AI for Access: Strategic Innovation at the Dutch National Library

Barbara van der Vaart and Carsten Schmidt

Abstract Digitisation and access to information became increasingly important during the last two decades. At the same time, the challenges for all kinds of public services and organisations, also National Libraries, to hire new qualified staff members became even more challenging. To cover the needs and overcome the hurdles at the same time, the National Library of the Netherlands decided to introduce AI solutions and incorporate them into the traditional workflows like cataloguing. In 2023, the "Retrotool-Project" was initiated. The goal is to develop an automated metadata tool to accelerate and streamline cataloguing while ensuring data security, transparency, and ethical AI use. This tool will enable human-AI interaction to support cataloguers by automating repetitive tasks, allowing human experts to focus on complex metadata refinement. The "Retrotool-Project" can be seen as an example to introduce AI into a National Library to improve efficiency, quality, and accessibility of metadata while maintaining human oversight.

Keywords The Netherlands · National Library · Koninklijke Bibliotheek · Artificial intelligence · AI · Automated Indexing · Large language models · Bibliographic data · Retrotool · AI-assisted metadata extraction · AI in cultural heritage

B. van der Vaart
LIBER Europe, The Hague, The Netherlands
e-mail: barbara.vandervaart@libereurope.org

C. Schmidt (✉)
Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany
e-mail: carsten.schmidt@uni-konstanz.de

I. Mergel, C. Schmidt (eds.), *AI Innovations in Public Services*,
https://doi.org/10.1007/978-3-032-01344-6_8

1 Introduction

1.1 The National Library of the Netherlands (KB)

The Koninklijke Bibliotheek (KB), the national library of the Netherlands, is headquartered in The Hague. Established in 1798 and officially designated as the Royal Library in 1806, the KB has functioned as an independent administrative entity since 1993, receiving its primary funding from the Ministry of Education, Culture, and Science. Its mission is to collect, preserve, and provide access to all publications related to the Netherlands, ranging from medieval manuscripts to modern digital resources.

Currently, the KB employs around 500 staff members and houses approximately seven million publications, including books, magazines, and newspapers. In addition to its physical collection, the KB offers digital services such as the national Online Library (providing e-books and audiobooks) and Delpher, a platform granting access to millions of digitised pages. Since 2015, the KB has a coordinating role for the Dutch public library network.

1.2 Metadata Generation at the KB

Koninklijke Bibliotheek explains the importance of metadata in library management on its Web site: “Good metadata are important for bringing people and information together. They give meaning to the collection, making it easier for users to search through it. Good metadata ensure that the KB’s collection will remain visible, preservable and usable in the future” [1].

The work on AI tools in KB started several years ago. One interviewee explained: “We started about eight years ago in the field of AI, and for four reasons: first, to improve our internal processes; second, to enhance our services; third, to contribute to the public debate on AI implementation; and fourth, to provide high-quality data for AI development.” To help cataloguers with their metadata work, the KB looked for solutions to automate the process of generating metadata. In 2020, the KB began to develop a tool called Demosaurus, an earlier project aimed at assigning authors and keywords to books in the cataloguing system that was previously used. This tool supported the completion of two complex data fields: it assisted in assigning keywords to titles, making it easier for users to find the information they are looking for, and it facilitated the linking of authors to the author thesaurus, a process known as thesaurisation [2]. In its initial phase, the tool specifically focused on children’s books. However, due to lack of funding, the project was discontinued and is no longer in use.

In 2023, the KB started the development and implementation of an automatic metadata generation tool called Retrotool. The project was introduced in 2023 and has since achieved significant operational milestones. This current Retrotool is a

new and improved version that covers a wider range of materials and applies automatic metadata generation to physical books. It uses Optical Character Recognition (OCR) and AI technologies, such as Large Language Models (LLMs), to extract bibliographic data directly from books and create basic catalogue records. This process offers greater efficiency and accuracy than its predecessor.

