

# Harnessing Power Apps and AI for Automated Cataloguing

## Innovations in Bibliographic Record Creation

**Hannes Lowagie**  0000-0002-0671-3568

Metadata Librarian, KBR (Royal Library of Belgium)

Received: 01 Oct 2024 | Published: 16 Dec 2024

### ABSTRACT

In the evolving landscape of library and information science, the integration of advanced technologies has become essential for efficient and accurate cataloguing. This talk explores the transformative potential of Microsoft Power Apps, focusing on the use of AI components for the detection and enhancement of metadata. We will delve into how Power Apps can streamline the cataloguing process by automating repetitive tasks and ensuring consistency in metadata creation.

Additionally, Power Apps can assist with various other crucial tasks in the realm of cataloguing. We will present initial tests demonstrating the use of Power Apps for creating an RDA application profile, which is linked to an automated validation file. This validation file can be used to ensure individual records conform to the RDA application profile standards.

**KEYWORDS** metadata enhancement; AI; cataloguing automation; application profiles; Power Apps; Power Automate; record validation

**CONTACT** Hannes Lowagie  hannes.lowagie@kbr.be  KBR (Royal Library Belgium)

## The Need for Automation in Cataloging

### Introduction

Libraries and metadata specialists are continuously seeking ways to streamline their workflows and enhance the accessibility of their collections. The sheer volume of materials requiring cataloging, especially in institutions with extensive holdings like the Royal Library of Belgium, presents significant challenges. Traditional cataloging methods are time-consuming and labor-intensive, making it difficult to keep pace with the growing influx of new and historical materials. Enter automation tools like Microsoft Power Apps and AI-powered technologies, which offer innovative solutions for automating cataloging processes. These tools not only simplify the creation of bibliographic records but also enhance metadata enrichment and validation, transforming the way libraries manage their vast resources. In this article, we explore

how Power Apps and AI are revolutionising cataloging, offering practical, customizable solutions that cater to the evolving needs of librarians and metadata specialists.

### **Challenges in Traditional Cataloging**

For institutions like the Royal Library of Belgium, with its vast collection of 5 million books and 3 million patrimonial objects, the task of cataloging represents a monumental challenge. Manually creating bibliographic records for each item is a labor-intensive process that demands significant time and effort. From medieval manuscripts to modern-day publications, each record requires meticulous attention to detail to ensure accuracy and completeness. Librarians and metadata specialists are often stretched thin, tasked with inputting data, verifying metadata, and organizing resources in a way that facilitates efficient retrieval. This manual approach, while thorough, can lead to bottlenecks, especially when managing large-scale collections. The growing volume of incoming materials only exacerbates the issue, making it difficult for libraries to keep their catalogs up to date and accessible to users.

### **Power Apps and AI-Builder**

#### **What is Power Apps?**

To address these challenges, we turned to Microsoft Power Apps, a low-code platform developed by Microsoft. It allows users to build custom applications designed to automate repetitive tasks without requiring extensive coding knowledge. Power Apps can offer a practical solution to develop cataloging tools without requiring advanced technical expertise. Its integration capabilities make it even more powerful, as it can easily connect with various tools and databases such as SharePoint, Microsoft 365, SQL Server, and Dataverse, but also with API and FTP.

Paired with AI, these tools streamline the cataloging process by automating metadata extraction and enrichment. For instance, AI models can identify key metadata elements such as the title, author, publisher, and publication date from scanned documents, reducing the need for manual input. By implementing these technologies, we can ensure faster access to our collections. Automation not only enhances efficiency but also enables librarians to shift their focus to more complex tasks, such as curation and improving user services.

#### **What is Power Automate?**

Power Automate is a cloud-based service from Microsoft that helps users automate workflows and business processes. It allows you to create automated workflows between different applications and services, such as Microsoft 365 apps, third-party services, and on-premises data sources. With Power Automate, you can automate repetitive tasks, streamline processes, and integrate data across various systems without needing extensive coding knowledge. There are various triggers that can start a Power Automate flow. One trigger could be a button click in Power Apps, but a flow

can also be initiated by creating a file (such as a scan) in SharePoint or OneDrive. In Power Automate you can also use the Microsoft AI-models

### **Microsoft AI-models in Power Apps and Power Automate**

Microsoft's AI Hub offers a variety of tools that libraries can leverage to build powerful AI models tailored to their cataloging needs. These models can be either pre-trained or custom-built, depending on the specific requirements of a library's collections. Pre-trained models, which are ready to use, can expedite implementation, while custom models allow for more precise handling of unique or specialised datasets. Both approaches offer libraries the flexibility to automate key aspects of cataloging, ultimately enhancing efficiency and accuracy in managing bibliographic records.

