Automated Subject Cataloguing at the German National Library

Ines Mergel, Justus Kühler, Anna-Lea Schumann, and Corinna Nitsch

Abstract The German National Library (DNB), as mandated by the German National Library Act of 2006, is responsible for collecting all German cultural artefacts, including digital and non-physical resources. To meet this requirement, DNB engaged in a Machine Learning (ML) project to improve metadata generation. After a while the project transitioned to a completely automated approach called subject cataloguing machine using Annif, an AI tool from the National Library of Finland, incorporating ML or subject indexing and metadata creation. This AI development relies heavily on the expertise of librarians, as AI tools are at times not fully accurate. While AI enhances efficiency, librarians and cataloguing specialists remain essential to ensure quality. The transition faced challenges, including limited resources, ethical issues, a diversified workforce's attitude towards AI, and the complexities of copyright laws. Despite these challenges, the integration of EMa has improved the generation and accessibility of metadata within DNB services. While AI can enrich librarians' skills, it cannot replace them. Mistakes encountered along the way should be viewed as learning opportunities, and future advancements should involve collaborative, diverse teams to continue refining the integration of AI in metadata management.

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

J. Kühler · A.-L. Schumann
Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany
e-mail: justus.kuehler@uni-konstanz.de; anna-lea.schumann@uni-konstanz.de

C. Nitsch
Ministry for the Environment, Climate and Energy Sector of Baden-Württemberg, Stuttgart, Germany

I. Mergel, C. Schmidt (eds.), *AI Innovations in Public Services*,
https://doi.org/10.1007/978-3-032-01344-6_3

Keywords German National Library · DNB · AI · Metadata · Cataloguing · Machine learning · Annif · EMa · Generative AI · Library innovation · Digital governance

1 Introduction

The German National Library (DNB) was founded in Leipzig in 1912 and expanded in 1947 with an additional location in Frankfurt [1]. As the central archive and information centre in Germany, the DNB provides access to knowledge and at the same time drives technological innovation, particularly in the field of artificial intelligence (AI). The tasks of the DNB are defined in the German National Library Act (DNBG) [2]. This law provides for the collection, cataloguing, and public accessibility of all media works published in Germany since 1913 in digital and printed form and media works published in German abroad. The DNB employs 617 civil servants and employees under collective agreements [3].

This case study shows that the DNB uses AI in projects to create added value for society in terms of inclusiveness and accessibility of resources and to optimise internal workflows. Two critical AI solutions were developed in succession: PETRUS, a software-based platform for automated subject indexing launched in 2009, formed the basis. The providing company could no longer efficiently provide a working application, and both sides decided to discontinue PETRUS in 2018: "*Over the course of the years, we had always made new requirements with them, and they continued to develop it for us. And the people who programmed and developed it, the first specialists, have all gradually migrated to Silicon Valley in the USA. In other words, we always had new people who had to learn the ropes, who had to understand what we wanted again.*"

For further development, the DNB decided to develop the project on their own to even further customise it to the DNB's needs. This was the starting point for EMa. In alignment with the German National Strategy for AI, this AI model for cataloguing was implemented in 2021/2022 and adapted the National Library of Finland's open-source Annif to the DNB's needs. This is a cataloguing engine for the automatic generation of descriptive metadata.

2 Description of the Main Project

The German National Library is testing and incorporating more than one AI project in its organisation. Next to the PETRUS and its successor, Project EMa, there is the so-called German Exile Archive 1933–1945 that allows visitors to interact with prerecorded interviews of holocaust survivors. Here, the German National Library

created an interactive exhibition with documents from authors in exile during *National Socialism* in Germany, as well as pre-recorded interviews with survivors of the Holocaust. Visitors can ask questions, and the AI searches through the more than 900 interviews and identifies matching responses by topic area. The library tested this application on a smaller scale to see how visitors react to this type of communication and interaction [4].

