Metadata & Discovery Group

Authentic and artificial cataloguing

Machine-readable cataloguing rules as a bridge to unified metadata practice

Hannes Lowagie 0000-0002-0671-3568

Head of Bibliographic Information Agency, Koninklijke Bibliotheek van België - Bibliothèque royale de Belgique (KBR)

Received: 17 April 2026 | Published: 16 September 2026

ABSTRACT

Cataloguers and metadata professionals are entering a period in which local cataloguing rules must be not only transparent and consistent, but also usable by machines. This article examines a metadata strategy developed at KBR, Royal Library of Belgium that transforms traditional cataloguing guidance into a modular, machine-readable ruleset structured in JSON. Rather than maintaining separate documentation for staff and systems, the JSON file functions as a single source of truth that supports two outputs: a dynamic HTML interface for human users and a machine-readable interface that can be integrated into generative AI workflows and other metadata processing environments.

The article explores how this "triangle model" – JSON as the core structure, HTML as the human interface, and prompt-ready data as the machine interface – offers a scalable approach to maintaining consistency, authority control, and transparency in cataloguing practice. It also considers the wider implications for interoperability across libraries and cultural heritage institutions. If local rules are encoded in comparable machine-readable forms, alignment between institutions may move beyond shared terminology towards shared cataloguing logic. The article concludes by reflecting on how metadata professionals can shape AI-enhanced cataloguing environments while preserving professional judgement, local context, and accountability.

KEYWORDS metadata standards; artificial intelligence; interoperability; JSON

CONTACT Hannes Lowagie Hannes.Lowagie@kbr.be Koninklijke Bibliotheek van België - Bibliothèque royale de Belgique

1. Introduction

Metadata work is changing rapidly as libraries and cultural heritage institutions begin to explore the possibilities of artificial intelligence and automation. Tasks that were once carried out entirely by human cataloguers are now increasingly supported by digital tools capable of extracting, analysing, and suggesting metadata at scale. This does not diminish the importance of professional expertise; rather, it changes the context in which that expertise is applied. Cataloguers and metadata specialists are now expected not only to understand descriptive standards and local practices, but

©2026 The Author(s). This article is licensed under a Creative Commons Attributions 4.0 License (CC BY 4.0)

also to work confidently within AI-assisted environments where speed, consistency, and data quality are essential.

At the same time, many institutions still rely on local cataloguing documentation presented in traditional formats such as printed manuals, PDF files, Word documents, or static web pages. While these formats remain useful for human reference, they are no longer sufficient in environments where systems need to interpret and apply rules automatically. Static documents can be difficult to update, challenging to search efficiently, and often disconnected from the tools used in day-to-day metadata creation. They may also lead to inconsistent interpretation when different staff members apply rules in slightly different ways.

This creates a clear need for guidance that is not only understandable to people but also structured in a way that machines can process. Modern metadata workflows require local rules that are consistent, transparent, reusable, and machine actionable. If institutions want to benefit from automation and generative AI while maintaining control over quality and local policy, their cataloguing guidance must evolve accordingly.

The method described in this article was developed using the Microsoft Power Platform. Power Platform is a low-code suite of tools that enables organisations to build applications, automate workflows, analyse data, and create intelligent processes without the need for extensive traditional software development. It includes tools such as Power Apps, Power Automate, Power BI, and AI Builder, which together make it possible to design flexible internal solutions quickly and efficiently[1].

The starting point for this project was retrospective cataloguing. Large volumes of legacy material still required catalogue records, and manual processing alone would have taken considerable time and resources. To address this, a Document Processing Model was implemented to extract metadata automatically from digitised sources such as title pages, scanned cards, and other structured documents. This model could identify and capture useful information including titles, names, dates, publishers, and other descriptive elements.

However, extracting metadata was only the first step. Raw data produced by automated systems does not automatically conform to local cataloguing rules, authority practices, or institutional standards. A separate tool was therefore needed to transform, validate, and enrich the extracted metadata so that it aligned with local cataloguing practices and could be imported reliably into the catalogue. Without this additional layer, automatically generated metadata would risk inconsistency, reduced quality, and limited usefulness.

The project therefore developed beyond simple extraction into a broader metadata governance solution: one that bridges the gap between automated data capture and

[1] For more guidance see Lowagie H. (2025) *AI Powered Cataloguing*. Facet Publishing.

professional cataloguing standards. This became the foundation for a system in which local rules could guide both human cataloguers and machine processes, ensuring that efficiency gains did not come at the expense of accuracy or coherence.