2 Description of the Project

2.1 Automatic Metadata Generation: Retrotool

The KB's extensive collection comprises millions of books, newspapers, and magazines, both in physical and digital formats. The Collection Data department, staffed by approximately 30 cataloguers, processes metadata for about 50,000 physical items annually. Metadata, such as title, author, genre, and publication year, play a crucial role in ensuring the accessibility and discoverability of the library's collection. It facilitates efficient searching, retrieval, and exchange of bibliographic records at national and international levels. The catalogue also records each book's location within the KB. Without this information, the book becomes untraceable and effectively lost [3].

Over time, a substantial backlog of uncatalogued materials has developed due to factors such as restricted access during the COVID-19 pandemic and the transition to a new library system with different metadata standards. To address this backlog, the KB introduced Retrotool, an AI-powered solution for cataloguing older publications lacking metadata. Using OCR, Retrotool captures key bibliographic details from the title page and colophon, processes them through an LLM (currently Microsoft's ChatGPT), and assigns them to appropriate cataloguing fields. The tool then cross-references the KB catalogue to determine if the item already exists. If not, a basic bibliographic record is generated for cataloguers to refine before the final entry into the system.

2.2 Development and Implementation

The development of Retrotool was inspired by similar projects at the National Libraries of Finland and Belgium. After an initial proof of concept was completed in October 2023, the first basic functionalities were tested. Collaboration with a technical partner, although new to library systems, allowed for an iterative development process where feedback from cataloguers was rapidly integrated.

By April 2024, a fully functional prototype was tested, leading to refinements such as the ability to generate XML files compatible with the KB's cataloguing system. By September 2024, Retrotool was considered ready for production use.

However, certain administrative processes, such as finalising long-term support contracts, remain in progress.

Currently, Retrotool operates as a *Software as a Service (SaaS)*, hosted externally by the technical development partner. In the future, the KB plans to migrate the system physically in-house to maintain full control and ownership.

2.2.1 Initiation of the Retrotool

The KB has long recognised the necessity of investing in automated cataloguing processes to enhance efficiency and address the existing challenges in bibliographic management. Recent government funding, allocated as part of broader initiatives to improve public communication following the Dutch healthcare benefits scandal, the so-called *toeslagenaffaire*, has made this possible. This political scandal involved false allegations of welfare fraud by the Tax and Customs Administration against thousands of families receiving childcare benefits. It drew widespread public outrage and media attention, and it had significant political consequences, ultimately leading to the resignation of the Rutte cabinet in 2021.

The government funding was subsequently used by the KB to improve automated processes aimed at addressing the significant cataloguing backlog. A key project is Retrotool, focusing on cataloguing older books, primarily from the 1800s onward, which lack existing data. Using text recognition, Retrotool extracts information directly from the books, reduces manual input errors, and helps identify duplicates. This allows the cataloguers to concentrate on refining more complex bibliographic details, speeding up the process of uploading accurate descriptions to the catalogue.

2.2.2 Operational Process of the Tool

New as well as old publications are continuously added to the collection. As one interviewee explained, it was only in 1974 that the KB officially adopted the goal of systematically collecting all publications published in or about the Netherlands that meet its inclusion criteria, inviting all publishers to provide copies. Before then, not all publications were added to the collection. Retrotool is used to address these gaps in the collection, which are often filled through donations of (private) book collections, totalling thousands of books per year.

The operational process of Retrotool is as follows: the tool scans the colophon and title page using OCR technology, after which it combines this information with data from the Internet and an LLM (currently Microsoft's ChatGPT) to generate a basic record. An interviewee said: "The main purpose of this tool is to remove the doubles from the retro collection, which is gigantic, and to create a basic description of each book. We only have to check it, make minor additions, and then it is ready

to be uploaded to the catalogue." Another interviewee explained: "The tool reads the title page and colophon of a book, extracts key information, checks if the book is already catalogued, and if not, generates a basic description that a cataloguer can refine before finalising the entry." The extracted fields are then automatically checked with the library's catalogue on whether the book is already listed. If this is not the case, eight fields serve as a foundation for further cataloguing by a cataloguer, who completes the fields and then adds the record to the catalogue. Additionally, the tool enhances basic descriptions, providing a more detailed catalogue entry. An interviewee explained: "It takes away most of the repetitive tasks and allows cataloguers to focus on the more complex parts of bibliographic description."

