HORIZON-CL2-2021
HORIZON-CL2-2021-HERITAGE-01-02

LibrarIN [101061516]: Value Co-creation and Social Innovation for a new Generation of European Libraries

D3.2 Digital Transformation and ICT v2.0

Project Reference No	LibrarIN -101061516
Deliverable	D3.2 Digital Transformation and ICT v2.0
Work package	WP3: Thematic and empirical work in selected co-creation areas
Type	R – Document, report
Dissemination Level	PU – Public
Date	31/10/2024
Status	Final V1.0
Editor(s)	Ines Mergel (UKON), Carsten Schmidt (UKON), Anna-Lea Schumann (UKON), Justus Kühler (UKON)
Contributor(s)	Antoine Henry (ULILLE), Francesco Molinari (LC), Alberto Peralta (UAH), Ada Scupola (RUC), Barbara van der Vaart (LIBER)
Reviewer(s)	Faiz Gallouj (ULILLE)
Document description	This document includes the updated version of the deliverable D3.1 "Digital transformation and ICT" of work package 3 which illustrate the cases selected and their relevance for the LibrarIN project.

Disclaimer

The LibrarIN project is funded by the European Union under grant agreement ID 101061516. The information and views set out in this publication are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
V0.1	07/07/2024	Write up of the initial version of the report	UKON
V0.2	23/09/2024	Final draft for reviewers	UKON
V0.3	28/09/2024	Suggestions by reviewers	ULILLE
V0.4	11/10/2024	Revised draft with partners revisions	UKON
V0.5	12/10/2024	Final version for submission to EC	UKON
V1.0	31/10/2024	Final review and submission	ATC

Executive Summary

National Libraries are beginning to explore the use of AI (artificial intelligence) tools to enhance their digital transformation efforts. Previous digital innovations have included augmented and virtual reality tools, eBooks instead of physical lending, and the use of social media to engage patrons. However, AI presents new risks and opportunities for transforming library services.

Libraries play a crucial role in preserving information for future generations and democratizing knowledge access. They promote digital literacy among patrons and library staff. Despite this, the use of AI in libraries has yet to be thoroughly studied. This report aims to understand how libraries integrate AI into their services, the main drivers of this adoption, and the barriers they face.

This deliverable contains the findings of year 2 of the EU LibrarIn work package 3 task 3.1 "Digital transformation and ICT v02". The deliverable includes a short overview of the methods that were used to a) identify the relevant case studies, b) a short case illustration of each case using a shared case illustration template, and c) a summative overview of the findings that will serve as a starting point for the cross-case analysis that is planned for year 3.

Based on a shared interview guide, the partners in this task conducted case interviews with national libraries in Europe to understand their adoption and use of artificial intelligence (AI) tools to fulfil their mission.

The participating partners include:

- UKON, Ines Mergel, Anna-Lea Schumann, Justus Kühler, Carsten Schmidt, University of Konstanz, Germany (UKON)
- ULILLE, Antoine Henry, Université de Lille, France (ULILLE)
- LC, Francesco Molinari, The Lisbon Council for Economic Competitiveness ASLB, Belgium (LC)
- UAH, Alberto Peralta, Universidad da Alcala, Spain (ALCALA)
- RUC, Ada Scupola, Roskilde Universitet, Denmark (RUC)
- LIBER, Barbara van der Vaart, Stichting Liber, Netherlands (LIBER)

As a group, we collected 13 case studies on the implementation of artificial intelligence in national libraries based on 70 interviews. The case studies were extracted from expert interviews with the executive directors of the international library associations (see deliverable 01).

Using criteria derived from extensive systematic literature on digital transformation in libraries as well as expert interviews, we identified 13 National Libraries for our case studies. Each library had at least one AI project underway or developed. We conducted case interviews with project leaders, information or cataloguing experts, and users, with librarians often being the primary users of AI services.

In an exploratory interpretative research design, we conducted expert and case study interviews in the following National Libraries:

1. Koninklijke Bibliotheek van België, BE (UKON)
2. Deutsche Nationalbibliothek, DE (UKON)
3. Det Kongelige Bibliotek, DK (RUC)
4. Eesti Rahvusraamatukogu, EE (UKON)
5. Biblioteca Nacional de España, ES (ALCALA)
6. Kansalliskirjasto, FI (UKON)
7. Bibliothèque nationale de France, FR (ULILLE)
8. Koninklijke Bibliotheek, NL (LIBER)
9. Nasjonalbiblioteket, NO (RUC)
10. Kungliga biblioteket, SE (RUC)
11. Narodna in univerzitetna knjižnica, SI (LC)
12. British Library, UK (UKON)
13. Library of Congress, US (UKON)

We found that Northern European libraries are the most advanced in adopting AI and started earlier than their European counterparts to experiment with AI in cataloguing, tagging, and converting historic text into digital text documents. They are followed by the Benelux countries and scarce Southern European National libraries. This peculiarity was already pointed out to us in the preceding interviews with the executive directors of the international library associations. Nevertheless, we made every effort also to include Eastern and Southern European cases.

Next, we will present our shared research design that was in large parts similarly designed to tasks 3.5 and 3.8 in the same work package but overall focused on different tasks and technologies that National libraries use to digitally transform themselves. We will then present our 13 case illustrations and end the report with an overview of the main findings.

At this point in the project, we won't be providing recommendations as deliverable 3.3 v02 was designed as an intermediary step on the way to the presentation of the cross-case analyses (deliverable v03 in year 3).

Table of Content

List of Terms and Abbreviations

Abbreviation	Definition
AI	Artificial intelligence
ASR	Automatic Speech Recognition
BE	Belgium
BL	British Library
BNE	Biblioteca Nacional de España
BNF	Bibliothèque Nationale de France
DE	Germany
DK	Denmark
DNB	Deutsche Nationalbibliothek
EE	Estonia
EMa	Erschließungsmaschine
ES	Spain
FI	Finland
FR	France
genAI	generative AI
HTR	Handwritten Text Recognition
KB	Koninklijke Bibliotheek
KBR	Royal Library of Belgium
LC	Library of Congress
LLM	Large Language Model
ML	Machine learning
NL	Netherlands
NLP	Natural Language Processing
NO	Norway
OCR	Optical Character Recognition
RDL	Det Kongelige Bibliotek
SE	Sweden
UK	United Kingdom
US	United States of America

1 Introduction

1.1 Purpose and Scope

The purpose of task 02 of deliverable 3.1 is to provide an overview of the collected cases and illustrate the findings. This means that this is an intermediary step on the way to the full cross-case analysis of task 03 that will be submitted after year 3.

1.2 Approach to Work Package and Relation to Other Work Packages and Deliverables

Work package 3 has three distinct tasks that all focus on ways in which libraries are innovating through different means. Task 3.1 focuses on the digital transformation and the use of ICT. As previously reported, we conducted an in-depth systematic literature review and synthesized the data using a structural topic modeling approach to identify the main topics related to digital transformation in libraries. We then interviewed the executive directors of the most prominent international library associations with the goals to identify what the main issues in the field of library digital transformation is.

The results were clear: the main challenge that has not been covered in the existing literature yet was the adoption of artificial intelligence (AI). Based on the input of the interviewees, we identified the types of libraries that have the resources and interest to experiment with the use of AI and the countries the experts identified as the most innovative in the field.

The following national libraries were identified through expert interviews using snowball and intensity sampling approaches: Belgium, Denmark, Estonia, Finland, France, Germany, Norway, Spain, Sweden, Netherlands, and the United Kingdom. We added the US Library of Congress because it was continuously mentioned in the case and expert interviews. From the initial expert interviews, we extracted the case selection criteria. The case illustrations focused on two main parts, each encompassing several subtopics to provide a comprehensive overview of the AI projects in the National Libraries.

1. Koninklijke Bibliotheek van België, BE
2. Deutsche Nationalbibliothek, DE
3. Det Kongelige Bibliotek, DK
4. Eesti Rahvusraamatukogu, EE
5. Biblioteca Nacional de España, ES
6. Kansalliskirjasto, FI
7. Bibliothèque nationale de France, FR
8. Koninklijke Bibliotheek, NL
9. Nasjonalbiblioteket, NO
10. Kungliga biblioteket, SE

11. Narodna in univerzitetna knjižnica, SI
12. British Library, UK
13. Library of Congress, US

1.3 AI technologies mentioned in the cases

In the case interviews, we noticed that the interviewees, who hailed from different technological and information science backgrounds, used the technological term AI or "the program" in rather broad terms. Others were very specific and were able to identify different types of technologies that fall under the AI umbrella terminology. We were able to identify the following types of technologies: automatic-semantic meta-tagging, machine learning (ML), automatic speech recognition (ASR), national language processing (NLP), scanning analogue documents, and optical character recognition (OCR), large language models (LLMs), and generative AI (genAI).

In the third version of this deliverable, we will analyse each of the technologies in more detail and aim to distinguish their use for different types of library tasks.

1.4 Methodology and Structure of the Deliverable

Based on the interviews with the executive directors of the international library associations, we developed a shared interview guide, that is included in appendix I.

The interview guide is divided into the following subsections:

1. Need(s) behind implementation
2. Actors involved
3. Challenges
4. Results

The interviewees were selected based on their involvement in the development, deployment, implementation, and use of the AI tools in their national library. Interviewees, therefore, included staff members from the IT department, the innovation or strategy departments, as well as users of the technology (internally or externally).

The interviews were recorded and transcribed verbatim to be able to analyse the statements of the participants.

For this report, we extracted the most poignant statements from the case interviews using a shared case illustration template. The goal was not to analyse and compare the insights at this stage but to illustrate the content of the case interviews.

1.5 Project description topics illustrated for each case

1.5.1 Need(s) Behind Implementation

This subtopic explores the specific needs and motivations driving the implementation of AI in the National Libraries. It explores the challenges and limitations of current practices that AI aims to address, such as improving efficiency, enhancing accuracy, reducing costs, and enabling more sophisticated data analysis. It also considers broader societal goals such as increasing accessibility, supporting digital literacy, and maintaining relevance in a rapidly evolving technological landscape.