However, this chapter mainly focuses on the subject cataloguing machine EMa, ("Erschließungsmaschine"). The German National Library, in general, distinguishes between automated content indexing and automated formal indexing of records. Automated content indexing involves the generation of metadata that supports information retrieval in the DNB holdings. The application's core task, therefore, is to open "*the possibility of classifying or verbally indexing publications that would otherwise not be indexed at all or only partially*" [5]. The DNB usually receives the formal indexing data from the publishers or the German Publishers and Booksellers Association. However, one interviewee emphasised: "*But the topic [formal indexing] is not off the table because there are also access points where we have no formal data. And that's why we're also looking at ways of extracting the formal data from the digital copy of the title page, for example, and assigning it correctly.*"

The first experiments with automated indexing started with the PETRUS project in 2009. The goal was to improve library operations through process-supporting software. They implemented four main scenarios: *automatic subject classification*, *keyword assignment*, *automatic linking of personal names*, and *parallel matching for print and online publications* [6]. Over the years, challenges arose, including personnel turnover and the need for modular solutions. In addition, the DNB worked closely with an external company that developed the software for them. In 2018, the company decided not to develop the software further. One interviewee stated: "*And that was no longer lucrative for the company, so they discontinued it, and we then had to see how we could do it with the next AI project we started. We then called it the cataloguing machine EMa and relaunched the entire process, so to speak. We've learned a lot over the 10 years and realised that this was actually our original goal, but we had strayed from it a little. We wanted a modular system that we could develop further and add to.*"

In 2018, the DNB initiated a new project called EMa, translated as "subject cataloguing machine". The project aimed to redesign the indexing system using the open-source toolkit Annif from the National Library of Finland, interdisciplinary expertise, and different in-house produced AI processes. EMa was enrolled in April 2022 and generated descriptive metadata to enrich the records in the catalogue [7].

EMa generates DDC subject categories (Dewey Decimal Classification, which identifies a numerical designation for the resource at its specific content location [8]), subject headings from GND ("Gemeinsame Normdatei" (German Integrated Authority File) can be used in the DNB and other libraries to index resources and describes entities, i.e. persons, bodies, subject areas, etc. [9]). Therefore, EMa creates text data by using ML.

A key advantage of the introduction of these machine approaches may be beneficial is the challenge of a lack of skilled labour force in libraries, which will enhance in the future. An interviewee stated that the use of AI in cataloguing systems might also mitigate the consequences of skilled labour force shortage in the library sector: "*A quantitative perspective would be that we will have to manage with fewer people.… If there are simply not as many people entering the industry*."

2.1 Need(s) Behind the Implementation

The DNB has the legal obligation to collect copies of almost all media published in Germany since 1913, which has led to an enormous accumulation of data. This is also the result of new legislation that came into force in 2006, the German National Library Act (DNBG). This law distinguishes between media works in physical form and non-physical form. Media in physical form are all representations on paper, electronic data carriers, and other media; non-physical form is all representations in public digital networks. This results in a huge amount of resources that need to be indexed, which is why collection habits have changed, together with the distribution between occurrences in printed and digital media, e.g. one-third printed and two-thirds digital media in 2022 [10].

To handle this huge amount of data and gain comprehensive access to its collections, DNB improved the content-related cataloguing and generated bibliographic data. By using ML to capture the semantic context, searchability is facilitated. In general, with the EMa application, it is now possible to extend to further data records, as one interviewee describes: "*And that means we can no longer just open up on a book basis as before, but we can extend to journal articles for articles. And that is a tremendous added value created by these automated processes, so to speak, and offers completely new possibilities.*"

Moreover, the top-level library management must drive and support the introduction of changes brought about by AI: "*This development only exists at the German National Library because it was initiated by the Directorate General. It was also strongly supported despite personnel changes. It wouldn't work without them because it has a bit to do with structural changes.*" Another interviewee stated: "*I would say that the drive mainly comes from within at the moment. We simply want to make better use of what we do and what we have.*" Therefore, the projects respond to the need for faster and more efficient library processes. The aim is to develop or acquire complex systems, e.g. EMa, to enhance these operations, aligning with the national library's mission to collect and provide accessible information.