2.2.3 Needs Behind the Implementation

Due to several factors such as the COVID-19 pandemic and the introduction of a new library system, the cataloguing department has accumulated a significant backlog. One interviewee said: "We have about 150,000 titles that we still have to describe in the catalogue. At our current rates, we would need about 15 years to catch up on that entire backlog if nothing else were to arrive at our library. So, we started to look at smarter solutions, and AI was one of the first things we considered removing repetitive tasks and allow us to focus on more complex issues." However, new items continue to arrive, so smarter solutions were explored, particularly in AI and automation, especially to streamline repetitive tasks. This would allow the cataloguers to focus more on complex cataloguing aspects and help to address the backlog more efficiently. However, human intervention remains essential in every stage, and the specific expertise of the cataloguers is what completes the process. Furthermore, an interviewee said: "One of the projects that is part of that larger whole is Retrotool, which I'm currently functioning as a project manager of. It consists mostly of books from 1800 onward that lack any existing metadata, requiring us to rely solely on text extraction from the book itself and external sources." Retrotool, therefore, is designed to assist, not replace, cataloguers—helping to make informed decisions faster and ultimately providing access to these works much sooner. An interviewee explained: "We are not replacing cataloguers with AI. Rather, we are giving them tools to make their work more efficient. AI handles repetitive tasks, but human expertise is still needed to ensure accuracy and contextual understanding."

2.2.4 Research and Development Approach to Innovation

The KB employs a *research and development-driven approach* to innovation, integrating AI and emerging technologies to enhance its services. This innovation process begins with exploratory research, guided by an internal strategic agenda that

identifies impactful areas for investigation. To support this approach, the KB establishes cross-departmental research groups, bringing together expertise from various divisions. In the case of Retrotool, collaboration was established between cataloguers from the Collection Data Department and the Development team to explore and evaluate potential automated metadata generation solutions. An interviewee said about the development process: "We use a structured approach: first, exploratory research, then proof of concept, then a minimal viable product (MVP), and finally, a fully operational prototype before full implementation."

Furthermore, the KB employs a *funnel approach* to innovation, progressing through distinct phases: exploration, proof of concept, and the development of a minimum viable product (MVP). Each project undergoes iterative testing and refinement, resulting in robust prototypes. Recently, Retrotool reached the prototype phase, enabling the automated generation of metadata descriptions. Throughout its development, and in close collaboration with a commercial technical partner aligned with the KB's AI principles, feedback has been systematically gathered from the Department Head, team leaders, and cataloguers within the Collection Data Department to refine the tool's functionality.

To ensure efficient technological implementation, the KB follows a *selective outsourcing strategy*, partnering with external providers for specialised expertise while retaining in-house development for core functionalities. This hybrid model allows the KB to incorporate cutting-edge technology while maintaining institutional control over critical processes. By early next year, Retrotool is expected to be fully integrated into the KB's standard cataloguing workflows. An interviewee said about future plans: "Retrotool currently operates as Software as a Service (SaaS), hosted by our partner company, but eventually, we plan to bring it in-house to ensure full ownership and control."

2.3 Actors Involved

The Retrotool project is based on the need to invest in automated cataloguing processes to enhance efficiency and address the existing challenges in bibliographic management. It is part of a broader initiative to improve public communication following the Dutch healthcare benefits scandal, the so-called toeslagenaffaire, described above. Table 1 provides greater detail on who the internal actors are in the Retrotool project and the activities that they developed in the project. The description follows the approach according to Mergel et al. (2025); co-production consists of five phases: co-commissioning, co-design, co-implementation, co-delivery, and co-assessment [4].

Table 1 Co-production activities in the Retrotool project at the Dutch National Library