1.5.2 Actors Involved

This section identifies and describes the various stakeholders participating in the AI projects. It includes:

- **Internal stakeholders**: Library staff, including librarians as specialists and future internal users, IT specialists, and project managers directly involved in the project's planning, development, and implementation.
- **External stakeholders**: Technology providers, consultants, and experts in artificial intelligence who offer their knowledge and tools to support the project.
- **End users**: Patrons and library users who will interact with the AI-enhanced services, providing feedback and helping to shape the final product.

1.5.3 Challenges

Here, the case study outlines the key challenges faced during the implementation of AI. These could include technical hurdles, such as data integration and system interoperability, as well as organisational and cultural challenges, such as resistance to change, lack of trust in AI, and the need for training and skill development among staff. It also considers external challenges like regulatory compliance and ethical considerations when using AI.

1.5.4 Results

1.5.4.1 Organisational Level

This subtopic examines the impact of the AI projects on the organisation as a whole. It looks at how adopting AI has influenced internal processes, workflows, and decision-making. It assesses changes in staff roles and responsibilities and how AI has been integrated into the library's strategic planning and operational frameworks.

1.5.4.2 *Value Created and/or Co-Created*

This section evaluates the tangible and intangible benefits generated by the AI projects. It includes:

- **Direct value**: Improvements in efficiency, accuracy, and cost-effectiveness of library operations and processes.
- **Indirect value**: Enhanced user experience, increased patron engagement, and better support for research and education.
- **Co-created value**: The collaborative benefits realised through partnerships with technology providers and the involvement of internal users in the development process leads to solutions that are more aligned with user needs and expectations.

1.5.5 Lessons Learned

This final subtopic summarises the insights and takeaways from the AI implementation. It highlights best practices, successful strategies, and areas for improvement in future iterations identified throughout the project. It also offers recommendations for other libraries considering similar initiatives, focusing on effectively managing change, engaging stakeholders, and ensuring the ethical and responsible use of AI.

In the following, we display a summary of each case following the above-mentioned case illustration items.

2 Case Illustrations

2.1 Overview / Summary Cases

Cases and project names			
• Retrocatalography at the Royal Library of Belgium (KBR) • The Digitalization of postcards collection project at the Royal Danish Library (RDL) • Automatic Indexing and enquiry answering service Ask Us at the National Library of Estonia • Annif and FintoAI at the National Library of Finland • Gallica at the National Library of France (BnF) • Petrus and EMa at the National Library of Germany (DNB) • Automatic Metadating project: Retrotool at the Royal Library of the Netherlands (KB) • National Library of Norway/Maken Project • Narodna in univerzitetna knjižnica, NUK - National and University Library, NUL • ETSO Project: Stylometry Applied to Golden Age Theater at the National Library of Spain (BNE) • Swedish National Library/ digitized postcards project (NLS) • The project is part of the Living with Machines project of The British Library of the United Kingdom (BL) • Exploring Computational Description at the Library of Congress (LC)			
Countries	*Number of employees*	*Type of AI solutions*	*Year and Maturity level*
• Belgium • Denmark • Estonia • Finland • France • Germany • Netherlands • Norway • Slovenia • Spain • Sweden • United Kingdom • United States	• **BE**: Approx. 250 employees • **DK**: 650 • **EE:** 250 • **FIN:** More than 230 employees with different backgrounds • **FR:** 2128 • **DE:** 617 employees located in Frankfurt am Main and Leipzig • **NL:**500	• **BE**: Cataloguing AI solution with the Microsoft Power Platform • **DK**: OCR/HRT recognition/ Google Vision • **EE:** Cataloguing AI solution • **FI:** Open-source toolkit called Annif for automated subject indexing and classification • **FR:** OCR, HTR and image/data mining • **DE:** Petrus is a process-supporting software that uses machine-based processes and AI for subject cataloguing. EMa ("Erschließungsmaschine") is a cataloging machine that generates	• **BE**: Application Go-live, 10/2022; Legal Deposit Updates, Spring 2024 • **DK**: Started in 2018/On the way to production • **EE:** 2023 / Prototype and major milestones accomplished • **FI:** Start of the Finto project, 2014; first prototype Annif, 2017; developing Annif, 2018; launch Finto AI, 2020 • **FR:** 1997, well underway/major milestones accomplished • **DE: 2009**: Project start Petrus; **2012:** Start of the implementation of Petrus; **2018:** Project start EMa; **2022:** Implementation of EMa

	• **NO:** 420 • **SI:** 134 • **ES:** 364 • **SE:** 350 • **UK: 1700 (2011)** • **US:** 3,172 permanent employees	descriptive metadata to enrich the records in the DNB's catalogue • **NL:** Creation of an AI tool for automatic meta-tagging of books • **NO:** recommendations system based on the content of books and images • **SI:** File format conversion; OCR and HTR • **ES:** Transkribus to automatically transcribe and modernize old prints and manuscripts with a high degree of accuracy. • **SE:** Swe-CLIP 2M • **UK:** Natural Language Processing/ OCR • **US:** Using AI and ML to create descriptive metadata for e-books that have not yet been catalogued	• **NL:** Starting in mid-2023, the Retrotool was developed and tested, and officially implemented in September 2024. • **NO:** Experimental stage/Demo launched on website • **SI:** 2020-2024 - MMA • **ES:** 2017, Implemented and alive • **SE:** Experimentation stage; a demo has been launched on the web • **UK:** Starting year 2018/2019 • **US:** Project start, 08/2022

Project descriptions

Royal Library of Belgium (KBR):
KBR developed a cataloguing AI solution in 10 days with a Belgian company named Inetum-Realdolmen. The project, called Retrocatalography, involved developing an application with the Microsoft Power Platform that can gather information from a book's cover, archive it, verify it, take a picture of it, and then analyse it using AI to check whether everything has been detected accurately. The Power Platform allows for low-code application development using Microsoft's pre-developed components.

Royal Danish Library (RDL):
Digitalization of approx. 350,000 topographical postcards from the period 1887 to the 1990s from all parts of Denmark.

National Library of Estonia:
The goal is to develop a prototype utilising machine learning and AI to automate publication content analysis and tagging.

National Library of Finland:
The national library has created an open-source Annif toolkit for automated subject indexing and classification. Annif is multilingual, works independently of the indexing vocabulary, is modular, and integrates different text algorithms such as Maui and a neural network based on TesnorFlow. Annif can be integrated into the Finto AI web user interface, an in-house-produced metadata management system. To evaluate Annif, they use manually indexed corpora, such as master and doctoral theses from the University Jyväskylä. Additionally, humans evaluate the results of the automated subject indexing. Finally, in May 2020, the national library launched Finto AI, which introduces automatically subject indexing into information systems and can be used for Finnish, Swedish, and English subjects. Finto AI is, for example, already implemented at the University of Jyväskylä, where students submit their master's thesis using the tool and receive suggestions from Annif. The students can use or discard the suggestions before a final check by librarians.

National Library of France (BnF):
Gallica is the digital library of the BNF. To enable people to access BnF's collections directly online, Gallica is engaged in digitization efforts. With over ten million documents and 60,000 daily users, it is among the largest digital libraries worldwide. Artificial intelligence has been utilized through technologies like Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) to enhance document readability and streamline search processes.

National Library of Germany (DNB):
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. Over the years, challenges arose, including personnel turnover and the need for modular solutions. In addition, DNB worked closely with an external company that developed the software for them. In 2018, the company decided not to develop the software further for the DNB.
In 2018, the project called cataloguing machine EMa ("*Erschließungsmaschine*") 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. With the EMa application, it is now possible to open up further data records. 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 category of medicine for media works in German and English.

Royal Library of the Netherland (KB):
The 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 organizes 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.

National Library of Norway:
Maken is a discovery engine to find similar books and similar images. Based on the pixels of the images, Maken finds related images. Based on the text of a book itself, Maken finds books that have some resemblance to each other

National Library of Slovenia:
Use of AI to significantly enhance the accessibility of digitised library content for blind and visually impaired communities through various SOTA technologies (mostly FLOSS).

National Library of Spain (BNE):
Thanks to the AI-powered tool Transkribus (READ-COOP, 2024) and its Handwritten Text Recognition (HTR) techniques, Dr. Cuellar and the ETSO team have trained three models, which are being used and refined by the research community, capable of automatically transcribing and modernizing these documents with a high degree of accuracy: approximately 97% accuracy for printed texts and 91% for manuscripts. Through these models, ETSO has processed around 1,800+ theatrical works contained in prints and manuscripts from numerous libraries, archives, and other digitized sources. The resulting transcriptions are now part of the CETSO Corpus and the TEXORO search engine, both part of the ETSO project. Moreover, they possess sufficient quality to undergo stylometric analysis, which reveals interesting authorship attributions.

National Library of Sweden (NLS):
Searching in postcard collections of ca. 17,000 digitized postcards

National Library of Great Britain (BL):
The project was initiated to explore the impact of AI and machine learning on historical research, focusing on the effects of mechanization in the long 19th century. It involved collaboration between the Alan Turing Institute, partner universities, and the British Library. Challenges included interdisciplinary collaboration, technical expertise and staff rotation, copyright and access issues, and project management. Despite these challenges, the project aimed to address the evolving landscape of historical research in the digital age.
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 digitization 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.

Library of Congress (LC):
The LC Labs and the Acquisitions and Bibliographic Access Directorate are currently working on a project called Exploring Computational Description to determine the feasibility of using AI and ML to create descriptive metadata for e-books that have not yet been catalogued. The project began in August 2022 and involves collaboration with an external vendor, Digirati. Digirati tested five different models for machine learning and used two workflows for assisted cataloguing, creating a subject assistant suggestion prototype. In September 2023, the first part of the project, called Toward Piloting Computational Description, was concluded, and the team has now moved on to the second part, which focuses on prototyping the use of machine learning in greater detail.

