Assessing Nineteenth-Century Library Collections with the "Living with Machines" Project at the British Library, United Kingdom

Justus Kühler and Ines Mergel

Abstract As part of its commitment to innovation and to improve the accessibility of the library's collection of resources from the long nineteenth century, the British Library ran the "Living with Machines" project between 2018 and 2023. The Library's project uses Artificial Intelligence (AI) technologies to improve access to historical collections. The library faced the challenge of making the extensive collections easier to interact and more accessible to users. The project was initiated to explore how AI and machine learning (ML) would impact historical research, with a particular focus on the impact of mechanisation in the nineteenth century. The project team was gathered from 2018 to July 2019. The project involved over 5500 volunteers and engaged the public through crowdsourcing tasks. The collaboration spanned various disciplines, including data scientists, historians and research software developers. The British Library also reached out to the public to ask for their help in commenting on the artefacts. Collaboration challenges arose when staff did not spend enough time understanding each other's disciplines. According to the interviewees, librarians will need a solid grounding in "information organisation" when it comes to working with AI solutions and outputs.

Keywords Digitisation of historical data · AI cataloguing · National Library · British Library · Artificial intelligence · AI · Living with Machines · Volunteers · User engagement

J. Kühler
Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany
e-mail: justus.kuehler@uni-konstanz.de

I. Mergel (✉)
Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany

Fraunhofer FOKUS, Fraunhofer Institute for Open Communication Systems, Berlin, Germany

School of Management, Public Management, University of Vaasa, Vaasa, Finland
e-mail: ines.mergel@uni-konstanz.de

I. Mergel, C. Schmidt (eds.), *AI Innovations in Public Services*,
https://doi.org/10.1007/978-3-032-01344-6_12

1 Introduction

The British Library, based in London, is the national library of the United Kingdom and one of the largest libraries in the world. It houses approximately 170–200 million items from various countries. It was founded on 1 July 1973 under the British Library Act 1972 and was formerly part of the British Museum. Today, the library employs around 1700 staff, and it had a budget of over 177.4 million euros in the financial year 2022/2023 [1].

As part of its programme for innovation and accessibility, the British Library launched the "Living with Machines" project running from 2018 to 2023. This initiative used machine learning (ML) to analyse and interpret extensive historical datasets [2]. Therefore, the project made use of solutions such as Computer Vision, a broad term for analysing and describing patterns in visual content [3]. Optical Character Recognition (OCR) and Natural Language Processing (NLP) were used to digitise newspaper materials: "*In Living with Machines, we did a lot of different experiments using things like computer vision and natural language processing that have specific uses for different kinds of research projects. We'll collaborate a lot with people where they're applying a particular method on a specific collection, and we help assemble the collection that they could use for that work, and then we can publish datasets that are derived from that project or machine learning models and other outputs on our research repository.*" By digitising archive material such as newspapers and maps, the project made these resources accessible for research purposes and at the same time tackled bias in historical data. During the project, the organisation developed and tested AI models for tasks such as named entity recognition and the digitisation of the library's extensive collections. The project strengthened interdisciplinary collaboration between specialists from different research fields such as historians, data scientists and linguists. In line with the British Library's mission, the project fostered collaboration between the Alan Turing Institute, partner universities and the British Library. User engagement was also at stake during the Living with Machines project. Users interacted with the library's existing datasets, helping to improve their accuracy and quality.

2 Description of the Project

The project in the British Library was initiated to address challenges in managing vast and diverse collections, driven by the need to enhance access, improve discoverability and understand the impact of digitisation on user engagement. It also was initiated to explore the impact of AI and ML on historical research, focusing on the effects of mechanisation in the mid-nineteenth century. It aimed to anticipate future transformations in research methodologies and access to historical data. It was launched through collaboration between the Alan Turing Institute, which is the British Institute for Data Science, including AI and the British Library. The project was funded by the Arts and Humanities Research Council (AHRC) [4].

Initiated in 2018, the team was assembled gradually, completing its composition by July of the following year. Involving over 5500 volunteers, the project engaged the public through crowdsourcing tasks. One interviewee described the involvement of diverse volunteers as follows: "*I think that the point was that people who volunteered on Zooniverse weren't necessarily self-consciously museum goers, unlike people who went to the exhibition. So, it was a really new group of audiences to the library and to the Turing who hadn't previously had any relationship with either of the institutions*." Collaboration spanned disciplines, including data scientists, historians and research software engineers, with members joining gradually between 2018 and 2019 within the team of the British Library [5].