Phase	Co-production type	Description
Co-initiation	Prospective co-production phase	The National Library of the Netherlands identified the need for AI-driven cataloguing due to a backlog of 150,000 uncatalogued books. The initiative was internally proposed by KB leadership, cataloguing staff, and AI researchers. The project was inspired by similar initiatives from other EU MS, e.g. FI and BE, with funding support from the Dutch government
Co-design	Concurrent co-production phases	Development of a prototype, integrating OCR and LLM (Microsoft ChatGPT, later moving to open-source LLMs) for metadata extraction. The co-design process involved cataloguers, AI specialists, and external developers, ensuring that the AI-assisted cataloguing process aligned with librarian expertise. Feedback loops were introduced to refine metadata accuracy
Co-implementation	Concurrent co-production phases	The prototype was tested by cataloguers and metadata specialists in early 2024, leading to adjustments in metadata field selection, XML export compatibility, and system integration. The tool was iteratively improved through librarian feedback, ensuring that AI-assisted cataloguing met NL bibliographic standards. Collaboration with external AI developers helped refine the tool's machine learning capabilities
Co-use/production moment	Concurrent co-production phases	Retrotool was officially deployed in September 2024, supporting cataloguers in processing new and historical materials. The system allows cataloguers to validate AI-generated metadata before the final entry. External research collaborations continue, with KB exploring further enhancements, such as automated thesaurisation and extended metadata fields
Co-evaluation	Retrospective co-production phase	KB conducts ongoing evaluations based on cataloguer feedback, metadata accuracy, and processing efficiency. External research institutions and AI experts are involved in assessing the tool's effectiveness, ethical considerations, and long-term sustainability. The transition to open-source AI models is a priority for future improvements

2.4 Challenges

While Retrotool is a significant step forward, several challenges remain:

1. **Use of Generative AI**: The KB currently employs Microsoft's ChatGPT to classify bibliographic data. However, due to concerns about data sourcing, transparency, and intellectual property rights, the goal is to transition to an open-source

large language model (LLM). The use of GPT-NL is being considered once this LLM is further developed.

2. **Employee Concerns**: Some cataloguers express apprehension about AI replacing human roles. However, Retrotool is designed to *assist rather than replace* human expertise, automating repetitive tasks to allow cataloguers to focus on more complex bibliographic descriptions. The expertise of cataloguers remains indispensable, as their input and guidance are crucial for the AI tool's effectiveness. Human oversight is essential when integrating information from an LLM into the catalogue, ensuring that all data generated by both the LLM and OCR is carefully reviewed. Ultimately, it is the responsibility of human experts to determine what should and should not be included in the catalogue. An interviewee described it that way: "There is still a very significant part of human intervention. AI removes repetitive tasks from bibliographic descriptions, allowing cataloguers to focus on the more nuanced aspects."

   The key to the success of Retrotool lies in striking the right balance between automation and manual labour and to apply AI responsibly. It is essential to identify the operations that can (and should) be automated, those that require additional attention from the cataloguer, and the functionalities that offer the greatest benefit to the cataloguers themselves [5].

3. **Ethical and Legal Considerations**: Questions surrounding the permissibility of AI-driven metadata generation within a national library context remain under discussion. An interviewee said: "The KB has explicitly said that we do not want to support commercial AI models that are built on illegally acquired text and information. While we currently use Microsoft's ChatGPT, we aim to transition to an open-source model in later versions of Retrotool." Ensuring compliance with data ethics and privacy regulations is a priority. It was precised by an interviewee: "We want to ensure that we're using AI ethically and transparently and moving away from commercial AI is a step in that direction." (…). Another interviewee commented: "In our case, we use a Microsoft version of ChatGPT, but we do not want it to share our data; instead, the data stay with us. We pose the questions, but these are not shared with the LLM for learning purposes; there is a checkbox for this that we can tick. We do not want our questions to be used to improve the LLM." Another interviewee explained: "We are developing an AI roadmap inspired by the French National Library's approach. This will help us systematically track our AI projects and ensure they align with our ethical principles."

   Regarding the legitimacy and unique role of national libraries in the future, an interviewee remarked: "I believe it is important for National Libraries and the library sector to establish clear standards for ethical values concerning AI applications and not leave this entirely to the commercial sector."

3 Results

3.1 Impact and Value Creation of the Retrotool at the KB

The introduction of Retrotool has influenced the workflow and development of metadata management inside the KB and caused significant positive results: Retrotool has demonstrated to provide important benefits for both the KB and its users, contributing to improved efficiency, accuracy in bibliographic data and by improved access to the collections. One interviewee said: "By making metadata generation more efficient, we are improving public access to books. If books are in our repository without metadata, they might as well not exist for the public."

