

AI Metadata Enrichment for Libraries

As the scope of cataloging, as well as the depth and complexity of cataloging standards evolve and grow, Ex Libris, part of Clarivate, is working with our community to build innovative solutions that ensure content quality, at scale. We are undertaking initiatives to harness advanced technologies to assist catalogers and libraries in their quest for better data and discoverability.

Libraries across the world, as well as content providers and aggregators, are managing vast volumes of new and constantly changing content, at a scale that is almost humanly impossible to maintain. The varied catalogs of many libraries offer additional challenges, leading catalogers to catalog subject matters in which they may not be experts and that require extensive research, on top of the challenges offered by the growing amounts of resources and resource types. At the same time, metadata quality continues to be a critical piece in collection management and development, as well as discovery and patron services, and therefore a necessity for all libraries to uphold. Various resource types such as images, sound recording and videos often require even more time and effort to catalog and expose in discovery, as they cannot be easily handled by traditional tools.

Using AI technology enables us – Ex Libris and our community members – to ensure that records contain the relevant metadata needed by the different stakeholders in the library, from assessing the collection and collaborating with partner libraries in resource sharing and shared print initiatives, to providing the best possible services for patrons.

AI Metadata Enrichment Principles

Ethics and Fair Use

We take great care to make sure that our metadata enrichment not only conforms to legal requirements, but also to ethical considerations. Our main focus areas for this are:

Respecting Copyrights

Our plans include supporting enrichment for various resource types, focused on allowing the library to use its own collections to generate the metadata.

Our enrichment processes for community records are only performed on materials for which we have the rights to do so, according to licenses and agreements with vendors.

Protecting Proprietary Information

The information shared with the AI Metadata Assistant is used only to generate the metadata required for the library's needs. It is not used to train AI models, and is not saved by us for later use.

Furthermore, the metadata returned by the AI is filtered to prevent accidental use of proprietary data such as system numbers or local fields.

High Quality of Metadata

The measures taken to maintain metadata quality are evaluated, assessed and refined in collaboration with the Alma community.

The available enrichments highly depend on many factors, such as the type of resource, the data provided to the AI, the requested information and formatting, and whether the data validation is automated or done by a cataloging expert. Clarivate is putting a lot of effort into assessing these differences and tailoring solutions for libraries' needs, that will ensure getting the most out of the AI capabilities.

In addition, the metadata generated by the AI Metadata Assistant is filtered and validated using automated tools, such as checking it against the authority vocabularies available in our platform, to maintain a high quality of data.

Leaving Control in the Library's Hands

The choice of whether or not to use the metadata suggested using AI is the library's. Whether by adding review and correction capabilities to enrichment workflows, or marking the enriched fields in community records for easy identification – we design the AI Metadata generation and enrichment processes to provide the library and its catalogers as much control over the use of this data as possible.

AI Metadata Assistant Plans

Enriching Community Records

In February 2024, an AI-based metadata generator for select bibliographic records in the Alma Community Zone was launched, to improve record quality, making them more discoverable and accessible for all who use the library management system.

Our AI metadata generator is live and enriches select ProQuest EBook Central records, by adding language, summary, and subject headings fields in alignment with the Library of Congress standards. These additions are mentioned in a Source of Description Note detailing the AI enrichment, to allow libraries to search for the enriched records. Our plans are to grow both the scope of records and the number of fields generated by the innovative AI metadata generator.

More information is available in the [Community Center Knowledge Article](#) and in the [Metadata Enrichment using AI 2024 Content Webinar](#).

588	__ a Part of the metadata in this record was created by AI, based on the text of the resource.
520	__ a Population Health in Challenging Times: Insights from Key Domains: Proceedings of a Workshop' is a collection of discussions about population health from a 2020 workshop held by the National Academies of Sciences, Engineering, and Medicine in the United States. This workshop focused on two major crises of 2020: the COVID-19 Pandemic and racial injustice. It aims to help health professionals, policy makers, and researchers by informing policy decisions and increasing understanding in science, engineering, and medicine. \$\$7 Generated by AI.
650	__0 a Public health. \$\$7 Generated by AI.