2.1 Need(s) Behind the Implementation

The British Library project uses AI to improve access to historical collections. The goal was to explore the impact of AI and ML on historical research. The interviewee states that there are two reasons which motivate the experimentation with AI that is to (1) improve the accessibility of collections and (2) to use these resources as data and create some sort of public value: "*There's two ways that I think about AI in the library; one is improving our ability to make our collections more findable, getting people to collection items that they can then use for their research, to find stories for creative uses, entrepreneurial uses, whatever. The other is people using our collections as datasets and studying them for their form rather than as individual items.*"

Therefore, it can be interpreted that it aims to make collections more accessible, provide contextual understanding and ensure inclusiveness in accessing historical materials. Also, for people not visiting the library that much: "*I think that the point was that people who volunteered on Zooniverse weren't necessarily self-consciously museum-goers, unlike people who went to the exhibition. So, it was a really new group of audiences to the library and to the Turing who hadn't previously had any relationship with either of the institutions*."

2.2 Actors Involved

The project was made possible by the collaboration of the Alan Turing Institute, the British Library, the Arts and Humanities Research Council (AHRC) and several universities in the UK.

Involving over 5500 volunteers, the project engaged the public through crowdsourcing tasks. One project member explained: "*The main point of public interaction was the crowdsourcing work that we did where we designed crowdsourcing tasks where the annotations that people/members of the public made would contribute to the building of a research dataset that would either be used to train a machine learning model and expand up our queries or used as an output in its own. The trick*

there was understanding what kinds of results would be useful for different disciplines in the project."

From an organisational perspective, the project was financed by the UK Research and Innovation (UKRI) Strategic Priority Fund, which comprises a total of around 988 million euros. Around 1% of the entire fund was allocated to the "Living with Machines" project at the British Library, around 11 million euros (approx. 9.2 million Pounds) [6].

The project was conducted highly interdisciplinary, including data science, history and research engineering. This fostered collaboration between the Alan Turing Institute, partner universities and the British Library. Based on Mergel et al. (2025) [7], we identified the following co-production phases in the implementation of AI (Table 1):

2.2.1 Organisational Level

The project received backing from the Alan Turing Institute and AHRC, securing funding from the UK Research and Innovation (UKRI) Strategic Priority Fund. In addition, the British Library collaborated with other UK institutions from the GLAM (Galleries, Libraries, Archives and Museums) sector, such as the National Football Museum and the National Railway Museum, to collect items researched and collected by the British Library's interdisciplinary experts [8].

Roles such as the library's project managers, data scientists, librarians and historians contributed to the project's success with their expertise and resources. This interdisciplinary approach contributed also to the main challenges of the project, missing understanding of the colleagues' work which is more time-consuming and considered to be rather boring: "*I think there was quite a lot of difficulty in finding because it was such an interdisciplinary project or multi-disciplinary project, finding things that resonated for everyone. We probably didn't spend enough time*

Table 1 Co-production process in the development and implementation of AI at the British Library

Co-initiation	Funding is provided by the UK Research and Innovation (UKRI) Strategic Priority Fund. This fund is an investment of 830 million pounds (approx. 988 million euros), which has been used to support 34 multidisciplinary and interdisciplinary research programmes in two investment waves. 9.2 million pounds (approx. 11 million euros) was made available for the "Living with Machines" project
Co-design	The project was conceived by the Alan Turing Institute, the British Library, the Arts and Humanities Research Council (AHRC) and several universities across the UK
Co-implementation	Experts from various fields such as computational linguists, historians, data scientists and librarians worked on the "Living with Machines" project and developed several data sets for further use and AI training
Co-use/production moment	The datasets were used both by the interdisciplinary library experts and by anonymous users who had free access to the datasets and used them for their own research purposes
Co-evaluation	–

sharing and really understanding what was inspiring for each other's disciplines and what the gaps were in each other's disciplines (…). So, it's not a technical issue, except for the scale of the collections, making anything a bit harder and slower. It's more of a collaboration issue where the time that it takes to coordinate and discuss takes away from time doing things, but it's also part of the work. I think we hadn't really anticipated—I think from the library, we knew, but I think for the researchers, they hadn't quite understood some of those compromises that would be necessary."