2.2 Actors Involved in Co-producing EMa

The project is supported by a multidisciplinary team, including experts in mathematics, computer science, and data science and librarians with domain expertise. Collaboration with external vendors, universities, and other (scientific) libraries enhances knowledge exchange and provides additional resources. Top-level support ensures funding and distribution of resources and facilitates coordination across departments. Additionally, legal experts contribute to developing policies for responsible AI use within the organisation.

2.3 Internal Actors Involved

The department "Acquisitions and Cataloguing" is responsible for collecting and producing cataloguing data. The domain includes the section "Automated Cataloguing and Online Publications", which collects and manages online publications such as eBooks, eJournals, and Web sites. It is also responsible for developing automatic processes for formal and content indexing. The section is therefore in charge of further developing EMa and consists of a cross-location team of around 30 people: Staff with different professional backgrounds, such as computer linguists, mathematicians, and experienced librarians, are involved in the AI project. The reason behind the broad scope of expertise was summarised by one interviewee: "*But I have to say that you always need a certain level of expertise, professional expertise. So you can't say you can do the whole thing without specialised staff.*" Most of the development and implementation of the AI project is carried out internally by the project staff. However, they are regularly supported by external partners.

The section "General collection management and use issues" is involved in experimenting and working with various types of innovation, from technical to social. This section is part of the "User Services and Preservation" domain and focuses on cutting-edge innovation and its implementation. They do not have any operational tasks regarding the acquisition of media, but they also seek to incorporate perspectives from other parts of the library.

Librarians' critical task regarding EMa is to monitor the AI outputs as described by one interviewee: "*In my view, AI or certain AI applications are very good at processing and structuring large amounts of information. […] the evaluation, the final assessment, is still a human matter. AI may be able to make recommendations, but I don't think AI can solve ethnic issues*." Furthermore, this interviewee stated: "*So the AI does something for us, but we look at the results. And we also evaluate these results […] and use them to feed a cycle, of course. […] They [librarians] don't look at all the results, but we have a quality assurance system*

for what is generated with AI support because we have a quality requirement […] And this has not become superfluous in all these years." Finally, another interviewee sums it up: "*And that was important to me right from the start, that you always see the interlocking, i.e., that you interlock the mechanical with the intellectual and make it clear from the beginning that you always need this intellectual [human] input.*"

2.4 External Actors

Further development in the area of AI requires intensive collaboration with external partners who are in possession of relevant expertise. Collaboration with universities, such as TU Darmstadt, Leipzig University, or Dresden University, with other libraries, such as the Technische Informationsbibliothek Hannover or other National Libraries, and with research institutions, such as the Fraunhofer Institutes, plays an important role. The academic institutions play a more central role in guiding the library scientifically in the development and exploration of the possibilities of AI. They provide special input when it comes to data processing, data quality, and the future development of artificial intelligence. One interviewee stated that the exchange of information with external partners is crucial for the library's development, not only in the field of AI: "*And I mean, outside of the AI project, we are always interested in exchanging information, including with other libraries, of course.*"

However, it is not only academic institutions that are involved in providing AI solutions in the German National Library. The DNB has collaborated with a startup from the City of Heidelberg, Aleph Alpha, to experiment on language models such as Aleph Alpha Luminous to improve the already mentioned GND. This language file describes entities in which the respective resource can be contextualised.

Bringing together diverse professional staff backgrounds and the expertise of external partners is known as co-production. According to Mergel et al. (2025), co-production consists of five phases: co-commissioning, co-design, co-implementation, co-delivery, and co-assessment [11]. Table 1 provides an overview of the different co-production phases in the development and implementation of EMa at DNB.

2.5 Skills and Competencies

The development and progress of AI at DNB is largely determined by the attitude towards AI and the error tolerance of employees. It must be noted that AI applications are not 100% accurate. Therefore, the expertise of librarians and resource indexing and cataloguing specialists is key. While expertise is important, motivation

Table 1 Co-production activities in the EMa project at the German National Library