Need(s) behind implementation	*Actors involved*	*Challenges*
BE: • Some books are still only described on paper cards, making it impossible for patrons to find these records in the library's online catalogue. The COVID-19 pandemic has highlighted the need to digitize these records. • There was a need to promote collaboration between departments within KBR to explore AI and develop an AI solution. **DK:** • The new strategy for wider dissemination of The Royal Danish library's collections – both in relation to target groups and accessibility; the user demands; Need to generate as much metadata as possible **EE:** • Tagging items is a labour-intensive task that demands significant resources. This process relies heavily on the knowledge and	**BE**: • Close collaboration with the Belgian company Inetum-Realdolmen. Collaboration with seven to eight volunteers who scan and validate records. In addition, an external company recently supported the volunteers. The library management finally supports exploring an AI-based solution. **DK**: • Curators in Special Collections, library management, digitalization department, Google Vision **EE:** • Head of development at the NLE Internal and external project managers, Experts with experience in AI, data management personnel, development team members, customers and vendors. **FI:** • The Annif project team developed the first prototype in 2017 and tested a use case at the University of Jyväskylä. The library management and the public showed interest in it, leading to further work on the project.	**BE**: • The main challenges are a staff shortage and a rising number of records. The library staff is finding it difficult to integrate AI into their daily work due to time constraints. Technical challenges include, for example, situations where the AI model is not functioning optimally and needs to be retrained automatically. Lastly, the budget for AI projects at KBR is a challenge. **DK**: • Copyright, GDPR, Technological Challenges, Data Bias, Quality of the generated metadata, Lack of time to experiment with technologies, accuracy of metadata provided to library users, high cost associated with generating sufficient metadata. **EE:** • It is vital to keep the people who are not working on a day-to-day basis in the project informed and involved. • It can be challenging to grasp what AI is doing in the library and what the librarian's role will

judgments of individual taggers. Additionally, the manual nature of tagging can be time-consuming and costly.

FI:

- Need for automated subject indexing arises due to the growing volume of digitally collected data and the extensive digitization of paper archives.
- Allows for indexing library collections where manual indexing is impractical
- Need an in-house solution due to the lack of AI support for multiple Finnish languages and dialects
- Need an in-house solution due to the lack of control over the AI system from commercial providers and the difficulty of integrating them into existing systems

FR:

- Information retrieval, cataloguing, ease the research inside the collections.
- Research of information from images or digitize documents (among the 10 million of the digital library)

DE:

- The law requires the acquisition of a copy of every media published in Germany, which results in the accumulation of vast amounts of data.

- The Annif project team collaborates closely with external stakeholders to develop and enhance Annif by using, testing, and modifying the tool's code. The Leibniz Information Centre for Economics (ZBW) is an important collaboration partner, for example, in technical development.
- Annif is implemented not only in the National Library of Finland but also in other libraries, such as the National Library of Germany, universities like the University of Jyväskylä, organizations like ZBW, and companies like the Finnish Broadcasting Company Yle.
- At the University of Jyväskylä, for example, students, librarians, and Annif work together to review Master's theses. Students upload their theses to the Finto AI tool, review the suggestions, and then a librarian gives it a final check. This process in Annif is a semi-automated indexing system.
- The Annif project team welcomes feedback and suggestions for improvement, for example, on the GitHub platform, where external partners can report bugs and make pull requests, or in the user forum called Annif-users. Additionally, they provide a hands-on tutorial with the ZBW.

FR:

- Mostly internal (IT, Collections and Cooperation departments), external partners like BNUS, INHA, INRIA or technical partners (i.e. Mistral)

DE:

- The "Acquisitions and Cataloguing Department" is responsible for developing the collection and producing cataloguing data, i.e., vulgo cataloguing. This department includes the "Automated Cataloguing and Online Publications" department.

be, especially with things like automatic cataloguing or coding examples.

FI:

- **First Challenge:** Over time, the model has become more accurate with fewer mistakes. However, there are concerns about hidden mistakes in assigning subject headings. Therefore, it's important for librarians to remain skilled in cataloguing and indexing subject headings.
- **Second Challenge:** Librarians require sufficient time to use the tool effectively to achieve satisfactory results.
- **Third Challenge:** Some librarians are concerned that the tools may replace human workers and lead to layoffs.
- **Fourth Challenge:** On the macro level, the development of AIs by big commercial companies is a challenge and even a dangerous undertaking because there is little transparency about what they are doing. Additionally, libraries could use AI to benefit society and counter bias and ethical issues.
- **Fifth Challenge:** There are challenges in using AI and dealing with copyright. The legal landscape is evolving, making it a grey area.
- Nevertheless, this approach takes a lot of resources, such as library staff with technical skills. That's why the library goes with the strategy to invest in their employees and their infrastructure.

FR:

- Moving from prototyping to implement and launch live
- Secure funding to achieve them

• New forms of publications 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. **NL:** • 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. **NO:** • Increasing difficulty in discovering the digital material; increasing amount of material; potential lack or insufficient metadata **SI:** • Compliance with WGAC • Better (re)usability of outputs **ES:** • To automatically transcribe the Spanish Golden Age documents for stylometric analyses able to establish trusted authorship attributions. **SE:** • Very limited access that the users had to this kind of collection; sometimes even the awareness of its existence. **UK:** • The project was initiated to explore how AI and machine	• The „Cross-cutting Matters of Use and Inventory Management" department is involved in experimenting and working with various types of innovation, from technical to social. • Staff with different professional backgrounds, such as computer linguists, mathematicians, and experienced librarians, are involved in the AI project. The project members have a certain affinity with the topic of AI. 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, such as universities, other libraries, research institutions, consultants, suppliers and start-ups. • Close cooperation between AI and librarians is required. **NL:** • KB Management • Collection data department: • Head of Department, two project leaders • Four cataloguers who tested and currently work with the tool • A technical company who developed the tool **NO:** • Top management, AI-lab, users, external digital agency **SI:** • The Library's research group • Blind/visually impaired users • Mobile/print disabled users **ES:**	**DE:** • Evaluating the information libraries receive from AI applications is crucial, as AI tools are neither complete nor always neutral. • The fear of cataloguers to lose their jobs • Limited availability of personal and material resources • Integrating AI into the organizational structure • Is uniform labelling of AI-generated output necessary? • Ethical considerations related to the use of AI • Copyright • Need of an incredible amount of energy **NL:** • Ethical challenges (use of LLM created by Microsoft is quite opaque).and fear by employees that AI would make their work obsolete. **NO:** • GDPR and privacy issues, Data Protection Impact Assessment, consent for the interviews **SI:** • Related to implementation • Related to take-up and adoption **ES:** • First Challenge: Automatic transcription using Optical Character Recognition (OCR) faces limitations due to historical typefaces and handwritten manuscripts from the XV-XVI centuries, making manual transcription more efficient. • Second Challenge: Modernizing spelling in the works is essential for meaningful comparison

learning would impact historical research, with a focus on the effects of mechanization during the 19th century. • 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 digitization on user engagement **US:** • In order to fulfil LC's goal of connecting all Americans to the library, LC needs to investigate new approaches that could bring significant changes to its practices. LC Labs are tasked with experimenting with new methods, approaches, and technologies, and collaborating with external stakeholders such as universities and research initiatives. Additionally, there is support from the library management and funding from the National Digital Trust Fund to experiment with AI and ML.	• Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid) • Their teams. • The research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona). • 200+ researchers, students, experts, and collectors from 20+ institutions and private collections. **SE:** • KBlab developers/Librarians/Archivist/designers; library users; copyright agency **UK:** • It involved collaboration between the Alan Turing Institute, partner universities, and the British Library. • Actors involved include internal teams like the Digital Research Team and external partners like the Alan Turing Institute. • Involving over 5,500 volunteers, the project engaged the public through crowdsourcing tasks. Collaboration spanned disciplines, including data scientists, historians, and research software engineers, with members joining gradually between 2018 and 2019. **US:** • The LC Labs serve as an experimental space for testing different projects and are part of the Digital Strategy Directorate. The Acquisitions and Bibliographic Access Directorate primarily catalogues books and e-books and is part of the Discovery and Preservation Services. These two departments collaborate and work with external vendors such as Digirati or with research institutions and academic partners. Furthermore, the Planning Framework, developed by the LC	with other modernized texts, even if OCR achieves precise transcription. • Third Challenge: Despite digitization, the vast quantity of Golden Age theater works remains inaccessible for analysis due to lack of transcription. • Fourth Challenge: In 2020, Transkribus lacked pre-existing models for recognizing Spanish. The creation of the initial recognition model, Spanish Golden Age Prints 1.0, allowed researchers to transcribe digitized manuscripts. A refined version of the model, incorporating modernized spelling, is also available via the READ-COOP platform. • Fifth Challenge: Despite having tools to recognize print from old works, errors persist due to factors like challenging handwriting, crossed-out sections, document dirtiness, translucency, and the AI tools' training data limitations. • Sixth Challenge: Manuscripts pose additional complexities due to their handwritten nature, resulting in more errors during automatic transcriptions. • Seventh Challenge: Stylometry can reveal authorship patterns in printed works and manuscripts, but caution is needed due to recognition tool errors and stylometric connections to specific authors, as seen in the case of Lope de Vega. **SE**: • Copyright, technical integration and accessibility challenges with the rest of the library system; the lack of resources and lack of strategic leadership to put the demo service into full production

	Labs, helps to connect library staff with patrons and the library community.	**UK:** • didn't 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. **US:** • Should patrons, mostly other libraries, be informed if a record was generated using AI and whether the record should contain details about the confidence level of the AI-generated record? • Can we use machine learning and AI to process content that is copyrighted? What will the machine do with the content after the modelling is complete? • What level of accuracy is considered acceptable for an AI-generated output? To ensure accuracy, a cataloguer is still required to identify possible mistakes and input the information correctly into the record. This is closely related to creating quality standards and policies specifically for AI. • AI struggles to identify the title and author of an e-book due to the varying design elements and typography. Additionally, predicting e-book content and identifying non-controlled vocabulary words remain challenging for AI. Additionally, the predicted words and terms may not always be useful or informative and cannot be corrected.