2.2.2 External Actors

The project engaged external vendors such as the Alan Turing Institute, an organisation working independently from the British Library but integrated in the same building. It was founded in 2015 as the National Institute for Data Science [8]. The close, also physical, relation between the Library and the Alan Turing Institute allowed for even closer collaboration facilitating the launch of the Living with Machines Project: "*the Alan Turing Institute is based physically in The British Library, so we knew we had an opportunity to work with all the data scientists and the cool people who were coming in to work with Turing.*"

The partners of the project were also other universities such as Cambridge, East Anglia, Exeter, Queen Mary University of London, the King's College, and the University of Edinburgh. Users and patrons, including researchers and historians, likely played a role by providing feedback, participating in crowdsourcing activities and utilising project tools for research to enrich the analytical scope and data quality. One interviewee describes this: "*They got to take part in that, and also the questions that they asked during the process really shaped how we thought about something of the work. Particularly, in the first crowdsourcing task that I designed, we asked does this accident involve the machine. Defining a machine actually turned out to be quite tricky because the 19th century definitions of the machine were different than what we would generally think of now. Even things like a potato peeler could be a machine, and we wouldn't think of a potato peeler as being a machine now.*"

2.3 Challenges

Despite the success of the project, it was also characterised by challenges. Challenges included interdisciplinary collaboration, technical expertise, staff rotation, copyright and access issues, and project management, but also the challenge of coordinating and agreeing on the core purpose of the project, as one of the interviewees from the library explained: "*Everyone had a different idea of what a research question is because everyone came from a different discipline*."

The large list of internal and external collaborators was challenging. The interviewees noticed that the collaborators did not spend enough time understanding each other's disciplines and the resulting differences in methods and theoretical

background: "*It's more of a collaboration issue where the time that it takes to coordinate and discuss takes away from time doing things, but it's also part of the work. I think we hadn't really anticipated—I think from the library, we knew, but I think for the researchers, they hadn't quite understood some of those compromises that would be necessary.*"

Essential but perceived "boring" tasks like cleaning datasets and matching metadata were often overlooked due to the pressure on researchers to publish papers and yielding for results quickly: "*So, it was not always easy to get as much attention on those boring bits when people were felt pressure because they were on this high-profile project to also be publishing papers all the time. In some ways, it was less productive for people who were looking at how many papers they could get to add to their CV because everything is slower when you're working across disciplines and when you're doing that negotiation around what's important to you.*"

2.4 New Skills for Librarians

Librarians involved in the AI projects of the British Library require a diverse skill set focusing on information literacy. Additionally, proficiency in digital tools and platforms is essential, enabling them to effectively navigate digital collections and utilise AI technologies like "machine learning" and "computer vision" tools for tasks such as "digitisation" and "metadata tagging": "*Then there's work around discovery; tagging images or tagging texts with entities/things detected by computer vision in images from books or things detected in text, people's names, places, concepts, whatever, as well as things that are more aimed at a metadata level for the whole item and not just looking at the text as a corpus. So, we bounce between different levels of where we apply these technologies and what they mean.*"

However, the interviewee also emphasised that collaboration does not change just because it is collaboration with AI, but that this type of technology also requires patterns of communication and collaboration that are already in use: "*(…) there's a continuity between AI and other methods. A lot of what we did ended up being AI, but it wasn't called AI when we started doing it, partly because AI is more of a marketing term than a technical term. Providing access to sources at scale, thinking about the logistics of ingesting things, running processes at scale, queries at scale, that doesn't really change with AI.*"

3 Results

3.1 Organisational Level

Overall, the project improved the accessibility and inclusivity of the library's collections, enhanced public engagement through exhibitions, workshops and crowdsourcing initiatives. But it has not fundamentally changed the organisation, as can

be seen from the statement above. The research results were oriented towards the feasibility within the complex project context and the many stakeholders. Although it was an AI project, many analytical tasks and methods remained unchanged and did not change fundamentally. The interdisciplinary collaboration and communication led to an adaptation of the methods used, but also did not change fundamentally: "*There's lots of different things that have happened, but they aren't necessarily as evident from the core. The way that you interact with the library hasn't necessarily changed that much as a result of these AI pieces of work.*"

