with LSE curators or *Women in Red* volunteers, and a colleague has suggested these 'hidden' women could be valuable candidates for inclusion in the *Dictionary of National Biography*, offering another avenue for recognition and scholarly engagement.

Impact

As we draw closer to the end of the project, demonstrable qualitative impacts have included:

- Strengthening Wikimedia skills within the team.
- Sharing the Suffrage Interviews metadata for re-use on a globally accessible, open platform.
- Creating linked data connections between LSE metadata and existing Wikidata entities.
- Enabling new ways to explore the collection through SPARQL queries and visualisations.

⁸ https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red

⁹ https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red/Missing_articles_by_occupation/Suffragists

- Enhancing discovery for potential users through Wikimedia platforms.
- Sharing project outputs via our new Digital Scholarship webpage¹⁰

For the conflict-of-interest reasons outlined above the project did not centre on active promotion of LSE content via Wikipedia, so investigating quantitative outcomes has not been the main focus of assessing project impact. While it is useful to explore whether any early conclusions can be drawn, several important caveats apply:

- Due to Google Analytics changes we are only able to access usage data for the main interviews page back to April 2023 and for interview downloads back to June 2023. We began the project in April 2023 so 'before and after' comparisons are difficult.
- Initial basic metadata was uploaded to Wikidata in automated batches, but ongoing metadata enhancement of those QIDs took place across subsequent and overlapping months, again making comparisons difficult without a longer period of analytics data.
- Wider ongoing project work is still in progress as of summer 2025 so it is too early to draw firm conclusions from Analytics without longer term usage data from the completed project.
- Use of our systems tends to follow the academic year with drops during vacation periods.
- The collection is small and specialised.

This means that the analytics we currently have cannot be considered reliably indicative of long-term trends. Nevertheless, as an interim assessment, they suggest small but significant positive gains from the work:

- Active users over the last 8 months, compared to the 8 months during which we carried out initial data uploads and enhancements to Wikidata, have increased by 41%, suggesting more users have been attracted to the site and engaged with it in some way on arrival.
- Traffic coming to the suffrage interviews homepage from Wikipedia has increased by 179% in the last 12 months compared to the previous 12 months.

The Wikidata Suffrage Interviews project has opened up new possibilities for representing and connecting the Library's unique and distinctive collections. It illustrates the way in which moving beyond traditional and familiar cataloguing workflows can enable the exploration of innovative approaches to modelling, creating, and enriching metadata, exposing connections between related entities, and contributing metadata to a global linked data ecosystem. The project's iterative and experimental approach, on a collaborative platform, has resulted in access to, and visualisation of, the collection in ways that were not previously possible. It

¹⁰ <https://www.lse.ac.uk/library/research-support/digital-scholarship>

demonstrates how we can extend our metadata skillsets to support discovery and re-use, not only within our own systems but across the wider web of data.

AI statement

I used Microsoft Copilot to proof-read my final draft of this article, incorporating some of its suggestions to improve the clarity of particular paragraphs. I also prompted it to help me find quotes which could be appropriated for the article title.

References

- LSE Library (no date) *The Suffrage Interviews*. Available at: <https://www.lse.ac.uk/library/collection-highlights/the-suffrage-interviews> [Accessed 8 July 2025]
- Pankhurst, E. (1908) 'We are here not because we are law breakers; we are here in our efforts to become law makers'. *Women's suffrage*. Available at: <https://www.npg.org.uk/schools-hub/womens-suffrage-deeds-not-words> [Accessed 8 July 2025]
- Wikidata (no date) *Help: Sources*. Available at: <https://www.wikidata.org/wiki/Help:Sources> [Accessed 8 July 2025]
- Wikidata (no date) *Oral evidence on the suffragette and suffragist movements: the Brian Harrison interviews*. Available at: <https://www.wikidata.org/wiki/Q100380678> [Accessed 8 July 2025]
- Wikidata (no date) *The Women's Library LSESuffrageInterviewsProject*. Available at: <https://www.wikidata.org/wiki/Q117322976> [Accessed 8 July 2025]
- Wikidata (no date). *WikiProject_PCC_Wikidata_Pilot*. Available at https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot [Accessed 8 July 2025]
- Wikipedia (2025) *Page views analysis: Leonora Cohen*. Available at: https://pageviews.wmcloud.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2024-07&end=2025-06&pages=Leonora_Cohen [Accessed 8 July 2025]
- Wikipedia (2025) *Page views analysis: Margaret Cole*. Available at: https://pageviews.wmcloud.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2024-07&end=2025-06&pages=Margaret_Cole [Accessed 8 July 2025]
- Wikipedia (2025) *Page views analysis: National Union of Women's Suffrage Societies*. Available at: https://pageviews.wmcloud.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2024-07&end=2025-06&pages=National_Union_of_Women%27s_Suffrage_Societies [Accessed 8 July 2025]
- Wikipedia (2025) *Page views analysis: Sybil Thorndike*. Available at: https://pageviews.wmcloud.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2024-07&end=2025-06&pages=Sybil_Thorndike [Accessed 8 July 2025]
- Wikipedia (2025) *Page views analysis: Women's Institute*. Available at: https://pageviews.wmcloud.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2024-07&end=2025-06&pages=Women%27s_Institute [Accessed 8 July 2025]
- Wikipedia (no date) *Wikipedia:WikiProject_Women_in_Red*. Available at: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red [Accessed 8 July 2025]

- Wikipedia (no date) *Wikipedia:WikiProject_Women_in_Red/Missing articles by occupation/Suffragists*. Available at: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red/Missing_articles_by_occupation/Suffragists [Accessed 8 July 2025]
- Williams, H. K. R. (2022) LSE's adventures in Wikidata-land: tears and triumphs down the rabbit hole. *Catalogue and Index*, 206, pp. 2-6. Available at: <https://eprints.lse.ac.uk/114976/> [Accessed 8 July 2025]
- Williams, H. K. R. (no date) Wikidata:WikiProject LSEThesisProject. Available at: https://www.wikidata.org/wiki/Wikidata:WikiProject_LSEThesisProject [Accessed 8 July 2025]
- Williams, H. K. R. (no date) *Wikidata:WikiProject_The_Women's_Library_LSESuffrageInterviewsProject*. Available at: https://www.wikidata.org/wiki/Wikidata:WikiProject_The_Women's_Library_LSESuffrageInterviewsProject [Accessed 8 July 2025]