		• How should we deal with harmful and disrespectful language in books and cultural heritage? This can be particularly problematic with old books that contain such language and when these are used to train ML models. • Training AI requires large volumes of data, including born-digital materials, which can be challenging. • There is a risk of ineffective ML and AI solutions, which is connected to the fear of investing resources into solutions that may not work or could worsen existing issues.
Results		
Organizational Level	***Value created and co-created***	***Lesson learned***
BE: • AI within the KBR revolutionizes how librarians work, making processes more efficient and allowing librarians to focus on intellectual work. However, not everyone in the library currently recognizes the value that AI brings. **DK:** • The value of AI at organizational level includes production of more metadata unlocking of data locked in handwriting text; better and increase precision of search services; potential for large scale OCR or handwriting text	**BE**: • The KBR now has a clearer overview of its collection, making it more user-friendly for patrons with specific requests. In addition, patrons can easily identify if a record in the catalogue was generated by AI, as this information is transparently displayed in the remarks section. Moreover, the AI solution makes it possible to catalogue more books efficiently and effortlessly, and the quality within the library will improve by the end. **DK**: • The cultural value of making existing collections available to the public; democratic value of increasing access to cultural heritage; the value of providing new search services and cleaner metadata to the public; potential of using data	**BE**: • The key competencies for excelling in AI include motivation and a strong interest in the field, as well as the necessary education to understand the concepts. **DK**: • Start in small scale and later scale it up to a bigger quantity of data; • Make sure to jump on the AI wagon if and how it is the right thing to do; employees' competences, e.g. employee in library collections need computation skills and vice versa computational experts need an understanding of the collections; awareness of the ethical, sustainability and bias issues in using AI **EE:**

recognition; freeing resources for other tasks; increasing cost-efficiency in generating metadata; better possibility to understand own collection data because it recognizes the value that AI brings.

EE:

- The project is coordinated by the Head of Library Services Center and his team. It is supported by the Estonian Government and part of the overarching AI strategy.

FI:

- The Annif project team first presented its ideas about Annif at a Library Networking Service meeting for management and the library colleges. They also conducted presentations about Annif's ideas outside the library. Additionally, the project team established a first use case with the University of Jyväskylä, demonstrating its effectiveness to management.
- An interviewee stated that they are experimenting with AI tools but have not fully integrated them organization-wide, and it is not yet implemented in the general process. Annif is performing its function, but it operates behind the scenes and still requires human assistance to yield valuable results.

collection in new ways and different stakeholder groups then wider public.

EE:

- The users and patrons are described as very open-minded. They are eager to use the new technologies. They are very supportive and provide valuable feedback on improving and moving forward.

FI:

- **First benefit:** AI needs custodianship and oversight, and the National Library of Finland plays a critical role in society in this regard.
- **Second benefit:** There is positive feedback from users, for example, in the use case, where students submit their theses to the repository, which in return encourages the library to work further on it.

FR:

- New services for users, new services for librarians and for partners (like other libraries), change at the process level to ensure a smoother experience as well for users and for employees.

DE:

- The DNB considers it its duty to uphold democratic values that are closely tied to thoughtful consideration of information sources. However, AI outputs are not always neutral and may threaten these values. Therefore, the DNB aims to serve as a platform where society can stay informed about various developments in AI. In addition, DNB is a reliable source of information for AI tools. 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

- It was reported that the employees of the NLE have a low level of trust in AI solutions. Therefore, keeping them informed about ongoing developments is crucial to build confidence and transparency in the process. Furthermore, it is important to do it step by step:
- Detailed Analysis
- Prototyping
- Production (This phase should only proceed when the AI solution works reliably and effectively)

FI:

- Automated subject indexing requires skills and competencies in manual subject indexing and description. For this reason, they advise that beginners or non-professionals avoid using the tool. In addition, some librarian staff are worried that the AI tool is seen as a way to cut off staff and provide quick but inadequate input, which could lead to incorrect information in the library database. Therefore, one interviewee recommended using the tool carefully and thoughtfully to improve metadata but urged more cautious usage. Furthermore, they already started embracing open access, data, and software. Progress has been slow, but there has been a gradual shift toward these principles.

FR:

- Importance of gathering people from various services to be part of AI projects to make sure they will contribute to it. Improve AI's literacy in the library.

DE:

- Need of more IT developers and specialists

- However, the National Library of Finland decided against developing an AI department within the library because they still believe that AI should be integrated across the organization and understand its basics from everyone in the library.
- There are lots of benefits of an open-source, in-house developed automated subject indexing tool.
- Automated subject indexing can assist manual indexing through semi-automated indexing, where Annif suggests subjects for a new record that is verified manually. Additionally, they test Annif for fully automated indexing. It minimizes the need for librarians to check and add subject indexing.
- In the future, the tool will become more communicative and, therefore, faster. This gives further potential to work with stakeholders together.
- An in-house solution benefits from the huge amount of knowledge in-house, and staff can immediately identify mistakes in the code or the automated record produced.

FR:

- A specific department is in charge of Gallica, this is the coordination department. The project has a sponsor at the political level and is

metadata via automated processes and import it into their catalogues.

- Furthermore, in our changing society, the library is increasingly becoming a protected space that is not only a place of learning but also a place to meet and linger. Libraries will remain vital in society, and AI adoption enhances accessibility, search functionalities, and personalized library services.
- The DNB staff's attitude has changed to some extent following the introduction of AI-supported methods.

NL:

- 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.

NO:

- Democratic value enacted by the easier access to digital books and images for the library users, contribution to the library strategy of experimenting with AI

SI:

- AI literacy
- Collaboration
- Focus on needs
- Designing
- Experimenting

ES:

- The ETSO project is highly collaborative at various levels: design, implementation, analysis of results, integration of stakeholders including experts, scholars from different disciplines, students,

- Teams with diverse professional backgrounds
- AI is not just a question of method and implementation
- Debate about whether metadata in the familiar form is still needed at all and whether the results of the automated processes are good enough
- It is crucial to approach ML and AI projects with naivety and impartiality and allow yourself to make mistakes.
- More than 10 years ago, DNB strategically used fully automated AI processes. This decision is currently being reviewed to determine whether a semi-structured automated process would be more advantageous, at least for certain applications.
- DNB should have invested even more in resources
- Experiences in the projects were valuable for further development of ML and AI in the DNB

NL:

- The focus of the innovation process remains on expanding and improving the tool, with feedback from cataloguers and other staff incorporated at various stages, and the process is progressing well.
- KB management has identified patron engagement in innovation projects in general as an area for improvement and further exploration.

NO:

- Importance of involving users in the process and the importance of understanding digital change in the libraries

SI:

an important project for the BnF related to AI.

DE:

- 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.
- 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. Ethical considerations, such as generating misleading or harmful information, have not been extensively explored.
- Five years after its inception, the Petrus project saw the initial outcomes implemented in daily library work. Four years after its inception, the EMa was implemented.

NL:

- Department of Collection Data

NO:

- User involvement and digital innovation/ transformation at organizational level

SI:

- Library's research group
- Increased user base, Reputational factors, Automation vs manual handling, Importance of FLOSS

associations, collectors, and the BNE. In addition, it emerged from the personal initiative of Dr. Cuéllar, who put together the large team involved in this co-creation effort.

- According to the evidence collected about the project, the major outcomes from their collaborative effort are:
  - The most advertised outcomes of the ETSO project are the attributions of La francesa Laura and Mujeres y criados to Lope de Vega. However, other works have been attributed to their right playwrights, such as La monja alférez to Juan Ruiz de Alarcón. Without automatic transcription, it would have been practically impossible to discover those works within the vast amount of Golden Age theater documents that exist. Therefore, these processes provide researchers with invaluable clues, which are afterwards investigated through traditional philological tools (ETSO, 2023).
  - ETSO has developed three models for automatic transcription using the Transkribus tool and its AI system. Additionally, the project has created a corpus of transcribed works, CETSO, that is being used by researchers to develop their investigations. Lastly, the TEXORO tool, is a search engine that allows exploring which works contain specific words or phrases. Currently, TEXORO contains the same texts as CETSO, but the intention is to expand in the future to include other types of texts beyond the

- It's not only about technology, but also how it is deployed. The challenge lies in the alignment between human, technological and organisational factors.

ES:

- The key lessons learned from the ETSO project, particularly regarding the use of AI in this collaborative initiative are:
  - AI Facilitates Large-Scale Projects but Requires Initial Human Input. The ETSO project has shown that AI, particularly through tools like Transkribus and Stylo, can dramatically accelerate the transcription of historical texts. However, the process still requires substantial initial human input to train the AI models to achieve high accuracy (up to 97% for printed texts). Human researchers remain critical for correcting errors and improving the AI models.
  - AI-Driven Collaboration Across Disciplines. The project involved collaboration between researchers, libraries, and digital service providers. AI tools enabled a shared platform where contributions from various scholars (philologists, historians, and technologists) could be pooled to enhance the transcription and analysis process. This demonstrates that AI can be a catalyst for cross-disciplinary cooperation.
  - Challenges in Historical Data Handling. The ETSO project encountered several

solutions, Time and resource constraints

ES:

- The work done by the ETSO project team did not produce any reported effect at the organizational level of the National Library of Spain.
- However, and since the BNE plays a pivotal role in understanding cultural phenomena, ETSO helps successfully achieve this role.
- By safeguarding the world's most significant collection of manuscripts and printed materials related to Golden Age theater, the BNE provides insights into the theater of that era. ETSO complements the efforts of the BNE and contributes to them revealing the keys to this theater's historical context and its status as a paramount cultural heritage. It also highlights its relevance in contemporary stages, scholarly studies, and the role of new technologies in dissemination, research, and even performance.
- In the words of the former BNE director, Ms. Ana Santos Aramburu, "we are aware that the digitization of culture serves to create wealth and has economic value in addition to its cultural significance" (Muñoz, 2018).
- In other words, the ETSO project exemplifies how AI can support theatrical domain, and it has a separate identity within the project.
    - Thanks to AI, the ETSO project significantly reduces the workload for researchers by mechanically transcribing extensive corpora of easily readable documents. This is especially relevant to analyse the hundreds of printed and theatrical manuscripts from the Spanish Golden Age and beyond that have remained forgotten and neglected, waiting to be worked on by philologists. Now, they can use texts that, if transcribed through traditional means, would have required a tremendous amount of effort.
    - For now, trained human researchers can still transcribe this type of document more accurately, albeit at a much slower pace. The work undertaken by ETSO would have taken several years for a large group of philologists. However, thanks to AI and its training, each work is transcribed in a matter of minutes. Digital Humanities provides invaluable assistance to researchers of Golden Age theater. These processes and tools, far from replacing the researchers, facilitate and propel their work to unprecedented levels.