2. The problem with traditional cataloguing documentation

For many years, local cataloguing guidance has been created primarily for human readers. Institutions have traditionally relied on printed manuals, internal policy documents, PDF handbooks, spreadsheets, Word files, and static intranet pages to explain how metadata should be created and maintained. These resources are often detailed and valuable, reflecting years of professional expertise and carefully developed local practice. They help cataloguers understand institutional preferences, interpret national and international standards, and apply rules consistently in day-to-day work. However, although these formats are readable and useful for staff, they are not designed for machines. A PDF or written manual may clearly explain a rule to an experienced cataloguer, but a computer system cannot easily interpret the same information in a reliable and structured way. Rules written in paragraphs of prose may contain exceptions, conditions, and nuanced decisions. Another significant challenge is the maintenance of multiple versions of local documentation. At KBR, we had for example a French version and a Dutch version that could differ. But we also accumulated guidance in several places: a master manual, separate departmental notes, training slides, older policy files, email instructions, and informal updates shared among staff. When rules change, every version ideally needs to be reviewed and amended. In practice, this can be slow and difficult to manage. Some documents may be updated immediately, while others remain outdated, creating uncertainty about which version is current. Staff may unknowingly rely on superseded instructions simply because they are easier to find or more familiar.

This fragmentation can lead to inconsistent interpretation by people. Even highly experienced cataloguers may apply a rule differently if wording is unclear or if several versions of guidance exist. New staff members may find local practice difficult to navigate, particularly when institutional decisions are scattered across numerous files. To address this, we moved away from separate PDF documents and created a single central HTML page that brings together all cataloguing guidance in one place, available in both French and Dutch, making access simpler, clearer, and easier to maintain.

Although moving from PDF documents to a central HTML page was an important improvement, HTML still has limitations. At its core, an HTML page is primarily a text-based format organised into headings, paragraphs, lists, and links. It is excellent for presenting information clearly to human readers, but it is not ideal as a structured rules engine for machines. The content may be easier to read and maintain than scattered PDF files, yet the logic behind cataloguing decisions often remains embedded in narrative text rather than expressed in a form that systems can readily interpret and apply. This became especially clear when I began experimenting with

prompts for generative AI tools. In order to guide the AI correctly, I repeatedly found myself copying and rewriting large sections of the same cataloguing guidance that already existed on the HTML page. Rules, exceptions, preferred wording, local decisions, and examples all had to be restated in prompts so that the system could understand what was required. Very quickly, it became obvious that this approach was inefficient and unsustainable. It created unnecessary duplication and increased the risk that the guidance used by staff and the instructions given to AI tools would gradually diverge. I wanted to unify both environments and avoid maintaining two parallel versions of the same cataloguing rules: one version written for human cataloguers and another rewritten for artificial intelligence. If a local rule changed, it should only need to be updated once. Both staff guidance and machine workflows should then reflect the same decision immediately and consistently. This need for a shared and maintainable foundation led to the development of what we call the triangle model.

3. The triangle model

The triangle model is built around the principle of a single authoritative source that can serve multiple audiences and systems at the same time. Rather than creating separate documentation for people and separate instructions for machines, the model uses structured data at the centre, from which different outputs can be generated according to need. At the heart of the model is a JSON file that acts as the single source of truth[2]. JSON is a lightweight structured data format that is both human-readable and machine-processable. It allows cataloguing rules to be stored in a clear and modular way, including rule names, descriptions, conditions, examples, language versions, exceptions, and workflow notes. Because the information is structured, it can be reused consistently across multiple platforms.

The first output of this central JSON source is the HTML interface used by staff. Instead of manually writing and updating static web pages, the HTML view is generated dynamically from the JSON content[3]. This creates a clear staff-facing interface where cataloguers can browse rules, search guidance, and consult examples in a familiar web environment. It also ensures that the information visible to staff always reflects the latest approved version of the rules.

The second output is prompt-ready structured data for AI systems and other automated tools. The same JSON content can be passed directly into prompt workflows, retrieval systems, or metadata processing applications. This means that AI tools no longer rely on loosely copied instructions or manually assembled prompts.

[2] The KBR's JSON file is available on github: https://github.com/kbrbe/cataloging-guidelines/blob/main/guidelines.json

[3] The human-readable interface is available here: https://kbrbe.github.io/cataloging-guidelines/cataloging-guidelines.html. The HTML and the scripts we use to fetch the information from the JSON and create the human-readable interface can be accessed via the site's source code or at https://github.com/kbrbe/cataloging-guidelines/blob/main/cataloging-guidelines.html.

Instead, they receive accurate, current, and institutionally approved guidance drawn from the same source used by staff.

The strength of the triangle model lies in its simplicity: one structured core, multiple reliable outputs. JSON provides the logic and governance layer, HTML provides the human interface, and AI-ready data provides the machine interface. This reduces duplication, improves consistency, and makes future development far easier.