Co-initiation	The DNB initiated the EMa project in the context of the German AI strategy
Co-design	The DNB designed EMa in-house and incorporated the AI application Annif which was developed by the Finnish National Library, to implement a modular AI solution that can be customised to the specific needs of the DNB. The DNB was supported by various external partners such as universities
Co-implementation	For implementing EMa the DNB used diverse professional staff backgrounds and was in exchange with external partners
Co-use/production moment	Librarians monitor the sample outputs of the AI solution
Co-assessment	The co-evaluation process takes place permanently by using samples of ML-Outputs and is benchmarked in-house by the standards of quality management. The interviewees state that there has been some improvement of the application

and willingness to experiment (with AI and in general) are crucial starting points. Above all, library management needs the skill of patience: "*Just saying something is not enough. And things really are more complex than you think.*" One interviewee sums it up by saying: "*In general, I would say that I think we can continue to do the majority of our business with the expertise we have built up. But I think you just have to have a keen eye and take note of things and not accept everything uncritically, including when it comes to AI.*"

The job description of librarians is also changing, as they have to deal more and more intensively with data management in the course of using AI: "*It's less about fulfilling library tasks intellectually, but you have to work much more intensively with the data, process the data in such a way that the methods, which are hopefully the right ones, lead to good results.*" Henceforth, librarians need to acquire some technical skills: "*And that means that librarians need to be tech-savvy. They don't have to be computer scientists, but they do need to be tech-savvy. We need people who can deal with large amounts of data and who have data management skills. And we need an awareness that the goals are different.*" As a consequence of this, an interviewee stated: "*(…) we need to open up the library to other people with other basic skills. But that's not really a new process. I think it started a long time ago with the introduction of information technologies in libraries. And of course, we also need to change the training of librarians themselves. Other priorities need to be set.*"

This is consistent with the "lifelong learning" skill that several interviewees mentioned: "*it's a case of continuous professional development in many places. I don't think it's ever been the case that you can make it to the end of your career with exactly the knowledge you started with.*"

Therefore, it can be summarised that AI-based solutions can enrich the skills and competences of librarians, but cannot replace them. Librarians will still have to make the final decision on the indexing of resources, especially critical and complex

resources. The AI solutions already in use are not able to do this, as one interviewee said: "*In my view, the evaluation, the final judgement, is still a human question.*"

3 Challenges to Implementing AI in the DNB

The introduction of AI-based solution in the German National Library also poses challenges. These challenges can be summarised in seven main fields.

The first challenge arises when the data used by AI applications is evaluated. AI tools are never consistently free from biases. At first glance, the access to knowledge is expanded, but the demands on librarians' information literacy increase as the information needs to be assessed: "*Where does this information come from? How is this information labeled? Do I remember where the source was? What was the source? Is the source actually 'clean'?*". In addition, one interviewee emphasised: "*I believe that libraries will face a debate about how we position ourselves as places where you can find reliable information and also information where it is clear how it came about.*"

The second challenge was already mentioned above when it comes to the perceived threat of job loss among librarians. Whereas the interviewee above mentioned that there will be the need of librarians collaborating with the AI application: "*And if this job changes now and disappears, then I can well understand that this is difficult for you (…)*". Related to this kind of challenge is also another aspect. This aspect mitigates the threat of job loss. Still, it stresses the enhancement of a librarian's job portfolio and thereby the need to update their AI literacy and competencies to face less accurate results of AI applications, as already mentioned above: "*the quality is getting worse, and you're embarrassing yourself to the outside world. So that has been a factor because, of course, we also have a lot of mistakes in there, so the processes are not perfect, and especially in the library context, this perfectionism was and is very pronounced, which is also right and important.*"

The third challenge is thereby characterised differently. The limited availability of resources is challenging. Enough funding can be a problem, and the constant battle between personnel and material resources is always a delicate matter. One interviewee stated that there is a lack of enough funding to train the AI models with data that has sufficient quality: "*This means that certain pre-selected training data has a certain quality and is made available. But if so, then it's more of a niche topic and not a broad one, because I don't think that individual libraries at least have enough money to do something all-encompassing.*" Furthermore, public tenders have the potential to complicate project implementation, often resulting in significant delays. There is a growing demand for greater agility within the administration to facilitate more efficient and timely project execution.