SE:

- The democratic value of accessibility to a postcard collection

UK:

challenges, such as the limitations of Optical Character Recognition (OCR) for older typefaces and manuscripts, and the need to modernize spelling. These challenges highlight that AI must be adapted to specific historical and linguistic contexts to be effective, particularly when dealing with non-modern languages and formats.

- AI as a Tool for Discovery. AI has been instrumental in discovering previously unattributed works, such as the attribution of La francesa Laura and Mujeres y criados to Lope de Vega. This shows AI's potential

SE:

- Develop internal competences to use AI responsibly and ethically; AI expertise gets integrated within the library and takes into consideration some of the informational challenges of large scale collections; building AI on a contextual understanding of libraries as a particular type of organisation with specific challenges; develop AI where it can add value to the library

UK:

- 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. These initiatives drive innovation, enhance public engagement, improve

public value creation, especially in cultural heritage. By transcribing vast amounts of manuscripts and printed works, AI made these historical texts accessible not just to academics but also to the public. This aligns with the mission of institutions like the BNE, highlighting how AI can serve broader societal and educational purposes by preserving and democratizing access to cultural assets. **SE:** • AI models relieve library employees and free resources for other jobs; KBlab's contribution to library innovation processes **UK:** • 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. **US:** • AI can provide extractive or abstractive summaries, which can assist cataloguers in	• Encompasses access and inclusion, education and awareness, community engagement, cultural enrichment, and fostering innovation. Initiatives such as exhibitions, crowdsourcing, workshops, and innovation **US:** • LC Labs conducts public affairs and presentations within the LC, and the LC has to report to Congress about its progress in experimenting with and implementing ML and AI. However, LC Labs lacks the resources to educate more staff about ML and AI technologies in the LC. In addition, the interviewee does not think AI is changing the mission and role of the library. Moreover, the expertise of the cataloguers is even more valuable in identifying wrongly predicted content and errors. It always requires librarians with the necessary skills to comprehend and apply cataloguing principles effectively. Nevertheless, library staff need at least a basic understanding of what ML and AI mean and how they work. • After the records are completed, they are sent to OCLC and become accessible in the LC catalogue and for other libraries. However, staff capacity and time constraints may delay the project's implementation until around spring 2025. • Adopting AI in libraries is seen as a way to create public value. However, one interviewee expressed concerns about the limited availability of library materials due to under-resourcing. This makes it difficult to connect with the public effectively.	collections access, build staff capacity, and align with strategic library goals. **US:** • When the LC had the opportunity to restart the project, it would provide the AI application with more content to improve its accuracy. However, in the current project phase, they have increased the number of records to 100,000 e-books, hoping this will result in a more precise model. • Staff needs to better understand machine learning. This basic knowledge can be acquired, for example, through conversations and workshops.

	understanding the essence of a book without having to go through the entire book. Additionally, AI can take care of routine tasks, such as identifying information on the title page, freeing up librarians to focus on intellectual tasks. This can lead to more efficient use of time and resources. Many library staff members are concerned that AI is replacing their jobs. However, this is not necessarily the case. In the short term, AI creates more work because someone has to ensure it is functioning correctly, train and revise it, and continually monitor it. This leaves library staff more time to focus on intellectual tasks such as reviewing the model or correcting records.		

2.2 Royal Library of Belgium (KBR)

Case and project name			
Retrocatalography at the Royal Library of Belgium (KBR)			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Belgium	Approx. 250 employees at KBR	Cataloguing AI solution with the Microsoft Power Platform	Application Go-live: October 2022 Legal Deposit Updates (Development): Spring 2024

Project description
KBR developed a cataloguing AI solution in 10 days with a Belgian company named Inetum-Realdolmen in 2022 and then started to feed the tool with data until 2024. The project, called Retrocatalography, involved developing an application with the Microsoft Power Platform that can gather information from a book's cover, archive it, verify it, take a picture of it, and then analyse it using AI to check whether everything has been detected accurately. The Power Platform allows for low-code application development using Microsoft's pre-developed components.

Need(s) behind implementation	Actors involved	Challenges
Some books are still only described on paper cards, making it impossible for patrons to find these records in the library's online catalogue. The COVID-19 pandemic has highlighted the need to digitize these records. There was a need to promote collaboration between departments within KBR to explore AI and develop an AI solution.	Close collaboration with the Belgian software company Inetum-Realdolmen. Collaboration with seven to eight volunteers who scan and validate records. In addition, an external company recently supported the volunteers. The library management supports exploring an AI-based solution.	The main challenges are a staff shortage and a rising number of records. The library staff is finding it difficult to integrate AI into their daily work due to time constraints. Technical challenges include, for example, situations where the AI model is not functioning optimally and needs to be retrained automatically. Lastly, the budget for AI projects at KBR is a challenge.

Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
AI within the KBR revolutionizes how librarians work, making processes more efficient and allowing librarians to focus on intellectual work. However, not everyone in the library currently recognizes the value that AI brings.	The KBR now has a clearer overview of its collection, making it more user-friendly for patrons with specific requests. In addition, patrons can easily identify if a record in the catalogue was generated by AI, as this information is transparently displayed in the remarks section. Moreover, the AI solution makes it possible to catalogue more books efficiently and effortlessly, and the quality within the library will improve by the end.	The key competencies for excelling in AI development include motivation and a strong interest in the field, as well as the necessary education to understand the concepts.

2.3 Royal Danish Library (RDL)

Case and project name			
Royal Danish Library/ The Digitalization of postcards collection project			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Denmark	650	OCR/HRT recognition/ Google Vision	Started in 2018/On the way to production

Project description		
Digitalization of approx. 350,000 topographical postcards from the period 1887 to the 1990s from all parts of Denmark.		
Need(s) behind implementation	**Actors involved**	**Challenges**
The new strategy for wider dissemination of The Royal Danish library's collections – both in relation to target groups and accessibility; the user demands; Need to generate as many metadata as possible	Curators in Special Collections, library management, digitalization department, Google Vision	Copyright, GDPR, Technological Challenges, Data Bias, Quality of the generated metadata, Lack of time to experiment with technologies, accuracy of metadata provided to library users, high cost associated with generating sufficient metadata

Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
The value of AI at the organizational level includes production of more metadata and extraction of data from the handwritten text; better and increase precisions of search services; potential for large scale OCR or handwriting text recognition; freeing resources for other tasks; increasing cost-efficiency in generating metadata; better possibility to understand own collection data	The cultural value of making existing collections available to the public; democratic value of increasing access to cultural heritage; the value of providing new search services and cleaner metadata to the public; potential of using data collection in new ways and different stakeholder groups then wider public.	1) Start in small and later scale-up to larger quantities of data; 2) Make sure to jump on the AI wagon if and how it is the right thing to do; employee's competences, e.g. employee in library collections need computation skills and vice versa computational experts need an understanding of the collections; awareness of the ethical, sustainability and bias issues in using AI

2.4 National Library of Estonia

Case and project name			
Automatic Indexing and inquiry answering service Ask Us			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Estonia	250	Multi-modal AI	2023 / Prototype and major milestones accomplished

Project description		
The goal is to develop a prototype utilising machine learning and AI to automate publication content analysis and tagging.		
Need(s) behind implementation	**Actors involved**	**Challenges**
Tagging items is a labour-intensive task that demands significant resources. This process relies heavily on the knowledge and judgments of individual taggers.	Head of development at the NLE Internal and external project managers, Experts with experience in AI, data management personnel, development team members, customers and vendors.	Librarians felt uninformed and feared losing their jobs and as a result mistrusted the use of AI.

Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
The project is coordinated by the Head of Library Services Center and his team. It is supported by the Estonian Government and part of the overarching AI strategy.	- The users and patrons are described as very open-minded. They are eager to use the new technologies. They are very supportive and provide valuable feedback on improving and moving forward. - It can be challenging to grasp what AI is doing in the library and what the librarian's role will be, especially with things like automatic cataloguing or coding examples.	It was reported that the employees of the NLE have a low level of trust in AI solutions. Therefore, keeping them informed about ongoing developments is crucial to build confidence and transparency in the process. Furthermore, it is important to do it step by step: - Detailed Analysis - Prototyping - Production (This phase should only proceed when the AI solution works reliably and effectively)

2.5 National Library of Finland

Case and project name			
Annif and FintoAI at the National Library of Finland			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Finland	+230	Open-source toolkit called Annif for automated subject indexing and classification	**2014:** Start of the Finto project **2017:** Start of the first prototype of Annif **2018:** Begin developing Annif on a more robust technical foundation **2020:** Launch of the Finto AI service that is intended for production use

Project description

The national library has created an open-source Annif toolkit for automated subject indexing and classification. Annif is multilingual, works independently of the indexing vocabulary, is modular, and integrates different text algorithms such as Maui and a neural network based on TesnorFlow. Annif can be integrated into the Finto AI web user interface, an in-house-produced metadata management system. To evaluate Annif, they use manually indexed corpora, such as master's and doctoral theses from the University Jyväskylä. Additionally, humans evaluate the results of the automated subject indexing. Finally, in May 2020, the national library launched Finto AI, which introduces automatic subject indexing into information systems and can be used for Finnish, Swedish, and English subjects. Finto AI is, for example, already implemented at the University of Jyväskylä, where students submit their master's thesis using the tool and receive suggestions from Annif. The students can use or discard the suggestions before a final check by librarians.