3.2 Value Created and Co-created

The focus of the Public Value creation within the British Library had its focus on increasing the accessibility of vast and extensive collections coming from the nineteenth century. Furthermore, there was also the creation of stronger interdisciplinary collaboration among humanities and data science. The synergies of both disciplines were put together in the library itself as one interviewee explained: "*One of the outcomes in the project, actually we did a few public workshops or workshops at conferences, but also we trial things with wherever we could in the digital scholarship training program so that people in the library were learning as the project learned different methods. But also so that hopefully it meant that if you did a workshop or a tutorial in public, it had a trial run with people who would give you useful feedback.*"

The already mentioned data-crowdsourcing task can be mentioned as one major outcome of the living with machines project: "*The main point of public interaction was the crowdsourcing work that we did where we designed crowdsourcing tasks where the annotations that people/members of the public made would contribute to the building of a research dataset that would either be used to train a machine learning model and expand up our queries or used as an output in its own. The results of these crowdsourcing tasks align with definition of Public Value Creation which is the result of a public sector organisation collaborating with its environment and also strengthens the individual and normative basics on which the activities of a public sector organisation rely such as accountability, accessibility and ethical behavior of employees"* [10].

3.3 Lesson Learned

The main lesson that was learnt is to promote project members' understanding of their tasks. This is critical when it comes to the production of research results and the phases that involve solving tasks that are perceived as rather "boring". In connection with this realisation, the interviewee also states that the time pressure in the research process promotes this development of a lack of understanding: "*I*

think there was quite a lot of difficulty in finding because it was such an interdisciplinary project or multi-disciplinary project, finding things that resonated for everyone. We probably didn't spend enough time sharing and really understanding what was inspiring for each other's disciplines and what the gaps were in each other's disciplines, because things that are really established in one field might be new in another. But also, a lot of the practical work, because everyone was a researcher, even the software engineers were publishing papers based on their work. Some of the really boring stuff isn't necessarily something that you can publish a paper on."

4 Conclusion

The Living with Machines Project started in 2018 with the collaboration of the British Library, the Alan Turing Institute and the Arts and Humanities Research Council (AHRC) with the goal of enhancing interdisciplinary collaboration by using new technologies such as AI. The narrower focus of the project was to explore resources and extensive collections from the British Library, coming from the "long nineteenth century", making resources such as maps and newspapers available for the broader research community.

Beginning in 2018/2019, the initiative utilised Computer Vision, natural language processing (NLP) and optical character recognition (OCR) technologies to achieve its goals. It brought together interdisciplinary teams, including the Digital Research Team, data scientists, historians and research software engineers, as well as over 5500 public volunteers engaged through crowdsourcing tasks. While the collaborative effort was key to the project's success, it also presented challenges. These included difficulties in achieving mutual understanding between disciplines, delays caused by coordination and negotiation efforts, and overlooked technical tasks, such as cleaning datasets and matching metadata, due to the pressure on researchers to publish. Additional challenges stemmed from staff rotation, copyright restrictions and balancing preparation with the delivery of project milestones.

Despite these obstacles, the project created substantial value. It increased the public engagement with historical data resources, strengthened interdisciplinary collaboration with external and internal actors and made the library's collection more accessible for patrons.

Several key lessons emerged from the initiative: Effective interdisciplinary collaboration required early resource allocation. Balancing the need for timely delivery with adequate preparation can mitigate delays and improve outcomes. Overall, the project underscores the potential of AI to transform historical research while maintaining a focus on inclusion and innovation in the digital age.

Acknowledgements The work for this chapter was supported by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement ID 101061516.

Appendix

See Table 2.