4. Wider implications for libraries

The approach described here has significance far beyond a single institution. While it was developed to solve local cataloguing and metadata management challenges, the same principles could be applied across libraries, archives, museums, and other cultural heritage organisations. If adopted more widely, machine-readable cataloguing rules could help move the sector towards a new level of collaboration, consistency, and interoperability.

One of the most promising outcomes is the potential for cross-institutional harmonisation. Libraries have long worked together through shared standards, cooperative cataloguing, authority control, and record exchange. However, collaboration has often focused primarily on sharing the metadata itself rather than sharing the logic behind how that metadata was created. Institutions may use similar standards while interpreting them differently through local policy statements, workflows, and cataloguing decisions. As a result, records can appear compatible on the surface while still reflecting significant differences in practice. A machine-readable rules framework offers the possibility of sharing not only data, but also the rules, decisions, and structures used to create that data. This is an important shift. If institutions can exchange metadata together with structured descriptions of the cataloguing logic behind it, receiving systems gain a clearer understanding of how records were produced, what local choices were made, and how those choices relate to broader standards. In this sense, transparency becomes as valuable as the metadata itself. This contributes directly to the long-standing vision of Universal Bibliographic Control (renewed in 2025) (International Federation of Library Associations and Institutions, 2025). The principle behind Universal Bibliographic Control is that bibliographic data should be created once, according to recognised standards, and then shared for the benefit of all. In practice, this goal has always been challenged by local variation, language differences, policy interpretations, and technological limitations. Machine-readable cataloguing rules offer a new pathway forward by making local variation explicit, structured, and easier to align across institutions. It could also support deeper integration with RDA Toolkit and institutional policy statements. Many organisations already maintain local application profiles of Resource Description and Access. If these policies were expressed in machine-readable formats, they could be connected more directly within the RDA Toolkit, local systems, training environments, and AI-assisted tools, reducing duplication and making updates easier to manage.

This model also makes a shared used of Artificial Intelligence tools possible. Rather than every library needing to build entirely separate systems, it becomes possible to imagine a common AI-enabled cataloguing platform that can be refined locally through institution-specific rulesets. In other words, one shared system could operate with local customisation. A library in Belgium, for example, could apply bilingual authority practices, while another institution elsewhere could apply its own language, descriptive traditions, or subject conventions using the same technical framework. This creates economies of scale while preserving local identity and professional standards. To make this possible, institutions could collaborate on a unified, machine-readable structure for expressing cataloguing rules. Formats such as JSON, persistent identifiers such as Uniform Resource Identifier, and linked data models could be used to represent rules in a consistent and reusable way. If the underlying structure is shared, individual institutions can adapt the content without losing interoperability.

5. Conclusion

The motivation for redesigning local cataloguing rules arises from both practical and strategic needs. On a practical level, libraries and other cultural heritage institutions need workflows that reduce repetitive manual tasks, improve consistency in metadata creation, and make policy updates easier to manage. Metadata infrastructures must now be capable of supporting future developments in automation and AI-assisted cataloguing. Library systems are no longer isolated local environments; they operate within increasingly connected networks of discovery platforms, data exchange services, repositories, and external knowledge systems. In this context, local rules that exist only in static documents are difficult to integrate into modern workflows. This article proposes a solution by making the cataloguing rules structured and ready to use in human and automated environments. The objective is to move towards a more sustainable and future-ready approach to cataloguing documentation. They move local policy from static reference files into active operational frameworks that can support both human expertise and technological innovation.

There is a strong case for collaborative development of interoperable, machine-readable rules across the sector. No single institution should need to solve these challenges in isolation. Shared models, common structures, and reusable approaches would benefit libraries of all sizes and create stronger foundations for collective progress. Now is the right time to act. Artificial intelligence is developing quickly, system expectations are changing, and institutions are already reconsidering how metadata work will be organised in the coming years. If libraries take the initiative now, they can shape these developments according to professional values of quality, transparency, access, and cooperation rather than adapting later to systems designed without them in mind.

References

International Federation of Library Associations and Institutions (2025) *IFLA Professional Statement on Universal Bibliographic Control*. Available at: https://repository.ifla.org/handle/20.500.14598/4259 [Accessed: 10 August 2026]

Lowagie, H. (2025) *AI Powered Cataloguing*. Facet Publishing.

Lowagie, H. and Van Woensel, J. (2025) *Royal Library of Belgium (KBR) - Bilingual Cataloging Guidelines* (version 0.1.0). Available at: https://github.com/kbrbe/cataloging-guidelines and https://doi.org/10.5281/zenodo.15018951 [Accessed: 18 August 2026]