Need(s) behind implementation	Actors involved	Challenges
1. Need for automated subject indexing arises due to the growing volume of digitally collected data and the extensive digitization of paper archives. 2. Allows for indexing library collections where manual indexing is impractical 3. Need an in-house solution due to the lack of AI support for multiple Finnish languages and dialects 4. Need an in-house solution due to the lack of control over the AI system from commercial providers and the difficulty of integrating them into existing systems	• The Annif project team developed the first prototype in 2017 and tested a use case at the University of Jyväskylä. The library management and the public showed interest in it, leading to further work on the project. • The Annif project team collaborates closely with external stakeholders to develop and enhance Annif by using, testing, and modifying the tool's code. The Leibniz Information Centre for Economics (ZBW) is an important collaboration partner, for example, in technical development. • Annif is implemented not only in the National Library of Finland but also in other libraries, such as the National	• Over time, the model has become more accurate with fewer mistakes. However, there are concerns about hidden mistakes in assigning subject headings. Therefore, it's important for librarians to remain skilled in cataloguing and indexing subject headings. • Librarians require sufficient time to use the tool effectively to achieve satisfactory results. • Some librarians are concerned that the tools may replace human workers and lead to layoffs. • On the macro level, the development of AIs by big commercial companies is a challenge and even a dangerous undertaking because there is little transparency about what they

	Library of Germany, universities like the University of Jyväskylä, organizations like ZBW, and companies like the Finnish Broadcasting Company Yle. • At the University of Jyväskylä, for example, students, librarians, and Annif work together to review Master's theses. Students upload their theses to the Finto AI tool, review the suggestions, and then a librarian gives it a final check. This process in Annif is a semi-automated indexing system. • The Annif project team welcomes feedback and suggestions for improvement, for example, on the GitHub platform, where external partners can report bugs and make pull requests, or in the user forum called annif-users. Additionally, they provide a hands-on tutorial with the ZBW.	are doing. Additionally, libraries could use AI to benefit society and counter bias and ethical issues. • There are challenges in using AI and dealing with copyright. The legal landscape is evolving, making it a grey area. • Nevertheless, this approach takes a lot of resources, such as library staff with technical skills. That's why the library goes with the strategy to invest in their employees and their infrastructure.

Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
The Annif project team first presented its ideas about Annif at a Library Networking Service meeting for management and the library colleges. They also conducted presentations about Annif's ideas outside the library. Additionally, the project team established a first use case with the University of Jyväskylä, demonstrating its effectiveness to management. An interviewee stated that they are experimenting with AI tools but have not fully integrated them organization-wide, and it is not yet implemented in the general process. Annif is performing its function, but it operates behind the scenes and still requires human assistance to yield valuable results. However, the National Library of Finland decided	1. AI needs custodianship and oversight, and the National Library of Finland plays a critical role in society in this regard. 2. There is positive feedback from users, for example, in the use case, where students submit their theses to the repository, which in return encourages the library to work further on it.	Automated subject indexing requires skills and competencies in manual subject indexing and description. For this reason, they advise that beginners or non-professionals avoid using the tool. In addition, some librarian staff are worried that the AI tool is seen as a way to cut off staff and provide quick but inadequate input, which could lead to incorrect information in the library database. Therefore, one interviewee recommended using the tool carefully and thoughtfully to improve metadata but urged more cautious usage. Furthermore, they already started embracing open access, data, and software. Progress has been slow, but there has been a gradual shift toward these principles.

against developing an AI department within the library because they still believe that AI should be integrated across the organization and understand its basics from everyone in the library. There are lots of benefits of an open-source, in-house developed automated subject indexing tool. Automated subject indexing can assist manual indexing through semi-automated indexing, where Annif suggests subjects for a new record that is verified manually. Additionally, they test Annif for fully automated indexing. It minimizes the need for librarians to check and add subject indexing. In the future, the tool will become more communicative and, therefore, faster. This gives further potential to work with stakeholders together. An in-house solution benefits from the huge amount of knowledge in-house, and staff can immediately identify mistakes in the code or the automated record produced.		

2.6 National Library of France (BnF)

Case and project name			
Gallica – French National Library			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
France	2128	OCR, HTR and image/data mining	1997, well underway/major milestones accomplished
Project description			
Gallica is the digital library of the BNF. This digital library is giving access to more than ten million free documents.			
Need(s) behind implementation	**Actors involved**	**Challenges**	
Information retrieval, cataloguing, ease the research inside the collections	Mostly internal (IT, Collections and Cooperation departments), external partners like BNUS, INHA, INRIA or technical partners (i.e. Mistral)	Moving from prototyping to implement and launch live Secure funding to achieve them	
Results			
Organizational Level	**Value created and co-created**	**Lesson learned**	
A specific department is in charge of Gallica, this is the coordination department. The project has a sponsor at the political level and is an important project for the BnF related to AI.	New services for users, new services for librarians and for partners (like other libraries), change at the process level to ensure a smoother experience as well for users and for employees.	Importance of gathering people from various services to be part of AI projects to make sure they will contribute to it. Improve AI's literacy in the library	

2.7 National Library of Germany (DNB)

Case and project name			
Petrus and EMa at the National Library of Germany (DNB)			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Germany	617 employees	Petrus is a process-supporting software that uses machine-based processes and AI for subject cataloguing. EMa ("*Erschließungsmaschine*") is a cataloguing machine that generates descriptive metadata to enrich the records in the DNB's catalogue.	**2009**: Project start Petrus; **2012:** Start of the implementation of Petrus; **2018:** Project start EMa; **2022:** Implementation of EMa

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. Over the years, challenges arose, including personnel turnover and the need for modular solutions. In addition, DNB worked closely with an external company that developed the software for them. In 2018, the company decided not to develop the software further for the DNB.

In 2018, the project called cataloguing machine EmA ("*Erschließungsmaschine*") 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. With the EMa application, it is now possible to open up further data records. 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 category of medicine for media works in German and English.

Need(s) behind implementation	Actors involved	Challenges
The law requires the acquisition of a copy of every media published in Germany, which results in the accumulation of vast amounts of data. New forms of publications 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.	The "Acquisitions and Cataloguing Department" is responsible for developing the collection and producing cataloguing data, i.e., vulgo cataloguing. This department includes the "Automated Cataloguing and Online Publications" department. The „Cross-cutting Matters of Use and Inventory Management" department is involved in experimenting and working with various types of innovation, from technical to social. Staff with different professional backgrounds, such as computer linguists, mathematicians, and experienced librarians, are involved in	1. Evaluating the information libraries receive from AI applications is crucial, as AI tools are neither complete nor always neutral. 2. The fear of cataloguers to lose their jobs 3. Limited availability of personal and material resources 4. Integrating AI into the organizational structure 5. Is uniform labelling of AI-generated output necessary? 6. Ethical considerations related to the use of AI 7. Copyright

	the AI project. The project members have a certain affinity with the topic of AI. 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, such as universities, other libraries, research institutions, consultants, suppliers and start-ups. Close cooperation between AI and librarians is required.	
Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
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. 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. Ethical considerations, such as generating misleading or harmful information, have not been extensively explored. Five years after its inception, the Petrus project saw the initial outcomes implemented in daily library work. Four years after its inception, the Ema was implemented.	The DNB considers it its duty to uphold democratic values that are closely tied to thoughtful consideration of information sources. However, AI outputs are not always neutral and may threaten these values. Therefore, the DNB aims to serve as a platform where society can stay informed about various developments in AI. In addition, DNB is a reliable source of information for AI tools. 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. Furthermore, in our changing society, the library is increasingly becoming a protected space that is not only a place of learning but also a place to meet and linger. Libraries will remain vital in society, and AI adoption enhances accessibility, search functionalities, and personalized library services. The DNB staff's attitude has changed to some extent following the introduction of AI-supported methods.	1. Need of more IT developers and 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 the results of the automated processes are good enough 5. It is crucial to approach ML and AI projects with naivety and impartiality and allow yourself to make mistakes 6. More than 10 years ago, DNB strategically used fully automated AI processes. This decision is currently being reviewed to determine whether a semi-structured automated process would be more advantageous, at least for certain applications. 7. DNB should have invested even more in resources 8. Experiences in the projects were valuable for further development of ML and AI in the DNB

2.8 Royal Library of the Netherlands (KB)

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 automatic metadating of books	Starting in mid-2023, the Retrotool was developed and tested, and officially implemented in September 2024.

Project description
The Retrotool was developed to address cataloging backlogs by processing the retro collection—old books that were acquired but not yet cataloged. Using a document camera (OCR), the tool captures information from the title page and colophon, and artificial intelligence organizes it into cataloging fields. The tool then checks if the book is already in the catalog. 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 (Interviews with vice director and head of research) Collection data department: Head of Department, two project leaders (interviews with project leader and interim project leader/cataloguer) Four cataloguers who tested and currently work with the tool A technical company who developed the tool	Ethical challenges (use of LLM created by Microsoft is quite opaque).and fear by employees that AI would make their work obsolete.

Results		
Organizational 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 feedback from cataloguers and other staff incorporated at various stages, and the process is progressing well. KB management has identified patron engagement in innovation projects in general as an area for improvement and further exploration.