Table 2 Overview of the AI project at the British Library

Case and project name			
British Library			
Country	**Number of employees**	**Type of AI solution**	**Year and maturity level**
United Kingdom	1700 (2011)	Natural Language Processing/OCR	Starting year 2018/2019
Project description			
The project at the British Library was initiated to address challenges in managing vast and diverse collections, driven by the need to enhance access, improve discoverability and understand the impact of digitisation on user engagement. Aligned with the library's mission, it aims to make collections more accessible, provide contextual understanding and ensure inclusivity in accessing historical materials			
Need(s) behind implementation	**Actors involved**	**Challenges**	
The project was initiated to address the challenges of managing large and diverse collections driven by the need to improve access, increase discoverability and understand the impact of digitisation on user engagement	It involved collaboration between the Alan Turing Institute, partner universities and the British Library Involving over 5500 volunteers, the project engaged the public through crowdsourcing tasks. Collaboration spanned disciplines, including data scientists, historians and research software engineers	Team members sometimes did not spend enough time understanding each other's disciplines and their gaps Mundane tasks like cleaning datasets and matching metadata were often overlooked due to the pressure on researchers to publish papers. Collaboration slowed progress due to the time needed for coordination and negotiation	
Results			
Organisational level	**Value created and co-created**	**Lesson learned**	
Overall, the collaboration required adapting to these methods but didn't fundamentally change due to AI Although many tasks involved AI, it wasn't initially labelled as such. The core methods and technical data management remained consistent, whether AI or other computational techniques were used	Encompasses access and inclusion, education and awareness, community engagement and fostering innovation It strengthened initiatives such as exhibitions, crowdsourcing, workshops and innovation These values are in accordance with the institution's strategic mission	Lessons learned include fostering interdisciplinary collaboration, early resource allocation, balancing speed with preparation, ensuring long-term technical support, raising staff awareness and promoting knowledge sharing	

References

1. British Library, *Corporate Report British Library Annual Report and Accounts 2022–23* (British Library, 2023)
2. R. Ahnert, The living with machines report (2023)
3. V. Wiley, T. Lucas, Computer vision and image processing: a paper review. Int. J. Artif. Intell. Res. **2**, 22 (2018)
4. T.A.T Institute, Living with machines (2025), https://www.turing.ac.uk/research/research-projects/living-machines. Accessed 19 Feb 2025
5. L.w. Machines, Understanding history with new eyes (2025), https://livingwithmachines.ac.uk/about/. Accessed 19 Feb 2025
6. U.R.a. Innovation, Living with machines (2025), https://gtr.ukri.org/projects?ref=AH%2FS01179X%2F1. Accessed 19 Feb 2025
7. I. Mergel, N. Edelmann, N. Haug, Co-production phases in the development and implementation of digital public services. Perspect. Public Manag. Govern. **8**(2), 93–105 (2025)
8. Living with Machines exhibition (2025), https://livingwithmachines.ac.uk/living-with-machines-exhibition/. Accessed 19 Feb 2025
9. Alan Turing Institute, About us (2025), https://www.turing.ac.uk/about-us#:~:text=The%20Alan%20Turing%20Institute%2C%20headquartered,artificial%20intelligence%20to%20our%20remit. Accessed 19 Feb 2025
10. N. Haug, I. Mergel, Public value co-creation in living labs—results from three case studies. Admin. Sci. **11**(3), 74 (2021)

Justus Kühler is studying for a bachelor's degree in political science and public administration at the University of Konstanz. He is particularly interested in the latest developments in public administration that reflect the trend towards democratisation. He is particularly interested in the application of new forms of participation at the local level. He has already gained initial experience with municipal administration through his involvement in local politics.

Ines Mergel is a University Professor of Public Administration in the Department of Politics and Public Administration at the University of Konstanz and a Fellow of the National Academy of Public Administration (Class of 2018). After holding positions at the Harvard Kennedy School of Government as a Doctoral Fellow and Postdoctoral Fellow (2002–2008), Professor Mergel was awarded tenure at the Maxwell School of Citizenship and Public Affairs (Syracuse University, NY) in 2014 (2008–2014) and then served as Associate Professor with tenure (2014–2016). Professor Mergel is a member of the supervisory board of the e-Governance Academy (Estonia). Other board memberships can be found here. Professor Mergel is a founding member of the international initiative Teaching Public Service in the Digital Age, which aims to integrate digital skills into the teaching and training of public managers. As part of this initiative, Professor Mergel was appointed a Schmidt Futures Innovation Fellow in 2022.