The fourth challenge describes the complex task of implementing AI solutions in the whole organisation. When experimenting with the AI solution in only parts of the library, the implementation is not as difficult as implementing it in the whole

organisation: "*If we really wanted to implement this, it would affect so many areas and involve so many dependencies, including the publication process itself. That's why I always like to use this comparison: In the lab environment, we have faster and simpler ways to achieve things than if we were to offer something for the DNB as a whole.*" Therefore, it must be addressed that integrating AI into the organisational structure is a multifaceted task.

The fifth challenge describes the transparent handling of the use of AI. The interviewees and experts in the German National Library already see some constraints in the broad use of AI. In the DNB, it is evident which data records were created by machine or intellectually by autopsy. Nevertheless, the question arises as to whether standardised labelling of AI-generated outputs is necessary. An interviewee described it as follows: "*Perhaps even a kind of stamp, a watermark, indicating what kind of information was generated by AI*", could help to make the use of AI more transparent.

The sixth challenge addresses ethical considerations related to the use of AI, especially in the context of language models. The necessity of considering ethical issues in developing and training AI models is emphasised by several interviewees. They address ethical biases, meaning that the resources used to enhance machine learning could contain harmful language or misinformation: "*But we absolutely need the expertise. I mean, if AI does something for me, I don't know. When you ask these ChatGPT things, you can see that some parts are terribly wrong, expressed in the most eloquent language.*" Furthermore, this issue highlights the dilemma institutions such as the DNB face regarding the decision to make their high-quality content available for training AI models. Especially in the German context, the danger of resources created during the time of *National Socialism* poses a substantial threat: "*But libraries, for example, still have collections from the times of National Socialism or colonialism, which means they also hold data that would normally need to be put into context before being made publicly available. If such data were simply released without curation, one would have to carefully consider in what form this could be done. But in some cases, that would certainly be quite challenging.*"

The seventh challenge, again, is related to aspects of training data. In this context, instead of judicial concerns, the organisation is mainly considering ethical challenges of training AI with existing data. The copyright status of most records imposes limitations on the extent to which the DNB can utilise them: "*Most of our collection, yes, is copyright protected, and there are legal, strict legal limits on what you can and cannot do with it.*" Also, when it comes to the internal use of different AI applications, the legal department of the library is worried: "*And if you read through some of the terms and conditions, it says (…) we immediately bag everything you enter and what comes out. And if any copyrights are infringed, which nobody really knows in the end because it hasn't been clarified by the courts yet, then we're out of it. Well, some of these terms and conditions are also written very flippantly. So, I just see a lot of risks because a lot of things haven't been clarified yet.*"

4 Results

4.1 Organisational Level

Four years after its inception, the EMa was implemented. The initial temporary project group has evolved into a whole organisational unit that is collaborating with several other project groups across the library.

The usage of AI in different departments of the DNB leads to new and more effective ways of working. This also includes interdisciplinary work and the breaking down of specialist silos: "*Everything is much more interrelated than it used to be. Even the individual subject areas, from a librarian's perspective, are merging more and more.*" However, it has to be stated that the project is still being improved and monitored. One interviewee states: "*So we are networked with various things and we really do research in that sense. We are really trying to go into depth and understand this in order to find out for ourselves, for our purposes, how we can best proceed in the future and improve productive operations. So we have already improved it enormously.*"

There are still unresolved issues, including how to deal with legal and ethical considerations, but clarification is already being worked on with several stakeholders inside and outside the DNB.

4.2 Value Created and Co-created

The German National Library considers it its duty to uphold democratic values that are closely tied to the thoughtful consideration of information sources. However, AI outputs are not necessarily unbiased and have the potential to compromise these values. Consequently, the DNB aims to serve as a platform through which society can stay informed about various developments in AI. In addition, the DNB is a reliable source of information for AI tools. The DNB's role as a data supplier is closely linked to this, e.g. by providing metadata for other academic and public libraries.