2.9 National Library of Norway

Case and project name			
National Library of Norway/Maken Project			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Norway	420	Recommendations system based on the content of books and images	Experimental stage/Demo launched on website…
Project description			
"Maken is a discovery engine to find similar books and similar images. Based on the pixels of the images, Maken finds related images. Based on the text of a book itself, Maken finds books that have some resemblance to each other"			
Need(s) behind implementation	**Actors involved**	**Challenges**	
Increasing difficulty in discovering the digital material; increasing amount of material; potential lack or insufficient metadata	Top management, AI-lab, users, external digital agency	GDPR and privacy issues, Data Protection Impact Assessment, consent for the interviews	
Results			
Organizational Level	**Value created and co-created**	**Lesson learned**	
User involvement and digital innovation/ transformation at organizational level	Democratic value enacted by the easier access to digital books and images for the library users, contribution to the library strategy of experimenting with AI	Importance of involving users in the process and the importance of understanding digital change in the libraries	

2.10 National Library of Slovenia

Case and project name			
Narodna in univerzitetna knjižnica, NUK - National and University Library, NUL			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Slovenia	134	File format conversion	Slovenia
Project description			
Use of AI to significantly enhance the accessibility of digitised library content for blind and visually impaired communities through various SOTA technologies (mostly FLOSS).			
Need(s) behind implementation	**Actors involved**	**Challenges**	
Compliance with WGAC Better (re)usability of outputs	The library's research group Blind/visually impaired users Mobile/print disabled users	Related to implementation Related to take-up and adoption	
Results			
Organizational Level	**Value created and co-created**	**Lesson learned**	
Increased user base Reputational factors Automation vs manual handling Importance of FLOSS solutions Time and resource constraints	AI literacy Collaboration Focus on needs Designing Experimenting	It's not only about technology, but also how it is deployed. The challenge lies in the alignment between human, technological and organisational factors.	

2.11 National Library of Spain (BNE)

Case and project name			
BNE - ETSO Project: Stylometry Applied to Golden Age Theater			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Spain	364	Transkribus to automatically transcribe and modernize old prints and manuscripts with a high degree of accuracy.	2017 Implemented and alive

Project description
Thanks to the AI-powered tool Transkribus (READ-COOP, 2024) and its Handwritten Text Recognition (HTR) techniques, Dr. Cuellar and the ETSO team have trained three models, which are being used and refined by the research community, capable of automatically transcribing and modernizing these documents with a high degree of accuracy: approximately 97% accuracy for printed texts and 91% for manuscripts. Through these models, ETSO has processed around 1,800+ theatrical works contained in prints and manuscripts from numerous libraries, archives, and other digitized sources. The resulting transcriptions are now part of the CETSO Corpus and the TEXORO search engine, both part of the ETSO project. Moreover, they possess sufficient quality to undergo stylometric analysis, which reveals interesting authorship attributions.

Need(s) behind implementation	Actors involved	Challenges
To automatically transcribe the Spanish Golden Age documents for stylometric analyses able to establish trusted authorship attributions.	• Dr. Álvaro Cuéllar (Universitat Autònoma de Barcelona) and Dr. Germán Vega García-Luengos (University of Valladolid) • Their teams. • The research group on Lope de Vega PROLOPE (Universidad Autónoma de Barcelona). • 200+ researchers, students, experts, and collectors from 20+ institutions and private collections.	1. Automatic transcription using Optical Character Recognition (OCR) faces limitations due to historical typefaces and handwritten manuscripts from the XV-XVI centuries, making manual transcription more efficient. 2. Modernizing spelling in the works is essential for meaningful comparison with other modernized texts, even if OCR achieves precise transcription. 3. Despite digitization, the vast quantity of Golden Age theater works remains inaccessible for analysis due to lack of transcription. 4. In 2020, Transkribus lacked pre-existing models for recognizing Spanish. The creation of the initial recognition model, Spanish Golden Age Prints 1.0, allowed researchers to transcribe digitized manuscripts. A refined version of the model, incorporating modernized spelling, is also available via the READ-COOP platform. 5. Despite having tools to recognize print from old works, errors persist due to factors like challenging handwriting, crossed-out sections, document dirtiness, translucency, and the AI tools' training data limitations.

		6. Manuscripts pose additional complexities due to their handwritten nature, resulting in more errors during automatic transcriptions. Seventh Challenge: Stylometry can reveal authorship patterns in printed works and manuscripts, but caution is needed due to recognition tool errors and stylometric connections to specific authors, as seen in the case of Lope de Vega.
Results		
Organizational Level	**Value created and co-created**	**Lessons learned**
The work done by the ETSO project team did not produce any reported effect at the organizational level of the National Library of Spain. However, and since the BNE plays a pivotal role in understanding cultural phenomena, ETSO helps successfully achieve this role. By safeguarding the world's most significant collection of manuscripts and printed materials related to Golden Age theater, the BNE provides insights into the theater of that era. ETSO complements the efforts of the BNE and contributes to them revealing the keys to this theater's historical context and its status as a paramount cultural heritage. It also highlights its relevance in contemporary stages, scholarly studies, and the role of new technologies in dissemination, research, and even performance. In the words of the former BNE director, Ms. Ana Santos Aramburu, "we are aware that the digitization of culture serves to create wealth and has economic value in addition to its cultural significance" (Muñoz, 2018). In other words, the ETSO project exemplifies how AI can support public value creation, especially in	The ETSO project is highly collaborative at various levels: design, implementation, analysis of results, integration of stakeholders including experts, scholars from different disciplines, students, associations, collectors, and the BNE. In addition, it emerged from the personal initiative of Dr. Cuéllar, who put together the large team involved in this co-creation effort. According to the evidence collected about the project, the major outcomes from their collaborative effort are: 1. The most advertised outcomes of the ETSO project are the attributions of La francesa Laura and Mujeres y criados to Lope de Vega. However, other works have been attributed to their right playwrights, such as La monja alférez to Juan Ruiz de Alarcón. Without automatic transcription, it would have been practically impossible to discover those works within the vast amount of Golden Age theater documents that exist. Therefore, these processes provide researchers with invaluable clues, which are afterwards investigated through traditional philological tools (ETSO, 2023). 2. 2. ETSO has developed three models for automatic transcription using the Transkribus tool and its AI system. Additionally, the project has created a corpus of transcribed works, CETSO, that is	The key lessons learned from the ETSO project, particularly regarding the use of AI in this collaborative initiative are: 1. AI Facilitates Large-Scale Projects but Requires Initial Human Input. The ETSO project has shown that AI, particularly through tools like Transkribus and Stylo, can dramatically accelerate the transcription of historical texts. However, the process still requires substantial initial human input to train the AI models to achieve high accuracy (up to 97% for printed texts). Human researchers remain critical for correcting errors and improving the AI models. 2. AI-Driven Collaboration Across Disciplines. The project involved collaboration between researchers, libraries, and digital service providers. AI tools enabled a shared platform where contributions from various scholars (philologists, historians, and technologists) could be pooled to enhance the transcription and analysis process. This demonstrates that AI can be a catalyst for cross-disciplinary cooperation.

cultural heritage. By transcribing vast amounts of manuscripts and printed works, AI made these historical texts accessible not just to academics but also to the public. This aligns with the mission of institutions like the BNE, highlighting how AI can serve broader societal and educational purposes by preserving and democratizing access to cultural assets.	being used by researchers to develop their investigations. Lastly, the TEXORO tool, is a search engine that allows exploring which works contain specific words or phrases. Currently, TEXORO contains the same texts as CETSO, but the intention is to expand in the future to include other types of texts beyond the theatrical domain, and it has a separate identity within the project. 3. Thanks to AI, the ETSO project significantly reduces the workload for researchers by mechanically transcribing extensive corpora of easily readable documents. This is especially relevant to analyse the hundreds of printed and theatrical manuscripts from the Spanish Golden Age and beyond that have remained forgotten and neglected, waiting to be worked on by philologists. Now, they can use texts that, if transcribed through traditional means, would have required a tremendous amount of effort. 4. For now, trained human researchers can still transcribe this type of document more accurately, albeit at a much slower pace. The work undertaken by ETSO would have taken several years for a large group of philologists. However, thanks to AI and its training, each work is transcribed in a matter of minutes. Digital Humanities provides invaluable assistance to researchers of Golden Age theater. These processes and tools, far from replacing the researchers, facilitate and propel their work to unprecedented levels.	3. Challenges in Historical Data Handling. The ETSO project encountered several challenges, such as the limitations of Optical Character Recognition (OCR) for older typefaces and manuscripts, and the need to modernize spelling. These challenges highlight that AI must be adapted to specific historical and linguistic contexts to be effective, particularly when dealing with non-modern languages and formats. 4. AI as a Tool for Discovery. AI has been instrumental in discovering previously unattributed works, such as the attribution of La Francesa Laura and Mujeres y criados to Lope de Vega.

