

Considering AI in the growth of RDA

Introduction

Libraries are at the forefront of information management and access. As the volume and variety of resources continue to grow, the need for efficient and accurate cataloging becomes increasingly crucial. Artificial Intelligence (AI) presents an opportunity to address these challenges, offering new ways to streamline processes and improve the quality of library catalogs. Meanwhile, we need to consider implementing AI responsibly and develop strategies to mitigate potential harms, such as hallucinations, bias, privacy issues, and other ethical concerns, in alignment with professional values. This approach is consistent with the principles outlined by organizations such as the IFLA Statement on Libraries and Artificial Intelligence¹, the Association of Research Libraries², IFLA's Professional Codes of Ethics for Librarians³ and the ALA Code of Ethics⁴.

AI can be used to streamline repetitive tasks, enhance the accuracy of resource description, and improve service delivery for resource-limited institutions. However, it is important to dispel misconceptions that AI will directly lead to cost savings through staffing cuts. Instead, AI should be viewed as a tool to complement human expertise, allowing catalogers to refocus their efforts on more intellectually stimulating tasks that require critical thinking and nuanced decision-making.

RDA is an important international metadata standard and the RDA Board and RDA Steering Committee have a responsibility to consider the implications of the use of AI in cataloging and to provide guidance for implementation in a way that encourages standardization in metadata while holding space for upholding our own communities' core values. Furthermore, the RDA Board and RDA Steering Committee would like to collaborate with other international library organizations, such as IFLA, PCC, and AI4LAM, to responsibly implement AI in cataloging worldwide.

AI as a Tool for Cataloging: Enhancing, Not Replacing Human Work

Cataloging in libraries often involves routine, repetitive tasks that consume much of a cataloger's time, such as assigning standardized metadata, updating records, and maintaining authority control. AI technologies, particularly machine learning and natural language processing (NLP), have the capacity to take on many of these tedious tasks, allowing catalogers to focus on work which at this point requires human intervention.

For example, AI can assist in the automated tagging of metadata by learning from existing catalog records and applying that knowledge to new materials. It can also aid in detecting inconsistencies in records and in maintaining authority control by flagging variations in names and terms. The key advantage of AI is its ability to process large volumes of information rapidly and consistently, freeing up catalogers to apply their expertise in tasks that require judgment, such as interpreting ambiguous resources, correcting incorrect subject terms, developing controlled vocabularies and thesauri, cross-referencing and linking

related resources, and working on the development and implementation of cataloging guidelines.

However, AI is not a substitute for human intelligence in cataloging. The success of AI tools is contingent on the quality and specificity of the training data they are provided with. AI can make automated suggestions, but human oversight is essential to ensure accuracy and cultural context, especially in diverse and interdisciplinary collections. Therefore, catalogers will remain indispensable as overseers, interpreters, and innovators, working alongside AI rather than being replaced by it.

AI and Large Language Models (LLM) need to be trained to create RDA-compatible metadata and to parse the massive amounts of metadata which they will encounter. While neither the RSC nor the Board have the capacity or mandate to validate metadata created by agencies worldwide, the RSC and RDA Board should consider making the [minimum description of a resource](#) guidance available in a form which can be used to train AI and LLM. Making this set available to machines has been discussed for some time and has gained in urgency due to the possible future use of AI, but is also needed for cataloging large quantities of resources that can no longer be managed intellectually.

Misconception: AI Implementation Will Lead to Cost Savings through Staffing Reductions

A common misconception is that AI adoption will result in cost savings by reducing the need for human catalogers. However, historical trends in library technology show that new tools often create more roles for staff rather than fewer. The introduction of online catalogs, digital repositories, and discovery systems did not eliminate the need for librarians; rather, it shifted their responsibilities, requiring new skills and expertise in technology management, digital preservation, and user engagement.

Similarly, AI will not reduce the need for catalogers and metadata librarians but will change the nature of their work. As AI handles the more mechanical aspects of cataloging, staff can take on new roles, such as managing AI systems, fine tuning training data, refining metadata schemas, or developing new cataloging practices for digital and emerging formats. Libraries will continue to need skilled professionals who can make critical decisions that AI cannot—such as navigating ambiguous metadata situations, ensuring inclusivity in resource description, and responding to the evolving needs of library users.