#### A - AI Models Based on Images

One of the most impactful uses of AI in cataloging is the ability to process images for metadata extraction. For example, the "Document Processing" model is highly effective in detecting pre-defined metadata from scanned images of books and manuscripts. The AI examines elements such as the structure of the title page, text placement, font size, and even content-specific information like edition statements. Impressively, these models require minimal training data—sometimes as few as five examples—to start generating accurate results. This makes the process of creating custom AI models more accessible, allowing us to quickly build and deploy tools that automate the extraction of critical metadata out of images of title pages, but also library cards, old maps, periodicals, and so on.

#### B - AI Models Based on Text

In addition to image-based models, AI also excels in processing textual data. An interesting type for libraries are the entity recognition models. These models, both pre-trained and custom, are capable of identifying key entities such as names, dates, and locations within text blocks (for examples out of back cover text, or parts of the summary or table of contents). Another useful model is the Category Classification model<sup>[1]</sup>, which helps classify your books. To use this model, you need to create a Dataverse table with classification or subject terms in column A and training texts (such as other summaries or texts) in column B. When you train your model, it will assign a term to an unknown text. It functions similarly to other proven tools like Annif<sup>[2]</sup>. Based on the classification or subject terms you provide, the model will categorize texts into appropriate categories based on detected metadata. Whether a library is dealing with books, manuscripts, or periodicals, these models enable quick and accurate classification, facilitating better organization and retrieval of materials. By employing a combination of image-based and text-based AI models, libraries can automate large portions of their cataloging workflows, leading to faster and more accurate metadata generation.

## Section 3: Power Apps as our solution for Retro-cataloguing in KBR

### Automated Metadata Detection

At the Royal Library of Belgium, retro-cataloguing presents a particularly challenging task due to the vast number of books and patrimonial objects that have yet to be fully digitised and cataloged. With over 3 million of the library's 5 million books already included in the online catalog, there remains a significant backlog of historical materials that require detailed bibliographic records. To address this, the library developed a retro-cataloguing application using Power Apps and Power Automate, integrated with above-mentioned Document Processing AI-model, designed specifically for extracting metadata from paper documents. The workflow is straightforward yet highly effective: scanned images of title pages are uploaded to SharePoint, where an AI-powered flow is triggered via Power Automate. The AI model automatically detects key metadata such as the title, author, and publisher from these scans, which is then saved into Dataverse and visualised within the Power Apps interface. In the application a human cataloguer can select a record. Then the scan of the title page is visualized next to the detected metadata (title, subtitle, author, place of publication, publisher, year, edition statement). The cataloguer can then correct (if needed) and validate the record. The final step involves exporting the validated records into a CSV format, where it can be imported into our LMS.

### Human Validation and Correction

Although the automation process significantly reduces the workload for librarians, we still chose to include the step of human validation in the workflow to ensure the quality of the cataloging data. Volunteers play an essential role in validating and correcting the automatically detected metadata. This project opened up opportunities for volunteers to engage with library work, contributing their time and expertise to refine the bibliographic records. The flexibility to incorporate human oversight allows libraries to maintain high standards of accuracy while still reaping the benefits of automation. By combining the strengths of AI with human expertise, libraries can create a more dynamic and efficient cataloging system.

## Next Phase: Legal Deposit and Other Projects

### Legal Deposit

In Belgium, legal deposit is the primary method of book acquisition. It's crucial to register incoming books promptly to confirm receipt to publishers and fulfill their legal obligations. To address this need, we have enhanced our application to handle not only older books but also new arrivals efficiently. Firstly, we added an AI model based on the colophon, because in contemporary books, most imported information is put in the colophon and not anymore on the title page. Secondly, -and this is one of the advantages of Power Automate flow- detected metadata can immediately be used in other processes, like HTTP queries. So, for example, if the AI-colophon model detects

an ISBN on the colophon, it can use that ISBN directly to search in other databases (using SRU queries) to see if they already have a record and if so, use that information to improve our own record. Thirdly, instead of using a CSV export, the application now sends the created record directly to our Library Management System (LMS) via FTP. As a result, just minutes after scanning, a record is created in our LMS, allowing publishers to see their deposited book online almost immediately. This tool is in production since September 2024 in KBR.

## **Other Projects**

### Periodicals

AI's role extends to the indexing of periodicals, where it can assist in article-level indexing. This application is particularly beneficial for school libraries and other institutions managing extensive collections of journals and magazines. By leveraging AI to identify and index individual articles, libraries can enhance the granularity of their catalogs, making it easier for users to locate specific articles and topics within periodicals.

### Old Maps

We will use a model specifically trained based on old maps to detect the metadata of those documents. The model can analyze and extract metadata written on top or bottom of the map, like title, map number, producer, scale, coordinates, and so on. This information is put in an Excel database so that corrections can be done more easily in batch by the Map Department.

### Handwritten library cards

We use the same workflow to detect the information on a collection of handwritten library cards that inventorize old prints and drawings. The model recognises and transcribes handwritten text. That way, the information is made accessible and searchable for researchers and historians.