2.12 National Library of Sweden (NLS)

Case and project name			
[Swedish National Library/ The project applies AI to search in a collection of 17,000 digitized postcards. It is possible to search by simple key words such as "Church", blue sky]			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
Sweden	350	Swe-CLIP 2M	Experimentation stage; a demo has been launched on the web in 2023

Project description		
Searching in postcard collections of ca. 17,000 digitized postcards		
Need(s) behind implementation	**Actors involved**	**Challenges**
Very limited access that the users had to this kind of collection; sometimes even the awareness of its existence.	KBlab developers/Librarians/Archivist/designers; library users; copyright agency	Copyright, technical integration and accessibility challenges with the rest of the library system; the lack of resources and lack of strategic leadership to put the demo service into full production

Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
AI models relieve library employees and free resources for other jobs; KBlab's contribution to library innovation processes	The democratic value of accessibility to a postcard collection	Develop internal competences to use AI responsibly and ethically; AI expertise gets integrated within the library and takes into consideration some of the informational challenges of large-scale collections; building AI on a contextual understanding of libraries as a particular type of organisation with specific challenges; develop AI where it can add value to the library

2.13 National Library of the United Kingdom (BL)

Case and project name			
The project was initiated to explore the impact of AI and machine learning on historical research, focusing on the effects of mechanization in the long 19th century. It involved collaboration between the Alan Turing Institute, partner universities, and the British Library. Challenges included interdisciplinary collaboration, technical expertise and staff rotation, copyright and access issues, and project management. Despite these challenges, the project aimed to address the evolving landscape of historical research in the digital age			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
United Kingdom	1300	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 digitization 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 explore how AI and machine learning would impact historical research, with a focus on the effects of mechanization during the 19th century. 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 digitization on user engagement	It involved collaboration between the Alan Turing Institute, partner universities, and the British Library. Actors involved include internal teams like the Digital Research Team and external partners like the Alan Turing Institute. Involving over 5,500 volunteers, the project engaged the public through crowdsourcing tasks. Collaboration spanned disciplines, including data scientists, historians, and research software engineers, with members joining gradually between 2018 and 2019.	didn't 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		
Organizational 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 labeled 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, cultural enrichment, and fostering innovation. Initiatives such as exhibitions, crowdsourcing, workshops, and innovation	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. These initiatives drive innovation, enhance public engagement, improve collections access, build staff capacity, and align with strategic library goals.

2.14 Library of Congress (LC)

Case and project name			
Exploring Computational Description at the Library of Congress (LC)			
Country	**Number of employees**	**Type of AI solution**	**Year and Maturity level**
USA	3,172 permanent employees	Using AI and ML to create descriptive metadata for e-books that have not yet been catalogued	August 2022: Project start; September 2023: End of first part of the project, called "Toward Piloting Computational Description"
Project description			
The LC Labs and the Acquisitions and Bibliographic Access Directorate are currently working on a project called Exploring Computational Description to determine the feasibility of using AI and ML to create descriptive metadata for e-books that have not yet been catalogued. The project began in August 2022 and involves collaboration with an external vendor, Digirati. Digirati tested five different models for machine learning and used two workflows for assisted cataloguing, creating a subject assistant suggestion prototype. In September 2023, the first part of the project, called Toward Piloting Computational Description, was concluded, and the team has now moved on to the second part, which focuses on prototyping the use of machine learning in greater detail.			
Need(s) behind implementation	**Actors involved**	**Challenges**	
In order to fulfil LC's goal of connecting all Americans to the library, LC needs to investigate new approaches that could bring significant changes to its practices. LC Labs are tasked with experimenting with new methods, approaches, and technologies, and collaborating with external stakeholders such as universities and research initiatives. Additionally, there is support from the library management and funding from the National Digital Trust Fund to experiment with AI and ML.	The LC Labs serve as an experimental space for testing different projects and are part of the Digital Strategy Directorate. The Acquisitions and Bibliographic Access Directorate primarily catalogues books and e-books and is part of the Discovery and Preservation Services. These two departments collaborate and work with external vendors such as Digirati or with research institutions and academic partners. Furthermore, the Planning Framework, developed by the LC Labs, helps to connect library staff with patrons and the library community.	1. Should patrons, mostly other libraries, be informed if a record was generated using AI and whether the record should contain details about the confidence level of the AI-generated record? 2. Can we use machine learning and AI to process content that is copyrighted? What will the machine do with the content after the modelling is complete? 3. What level of accuracy is considered acceptable for an AI-generated output? To ensure accuracy, a cataloguer is still required to identify possible mistakes and input the information correctly into the record. This is closely related to creating quality standards and policies specifically for AI. 4. AI struggles to identify the title and author of an e-book due to the varying design elements and typography. Additionally, predicting e-book content and identifying non-controlled vocabulary words remain challenging for AI. Additionally, the predicted words and terms may not always be useful or informative and cannot be corrected.	

		5. How should we deal with harmful and disrespectful language in books and cultural heritage? This can be particularly problematic with old books that contain such language and when these are used to train ML models. 6. Training AI requires large volumes of data, including born-digital materials, which can be challenging. 7. There is a risk of ineffective ML and AI solutions, which is connected to the fear of investing resources into solutions that may not work or could worsen existing issues.
Results		
Organizational Level	**Value created and co-created**	**Lesson learned**
AI can provide extractive or abstractive summaries, which can assist cataloguers in understanding the essence of a book without having to go through the entire book. Additionally, AI can take care of routine tasks, such as identifying information on the title page, freeing up librarians to focus on intellectual tasks. This can lead to more efficient use of time and resources. Many library staff members are concerned that AI is replacing their jobs. However, this is not necessarily the case. In the short term, AI creates more work because someone has to ensure it is functioning correctly, train and revise it, and continually monitor it. This leaves library staff more time to focus on intellectual tasks such as reviewing the model or correcting records.	LC Labs conducts public affairs has to report to Congress about its progress in experimenting with and implementing ML and AI. However, LC Labs lacks the resources to educate more staff about ML and AI technologies in the LC. In addition, the interviewee does not think AI is changing the mission and role of the library. Moreover, the expertise of the cataloguers is even more valuable in identifying wrongly predicted content and errors. It always requires librarians with the necessary skills to comprehend and apply cataloguing principles effectively. Nevertheless, library staff need at least a basic understanding of what ML and AI mean and how they work. After the records are completed, they are sent to OCLC and become accessible in the LC catalogue and for other libraries. However, staff capacity and time constraints may delay the project's implementation until around spring 2025. Adopting AI in libraries is seen as a way to create public value. However, one interviewee expressed concerns about the limited availability of library materials due to under-resourcing. This makes it difficult to connect with the public effectively.	When the LC had the opportunity to restart the project, it would provide the AI application with more content to improve its accuracy. However, in the current project phase, they have increased the number of records to 100,000 e-books, hoping this will result in a more precise model. Create a guideline on how to deal with harmful and disrespectful language use. Start with a large-enough data set, so that the AI tool can learn and provide more accurate results over time.

3 Next steps in the cross-case analysis for task 3.3 v03

As a next step, we will use the case illustrations to compare and contrast the findings across cases to understand how AI use in National libraries contributes to their digital transformation. In addition, we will derive recommendations for both practice and research.

4 Annex I - Interview Guide

Interview Guide

Objective:

- To gain an understanding of the implementation steps of AI in National Libraries;
- To understand how AI is intended to support National Libraries in delivering value;
- To understand how different stakeholders are involved in the implementation of AI solutions and what value(s) they perceive;
- To understand how cultural heritage data could support the development of ethical and trustworthy AI solutions and
- To provide a knowledge base for the development of action strategies at the intersection of the humanities and AI.

For internal use only, please give information about the interviewer and interviewees. Information about interviewees is only for internal use during the project period.

Name and role of interviewee	
Name of national library	
Name of interviewer and date	

Part I. General questions

1. At the start, can you please tell me what your **current position and responsibilities** are?

2. Could you please briefly describe the National Library where you work: its history and current focus?
   Follow-up question:
   - *In your words, what do you consider the main role of National Libraries?*

3. One of the key areas that we want to investigate is the use of **AI in libraries**: why have you decided to start using this technology?

4. Could you please describe the AI project you are currently working on?
   Follow-up questions:
   - *Why was this project initiated, and by whom?*
   - *Why have you decided to rely on this specific AI solution?*
   - *How will the AI solution be useful for library services?*
   - *How many employees have worked on this project?*

Part II. Questions about AI impacts on library mission and strategy

5. How does the implementation of the AI solution support your **library´s mission**? [strategy; AI strategy]

6. How will AI support the **library in maintaining or developing…**
   - *cultural and democratic values?*
   - *Preservation and accessibility of the information?*
   - *Inclusion and ethical issues?*
   - *Data privacy?*

Part III. AI implementation within the library

7. What is the expected outcome of the current AI project you are implementing? [user services; internal process; data & collection management; citizen value; administrative value]

8. Can you walk me through the **AI implementation process**?
   Follow-up questions:
   - *How is this technology finding its way into the library (pilot projects, systemic implementations…)?*
   - *What are the **key ingredients** – already in place or to be acquired – for the **proper adoption of AI**?*

9. Which new **limitations** does AI bring? [ethical concerns; bias in data; …]

Part IV. Value creation and co-creation through AI

10. How are the different stakeholders of the library involved in the implementation of the AI solution?
    - Patrons?
    - Librarians?
    - Other organizations?
    - Others?

11. How do you think the implementation of the **AI solution is affecting how the library collaborates** with [co-production by individual/group/collective; co-designing; co-delivery; constellations of relations]:
    - the society
    - patrons
    - other libraries
    - external providers
    - research community

Could you please provide specific examples?

12. How has the implementation of the AI project changed the representation of knowledge? [multi-modal way of accessing knowledge]

13. What is the role of library data heritage for AI?

14. How can this project and the adoption of AI help to address AI literacy in society?

Part V. Organizational change as a result of AI introduction

15. Library are knowledge organizations: how is AI related to your internal knowledge management or the data management of your artifacts?

16. How do the tasks of librarians change as a result?

17. AI can perform many tasks independently but it also serves to **support human activities**: can you describe to me how the **interactions between humans and AI** is structured within the library? [human-machine relationship]
    Follow-up question:
    - *Could you give some examples of what you are saying?*

18. What new organizational roles have emerged through the implementation of AI?

19. What **skills and competencies are** needed to reap the effects of AI? [skills; digital skills; AI literacy]
    - Librarians
    - Leadership
    - Other library employees
    - Patrons

20. How do you manage the transfer of AI competencies from the project team to the entire library?

Part VI. Outlook

21. What would you do **differently next time you have to start a new AI project**?

22. How do you think the introduction of **AI will change the role of libraries in the next 5-10 years**?
    Follow-up question:
    - *What do libraries need to get there?*

23. How can libraries, as a sector, adopt AI solutions in a value-driven way?

24. What can the role of National Libraries be in the development and implementation of AI?

These were all the questions we/I had.
Is there anything I forgot to ask or you would like to add?
Thank you for answering our questions!