3.1.1 Efficiency Gains

One of the most notable advantages of Retrotool is its ability to accelerate the cataloguing process, thereby significantly reducing backlog processing time. By taking over tasks in metadata generation, the tool enables cataloguers to work more efficiently, allowing for a higher volume of materials to be processed in a shorter timeframe. An interviewee said: "The tool allows us to describe and catalogue books at a much faster rate, making previously inaccessible books available to the public sooner." This is particularly important given KB's extensive collection, which includes both newly acquired and historical publications, and the significant backlog that has accumulated over the years.

3.1.2 Enhanced Accessibility

The timely inclusion of records in the KB catalogue ensures greater public access to historical materials. Previously uncatalogued or inaccessible items can now be integrated into the library's digital cataloguing systems, allowing researchers, scholars, and the public to discover and utilise a more comprehensive range of resources. This aligns with the KB's mission to preserve and provide access to Dutch cultural heritage.

3.1.3 Value Creation for the KB

Retrotool has created substantial value in two primary areas: internal cataloguing processes and data and collection management. By efficiently handling basic bibliographic descriptions, the tool optimises workflow efficiency without replacing

human cataloguers. As it takes away repetitive tasks from cataloguers, it allows them to apply their valuable expertise where it truly matters: to focus on more complex aspects of bibliographic description, such as subject classification, metadata refinement, and curatorial decision-making. This targeted use of automation ensures that human expertise remains central to the cataloguing process while alleviating repetitive, time-consuming tasks.

3.1.4 Broader Societal Impact

Beyond its institutional benefits, Retrotool enhances public access to the KB's collections, ensuring that previously unprocessed materials are available for academic and public use. By improving the discoverability of historical documents and rare books, the tool plays a crucial role in supporting research, education, and cultural preservation.

In conclusion, Retrotool marks a significant advancement in metadata management at the KB, successfully integrating automation with human expertise. By enhancing cataloguing efficiency and improving public accessibility, the tool serves as an important new asset to the modernisation of the KB's cataloguing system. However, while automation streamlines the processes, the expertise of the KB's skilled cataloguers remains indispensable, in order to ensure the accuracy, contextualisation, and intellectual integrity of bibliographic records. An interviewee said: "We still rely on cataloguers' expertise, but we now have a way to process and integrate descriptions more efficiently, helping to clear the backlog."

3.2 Lessons Learned

Retrotool was developed in an iterative project style, which started with a transition phase and piloting and led to full implementation. During the different steps of implementation, several key lessons are identified:

1. *Iterative development is crucial*: Close collaboration between cataloguers and technical developers by rapidly integrating feedback from cataloguers into the tool led to a system tailored to the library-specific needs.
2. *AI awareness and training are essential*: Addressing staff concerns through training and transparency is necessary for successful adoption of the AI tool by staff.

3. *User involvement may further improve outcomes in future AI projects*: Cataloguers have been actively and successfully engaged in the development process. Involving library patrons in future innovation projects could further enhance KB's service offerings.

   Even if Retrotool is already fully implemented, development and lessons learned are not finished. Looking forward, continued refinements, including new AI functionalities (such as text recognition of handwritten manuscripts) and expanded metadata fields, will further optimise Retrotool's impact in national bibliographic management.

4 Conclusion

The implementation of Retrotool at the KB demonstrates the potential of AI as a valuable support tool in cataloguing, by streamlining repetitive metadata tasks while ensuring that human expertise remains central to complex decision-making. Rather than replacing professional cataloguers, the tool enhances their efficiency, allowing them to focus on higher-order intellectual tasks. By accelerating the cataloguing process, Retrotool has played a crucial role in reducing the backlog of 150,000 books, significantly cutting down the time required for metadata generation.

Despite these efficiency gains, human oversight remains essential. AI-generated metadata must be carefully reviewed and refined to maintain the accuracy and quality of catalogue records. The KB recognises the ethical implications of AI adoption and is committed to transparency. As part of this effort, the institution aims to transition from proprietary AI solutions, such as Microsoft's ChatGPT, to an open-source language model, to safeguard greater alignment with ethical standards and institutional values.