Furthermore, in a changing society, libraries are increasingly evolving into a protected space that is not merely a place of education but also a place to meet and exchange. The vital role of libraries in society is further reinforced by AI adoption, which enhances accessibility, search functionalities, and personalised library services. Nevertheless, it is emphasised: "*[…] we need to make people more aware that we have these skills, materials, content, collections and also this knowledge and that we can communicate it better, that we also have expertise on AI, for example, and that we can also be consulted on questions relating to AI as such.*"

A certain proportion of the DNB staff's attitude has shifted following the introduction of AI-supported methods: "*Younger colleagues have a different attitude to this than many older colleagues, and it very much depends on their personal attitude. Many colleagues are going along with the path, but the big problem is that there are also large areas that are not going along because the results are not yet such that they can be fully satisfied. This problem of acceptance and integration into everyday library work is a much bigger task than many people imagine.*" However, changes in technological approaches might also support the development of new standards for cataloguing within the library community.

4.3 Lesson Learned

The experimentation and implementation of new technologies, such as AI-based solutions like EMa, are always associated with lessons learned. These experiences, which are presented below, help DNB to learn and develop its AI-based solutions further.

The DNB has a reasonably large IT Department, but more developers and AI specialists are needed for the automatic cataloguing unit. Moreover, it is crucial that the IT department stays up to date and has access to training and education to properly support the automatic cataloguing unit.

The team's composition was influenced by an employee with an IT background joining: "*It wasn't planned, but it had a very positive effect on our development. The IT perspective offered new insights into methods and processes, enabling us to set a course that IT, which was too far removed from our specialist topics, could not achieve at the time.*" As the project progressed, highly skilled specialists were intentionally recruited as a result.

Also, the DNB must be aware of AI's inherent complexity, which extends beyond the realms of methods and implementation. One interviewee summarises this: "*Looking back, I would say that it is very important that the institution is aware that this is not a sprint, but a marathon. It's not enough to select a method and implement it. That's a bit of a suggestion because the technology is developing so quickly. The methods are only a small part of the whole.*" No technical applications should be implemented solely based on their current popularity. Instead, only technical applications that provide demonstrable added value for the DNB should be considered for implementation. However, one interviewee elaborated: "*We don't need to adopt technical applications that don't add value for us. We don't have to follow every trend. However, as a national library, we need to be knowledgeable about this environment to have a say. What is effective, what can be effective, and what might.*" Ultimately, a rethinking process towards AI is essential at all levels in the library, which takes time: "*People expect quick results and are not prepared for the fact that it is so time-consuming and that the result will also look different to what they were used to before.*"

When EMa was introduced, there was a debate within the DNB as to whether metadata in the familiar form is still needed at all. Over time, the DNB finally decided that metadata is still necessary but no longer needs to be created according to library rules in the EMa-Project. One interviewee described the discourse: "*This also led to controversial positions because colleagues who are trained accordingly naturally have a completely different relationship to the library rules. And the discussion as to whether the results of the automated processes are good enough was also conducted from the point of view of whether the bibliographic rules are being adhered to. No, they are not adhered to, but the aim of the automated processes was exclusively to find and support.*"

It is crucial to approach AI projects with a fresh and unbiased perspective, recognising the value in embracing mistakes as a part of the learning process. Consequently, top management must foster an environment that embraces experimentation and a culture of failure. It is important to recognise that such projects are iterative processes, necessitating the establishment of effective team dynamics as well. One interviewee summarised the AI experience: "*But because we had the chance to rebuild and redo everything, I would say today that I wouldn't do anything differently. We needed this experience. We worked with an external company, learned a lot, and were able to redo everything. […] You also need the mistakes to learn from them. And the good thing for us was that we had a managing director at the time who gave us this freedom, who invested money, who invested people, who gave us the freedom to tackle the issues.*"

More than 10 years ago, DNB strategically used the first steps in ML processes, which then led step by step to AI processes. This decision is currently being reviewed to determine whether a semi-automated process would be more advantageous, at least for certain applications. One interviewee commented: "*This is all justified, but experience shows that fully automated processes do not deliver the results that we would like. That's why these feedback processes with intellectual evaluation are more important.*"

Finally, the allocation of more resources towards the AI project by the DNB would have been preferable: "*[…] we should, or we should have put even more resources into it, yes. We should have invested, both in personnel and financially, as well as simply something like hardware, yes, and yes, I think that's what I would do differently.*"