Furthermore, implementing AI is not a cost-neutral endeavor. It requires investment in infrastructure, software, training, and ongoing maintenance. Therefore, AI is unlikely to generate immediate cost savings, and it should not be seen as a way to justify staffing reductions. Rather, the goal should be to enhance the capabilities of existing staff, enabling them to perform higher-level tasks that machines cannot do.

AI as a Tool for Libraries with Fewer Resources

We align with the principles in 'Policy and Investment Recommendations for Trustworthy AI,'⁵ which emphasize that AI literacy initiatives should be accessible to all. We aim to apply this principle to metadata operations to benefit less advantaged groups.

Libraries with limited resources, such as smaller libraries with sparse budgets, stand to benefit significantly from AI assistance. AI can help minimize errors in transcribed metadata, even when resources for quality control are limited.

By combining AI with human expertise, libraries can achieve greater operational efficiency and improve the overall user experience without needing to significantly expand their staff. However, the successful implementation of AI in these contexts will depend on careful integration with existing workflows and the ongoing involvement of skilled catalogers, from cataloging organizations or communities as part of the support infrastructure.

Large Language Models and the Importance of Clear Guidelines for Metadata Training

The effectiveness of AI in cataloging depends heavily on the quality of the data used to train the system. Large Language Models (LLMs), a subset of AI that excels at tasks like natural language understanding and generation, are most effective when trained with clear guidelines and structured metadata. We believe that the closer the training data adhere to well-defined, machine-ingestible standards, the better the AI will perform in automating cataloging tasks, in accordance with best practices for large language models, such as OpenAI's GPT guidelines for fine-tuning and training data.⁶

For example, metadata sets that follow explicit, standardized formats such as RDA provide a strong foundation for training AI models. By feeding these models high-quality, standardized metadata, AI tools can learn to generate more accurate and reliable cataloging data. Conversely, when metadata is inconsistent or poorly defined, AI systems struggle to make meaningful inferences, leading to suboptimal results.

To maximize the potential of AI in cataloging, libraries should invest in creating and maintaining well-structured metadata sets that adhere to established standards. Moreover, clear guidelines should accompany the training corpus to ensure that the AI is interpreting and applying metadata in the intended way. The Semantic Web, and especially the metadata

provided by well-trained librarians, can ensure that AI technologies provide us with more reliable (verified/controlled) data (instead of providing answers based solely on pattern recognition technologies, statistical algorithms, etc.). Collaboration between cataloging experts and AI developers will be required to refine and optimize AI tools for the specific needs of the library community.

The role of RDA

AI presents unique challenges and opportunities for RDA.

AI is likely to introduce new ways of interacting with digital materials and emerging data types. The RSC will have to deal with the implications of AI generated content within the constraints of its theoretical underpinning, IFLA-LRM.

LLMs are most effective when trained with clear guidelines and well-defined metadata sets. The closer the training data adheres to explicit, machine-ingestible standards, the better these models perform in describing and categorizing resources. The RSC will have to ensure a comprehensive and explicit set of rules that can be used to train these systems to better understand and catalog resources.

AI can assist in identifying new forms of resources that may require updates to the RDA framework; there will need to be a mechanism to return this information to the RSC for consideration.

The implementation of AI and other machine assisted cataloging applications by the major library system vendors may open new vistas for implementation of RDA and for the switch from the Original Toolkit to the Official. The RDA Board will need to prepare for AI and LLMs ingesting the guidelines and vocabularies outside the standard subscription framework while creating a pathway for smaller libraries to adopt Official RDA in a lighter, more feasible, AI-based implementation.

References

1. IFLA Statement on Libraries and Artificial Intelligence, <https://repository.ifla.org/server/api/core/bitstreams/8d21df6c-e61b-41a5-b089-0353a5552dbd/content>
2. Henry, G., 2019, "Research Librarians as Guides and Navigators for AI Policies at Universities." *Research Library Issues*, no. 299 (2019): 47–65, <https://doi.org/10.29242/rli.299.4>
3. IFLA Professional Codes of Ethics for Librarians, <https://www.ifla.org/g/faife/professional-codes-of-ethics-for-librarians/>
4. ALA Code of Ethics, <https://www.ala.org/tools/ethics>

5. Policy and Investment Recommendations for Trustworthy AI,
https://www.europarl.europa.eu/cmsdata/196378/AI%20HLEG_Policy%20and%20Investment%20Recommendations.pdf
6. OpenAI Developer Platform documentation,
<https://platform.openai.com/docs/overview>