Looking ahead, Retrotool's scalable architecture provides a basis for further refinements, including expanding metadata capabilities and enhancing automation while preserving human professional oversight. Collaboration with the Royal Library of Belgium and other institutions highlights the potential for shared learning in AI-driven cataloguing innovations. By bringing Retrotool in-house in the near future, the KB aims to ensure long-term control, enhance customisation, and sustain the tool within its metadata workflows.

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 Retrotool project at the National Library of the Netherlands

Case and project name			
Automatic metadating project: Retrotool			
Country	**Number of employees**	**Type of AI solution**	**Year and maturity level**
The Netherlands	500	Creation of an AI tool for the automatic metadating of book publications from 1800	The project was initiated in mid-2023. Retrotool was developed and tested, and in September 2024 it was officially implemented
Project description			
Retrotool was developed to address cataloguing backlogs by processing the retro collection—old books that were acquired but not yet catalogued. Using a document camera (OCR), the tool captures information from the title page and colophon, and artificial intelligence organises it into cataloguing fields. The tool then checks if the book is already in the catalogue. If not, it adds the book with a basic description, which a KB staff member can further refine			
Need(s) behind implementation	**Actors involved**	**Challenges**	
Diminishing the backlog in the cataloguing of books, with the additional benefit that repetitive tasks are taken over, allowing the cataloguers to focus more on complex cataloguing aspects	KB Management: Deputy director (interviewed) and the head of research (interviewed) Development team Collection Data Department: Head of Department, Innovation Coordinator (interviewed), and Interim Innovation Coordinator (interviewed) Four cataloguers who tested and currently work with the tool Technical company that developed the tool	Ethical challenges (use of LLM created by Microsoft is opaque) and apprehension by employees that AI would make their work obsolete	
Results			
Organisational level	**Value created and co-created**	**Lesson learned**	
Department of Collection Data	This initiative expands public access to a broader collection of KB publications while enabling cataloguers to focus on the more critical aspects of bibliographical descriptions. By automating repetitive and tedious tasks, it allows them to apply their valuable expertise where it truly matters	The focus of the innovation process remains on expanding and improving the tool, with iterative feedback from cataloguers and other staff incorporated at various stages AI awareness and training is essential; addressing concerns by staff through training and transparency is necessary for successful adoption of the AI tool by staff	

References

1. Koninklijke Bibliotheek, *Working with Metadata* (KB, National Library of the Netherlands, n.d.), https://www.kb.nl/en/over-ons/expertises/metadata. Accessed 10 Feb 2025
2. D. Haagsma, Uitzoekassistentie van de Demosaurus. Vakblad voor Informatieprofessionals **08** (2021) https://zenodo.org/records/5705945
3. W. Van Wezenbeek, *Wilma Ontdekt: 50.000 Boeken, Kranten en Tijdschriften per Jaar* (KB, de Nationale Bibliotheek, 2024), https://www.kb.nl/blogs/wilma-ontdekt-50000-boeken-kranten-en-tijdschriften-jaar
4. I. Mergel, N. Edelmann, N. Haug, Co-production phases in the development and implementation of digital public services. Perspect. Public Manag. Govern. **8**(2), 1–13 (2025)
5. M. Buesink, AI gedreven ondersteuning voor en door catalogiseerders. Vakblad voor Informatieprofessionals **05** (2024) https://informatieprofessional.maglr.com/kb-onderzoekskroniek12

Barbara van der Vaart has many years of experience in international project management and has worked as a project coordinator and policy officer in academic research projects and governmental and international organisations. Before joining LIBER, she worked on two ERC Grant projects at Delft University and Leiden University. She obtained an MSc degree in Cultural Anthropology at Utrecht University, where she also studied Child and Adolescent Psychology.

Carsten Schmidt focuses on concepts and methodologies for e-government in the European Union and the associated countries, as well as the related aspects of digital infrastructure. Particular attention is paid to the areas of the once-only principle, European large-scale pilot projects, and sustainability. Before, he was involved in different positions (Project Director, Work Package Manager, etc.) in European large-scale projects like mGov4EU, TOOP, e-SENS, and e-CODEX. He worked for the Ministry of Justice NRW, Germany, the Technical University of Tallinn (TalTech), and the University of Tartu, Estonia.