5 Conclusion

The German National Library is Germany's central archive and library institution. The DNB has been using AI in various forms since 2009 to efficiently catalogue and facilitate accessibility of its extensive collection of media. The central projects

were *PETRUS, starting in 2009* and *EMa* ("subject cataloguing machine"). *PETRUS* was initiated as a software-supported solution in 2009. The development of *EMa* began in 2018. Based on the National Library of Finland's open-source toolkit *Annif*, *EMa* was launched in 2022. It is a modular tool, meaning that its composition and function can be adjusted to the special needs of the library and create metadata such as DDC subject categories and GND keywords for German and English-language media.

The use of AI became key to efficiently managing the growing number of publications necessitated by the legal obligation. This involved the development of new approaches for modifying publication formats and the establishment of interdisciplinary teams of librarians, IT specialists, and other experts. The biggest challenges encompassed limited resources, ethical issues, and resistance to change processes within the workforce. The legal requirements of copyright and the processing of copyright-protected resources made the implementation even more complex.

Despite these challenges, the projects have influenced the DNB's operational processes. The integration of *EMa* has enabled an efficient generation of metadata, thereby facilitating its accessibility to other libraries via DNB services. This has not only enhanced efficiency but also strengthened the role of the library as a protected space for learning. In addition, the use of AI offers new opportunities for the personalisation of library services and the improvement of search and access processes.

The projects have provided valuable insights, such as the importance of an interdisciplinary and autonomous approach by strengthening internal processes. Mistakes along the way should be regarded as learning opportunities, and the involvement of diverse teams is recommended. The feasibility of semi-automated processes for specific applications is being evaluated. Long-term investment in resources and further development of AI technologies are crucial to ensure that the DNB can meet the demands of the future and fulfil its social obligation as a central source of information in Germany.

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 EMa project at the German National Library

Case and project name			
EMa (subject cataloguing machine) at the National Library of Germany (DNB)			
Country	**Number of employees**	**Type of AI solution**	**Year and maturity level**
Germany	617 employees located in Frankfurt am Main and Leipzig	EMa (subject cataloguing machine) is the abbreviation for "*Erschließungsmaschine*". EMa is an automatic subject cataloguing machine that generates descriptive metadata to enrich the records in the DNB's catalogue	2009: Project start PETRUS 2019: Project start EMa 2022: EMa went in operation
Project description			
The first experiments with automated indexing started in the Petrus project in 2009. The goal was to improve library operations through process-supporting software. They implemented four main scenarios: automatic subject classification, keyword assignment, automatic linking of personal names, and parallel matching for print and online publications. In 2018, it was decided not to follow the PETRUS application any longer. In 2018, the project called cataloguing machine EMa began to rebuild the indexing system using the open-source Annif from the National Library of Finland, interdisciplinary expertise, and different in-house produced AI processes. EMa was enrolled in April 2022 and generated descriptive metadata to enrich the records in the catalogue. The metadata is available for external applications via the DNB data services and, in general, EMa-generated DDC subject categories, subject headings from GND, and DDC short numbers in the subject categories			
Need(s) behind implementation	**Actors involved**	**Challenges**	
Since 1913 there is a legal obligation that requires the acquisition of a copy of every media published in Germany, which results in the accumulation of vast amounts of data. Moreover, the German National Library Act (DNGB) came into force in 2006 New forms of publications, especially electronic ones, mean that the established procurement channels no longer work Support from the library management behind the introduction of changes brought about by ML and AI	A broad multiprofessional in-house team consisting of librarians, mathematicians, and computer linguists was involved in the process However, it is important to mention that there must be always a close cooperation between AI and librarians to achieve valuable results The in-house team was supported by a broad scope of external actors consisting of, e.g. universities and research organisations	1. Evaluating the AI-Output 2. Fear of losing the job 3. Scarce resources 4. Integration in the organisational structure 5. Labelling of AI-Outputs? 6. Ethical considerations 7. Copyright	