As these technologies continue to evolve, their potential to transform library cataloging and metadata management grows. The integration of AI with platforms like Power Apps not only promises to streamline workflows but also to unlock new possibilities for managing diverse and complex collections. By embracing these advancements, libraries can stay at the forefront of innovation, ensuring that they meet the changing needs of their users while preserving and enhancing their valuable collections.

## **RDA Application Profiles and Validation**

Resource Description and Access (RDA) is a framework for cataloging that provides guidelines for describing and accessing resources. It is essential in ensuring that bibliographic records are consistent, comprehensive, and useful across different

---

systems and institutions. In a test phase we now use Power Apps to create custom RDA Application Profiles tailored to our specific needs. The idea is that we use the form in Power Apps to select the elements we want to describe for each entity, together with the specifications like if the element is mandatory, repeatable, if it is unstructured/structured/with identifier/with IRI, if it is linked to a vocabulary, and so on. But the most important feature of this app is that it also creates a file that can be used to validate our records, to control if they fit the definition of the Application Profile. For that, we need to take a look at Shacl4Bib and QA Catalog.

### **Shacl4Bib, QA Catalogue and Custom Validation**

The Shapes Constraint Language (SHACL) is a formal language for validating RDF graphs against a set of conditions. Following this idea and implementing a subset of the language, the Metadata Quality Assessment Framework provides Shacl4Bib: a mechanism to define SHACL-like rules for data sources in non-RDF based formats, such as XML, CSV and JSON. QA catalogue extends this concept further to MARC21, UNIMARC and PICA. The criteria can be defined either with YAML or JSON configuration files or with Java code. Libraries can validate their data against criteria expressed in a unified language, that improves the clarity and the reusability of custom validation processes[3].

QA Catalogue is a metadata quality assessment tool designed for evaluating library catalog records[4]. At KBR, with a catalog encompassing approximately 4.7 million title descriptions—many of which originate from older catalogs or retro-cataloguing projects—we face considerable challenges in data cleanup and enrichment.

To address these challenges, we employ QA Catalogue, an open-source tool that helps us assess and improve our records. This tool ensures that our data complies with MARC21 standards by identifying and rectifying issues such as invalid codes, incorrect values, obsolete codes, undefined fields, and the repetition of non-repeatable fields. This process is crucial for maintaining the accuracy and reliability of our records.

Additionally, we have integrated the new feature in QA Catalogue that allows custom validation using Shacl4Bib. By creating a tailored Application Profile in Power Apps, we translate these profiles into rules and specifications, formulated in the YAML language. This file we can upload and it is used to check the records against those rules. The result is put in the QA Catalogue dashboard where we can see how many records do not fit the rules (with the possibility in downloading the identifiers of those incorrect records)[5].

Although this project is still in the testing phase, it highlights the significant impact of Power Apps and demonstrates its potential to enhance various projects and processes at KBR.

## Conclusion

The integration of Power Apps and AI into cataloging processes offers significant benefits, transforming the way libraries manage their vast collections. By automating repetitive tasks, such as metadata extraction and classification, these technologies enhance efficiency, accuracy, and consistency in cataloging. Power Apps provides a customizable platform for creating tailored applications that fit the unique needs of each institution, while AI models offer advanced capabilities for processing and enriching metadata. Together, they streamline workflows, improve data quality, and enable librarians to focus on more strategic tasks.

Looking ahead, the potential for further developments in AI and automation is substantial. As these technologies continue to evolve, they will likely offer even more sophisticated tools for metadata management, further reducing manual effort and improving data integration. The evolving role of technology promises to make cataloging more dynamic and responsive to the changing needs of libraries and their patrons.

Metadata specialists and librarians are encouraged to consider adopting these innovative tools to enhance their cataloging processes. By embracing Power Apps and AI, or similar tools, libraries can achieve greater operational efficiency and better serve their users. That is why we invite readers to stay informed about future advancements in AI and library automation. Embracing these technologies can significantly improve cataloging workflows and contribute to more effective metadata management.

## References

- [1] Microsoft (2024) *Overview of the category classification custom model*. Available at: <https://learn.microsoft.com/en-us/ai-builder/text-classification-overview>
- [2] National Library of Finland (2020-) *Annif*. Available at: <https://annif.org/>. See also Osma Suominen, O., Koskenniemi, I. (2022) 'Annif Analyzer Shootout: Comparing text lemmatization methods for automated subject indexing', *Code4Lib Journal*, 2022-08-29 (issue 54). Available at: <https://journal.code4lib.org/articles/16719>
- [3] Király, P. (2024) *Shacl4Bib: custom validation of library data*. Available at: <https://doi.org/10.48550/arXiv.2405.09177>
- [4] Király, P. (2024) *qa-catalogue* (0.7.0). Available at: <https://github.com/pkiral/qa-catalogue>
- [5] <https://qa-data.kbr.be/?tab=shacl&lang=en&groupId=0>