Results		
Organisational level	**Value created and co-created**	**Lesson learned**
The usage of AI leads to interdisciplinary work and the breaking down of specialist silos The risks in the library sector are generally manageable, as they primarily involve the quality of results and the changing nature of the librarian profession	The DNB goal is to serve as a platform where society can stay informed about various media with the help of AI DNB's role as a data supplier is closely linked to this. It provides metadata for other libraries in the academic and public library networks. This means that DNB indexes the publications, and the libraries can obtain the metadata via automated processes and import it into their catalogues	1. Need for more data specialists 2. Teams with diverse professional backgrounds 3. AI is not just a question of method and implementation 4. Debate about whether metadata in the familiar form is still needed at all and whether AI-Output is accurate 5. Projects should be approached with curiosity and a form of naivety 6. Evaluation of trade-offs between fully automated and semi-automated subject cataloguing solutions 7. More investment in resources

References

1. Deutsche Nationalbibliothek, Eine Bibliothek zwei Standorte (2025), https://www.dnb.de/DE/Benutzung/Oeffnungszeiten/oeffnungszeiten_node.html#:~:text=Die%20Deutsche%20Nationalbibliothek%20hat%20zwei,kulturellen%20oder%20fachlichen%20Veranstaltungen%20besuchen
2. Bundesministerium der Justiz (ed.), *Gesetz über die Deutsche Nationalbibliothek (DNBG)* (Bundesministerium der Justiz, Berlin, 2006)
3. Deutsche Nationalbibliothek. ORGANISATION—Beschäftigte (2024), https://www.dnb.de/DE/Ueber-uns/Organisation/organisation_node.html
4. K. Hendinger, DIGITALE ERINNERUNGSKULTUR Deutsche Nationalbibliothek: Wie KI hilft, mit Holocaust-Überlebenden zu reden, in *MDR Kulturdesk*, (2024)
5. C. Poley, U. Sandro, F. Busse, J.-H. Jacobs, M. Kähler, M. Nagelschmidt, M. Schumacher, Automatic subject cataloguing offers the possibility of classifying or verbally indexing publications that would otherwise not be indexed at all or only partially. LIBER Q. **35**, 1–29 (2025)
6. C. Schöning-Walter, PETRUS Prozessunterstützende Software für die digitale Deutsche Nationalbibliothek, in *Dialog mit Bibliotheken 2010/1*, (Deutsche Nationalbibliothek, 2010)
7. Start der Erschliessungmaschine EMa (2023), https://jahresbericht.dnb.de/Webs/jahresbericht/DE/2022/Hoehepunkte/Erschliessungsmaschine/erschliessungsmaschine_node.html
8. B. Lund, D. Agbaji, Use of Dewey decimal classification by academic libraries in the United States. Cat. Classif. Q. **56**(7), 653–661 (2018)
9. Deutsche Nationalbibliothek (ed.), *Kooperationsvereinbarung zur Gemeinsamen Normdatei* (2024)
10. Deutsche Nationalbibliothek, *Jahresbericht 2022* (Deutsche Nationalbibliothek, 2022)
11. I. Mergel, N. Edelmann, N. Haug, Co-production phases in the development and implementation of digital public services. Perspect. Public Manag. Govern. **8**, 93 (2025)

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.

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.

Anna-Lea Schumann is currently pursuing a master's degree in politics and public administration at the University of Konstanz. Her research interests focus on applying modern working practices within public administration and their implications for public value creation. In her bachelor's thesis, she explored the use and implementation of design thinking in libraries. She has gained initial practical experience through various internships in public administration, providing her with valuable insights into the operational methods and processes of the public sector.

Corinna Nitsch currently serves as the Chief Information Security Officer in the Baden-Württemberg Ministry for the Environment, Climate and Energy Sector. In both her bachelor's degree in public management and her master's degree in politics and public administration, she concentrated her studies on matters pertaining to digitalisation in public administration. In her theses and part-time roles, she addressed a range of issues and projects, including 5G expansion in municipalities, the integration of artificial intelligence in public libraries, and inequities in online information seeking. She gained experience in several roles at the Kehl Institute for Applied Research and the Cluster of Excellence at the University of Konstanz, among other institutions.