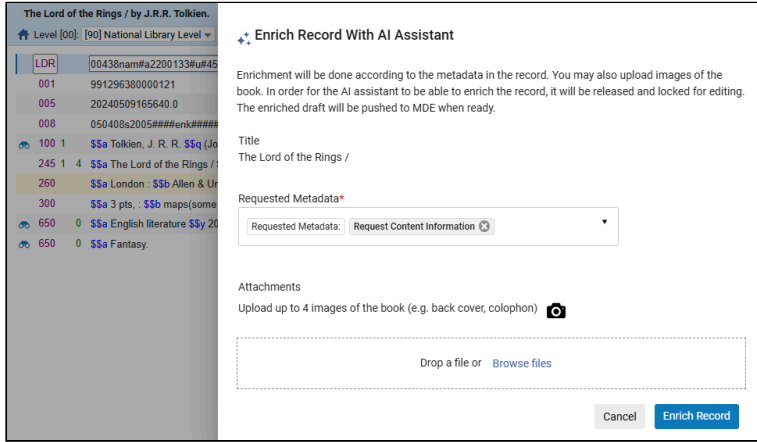
AI enriched data in community records is clearly marked for easy identification by libraries

Assisting Catalogers' Workflows in the Metadata Editor

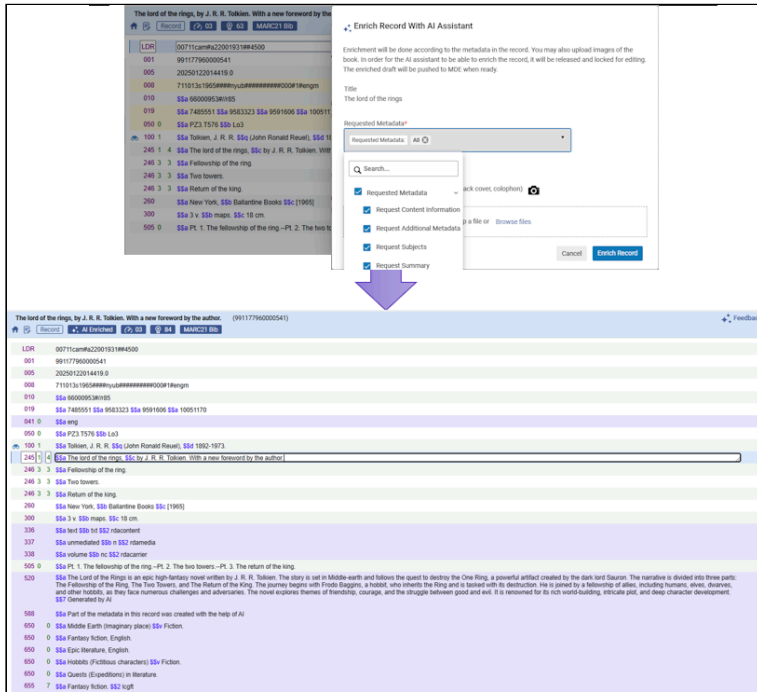
The AI Metadata Assistant in Alma's Metadata Editor helps catalogers in their work by suggesting metadata they can use

when cataloging a resource, saving time and effort on researching and searching information and freeing catalogers' time to handle more complex cataloging tasks.

The cataloger can easily create a new bibliographic record or enrich an existing brief record using the resource's metadata (such as title, author, content note, etc.) and/or images (such as a book's back cover, title page, or title verso). After receiving the AI's suggested metadata generated from the provided information, they're able to accept, correct or discard the suggested changes.



Alma's AI Metadata Assistant enrichment workflow is embedded in the Metadata Editor



AI generated metadata is clearly marked for the cataloger to review and accept, correct or reject



Creating a new record using the AI Metadata Assistant in Alma's Metadata Editor - cataloger can easily distinguish AI generated metadata from fields coming from the library's template

Catalogers with the AI Assisted Cataloging role can also use the Alma Mobile 2 app (which allows librarians to use a mobile device camera to process library resources) to easily take pictures of library resources and send them to the AI Metadata Assistant for processing. Once the assistant completes processing the images, the draft containing the AI assistant's suggested metadata is pushed to the Metadata Editor for the cataloger to review and accept, correct or dismiss it.

Today, the AI Metadata Assistant helps with creating English MARC 21 records, using Library of Congress subject headings. Support for more languages, vocabularies and formats will be gradually added as we collaborate with our global community to ensure quality metadata generation for more varied resources.

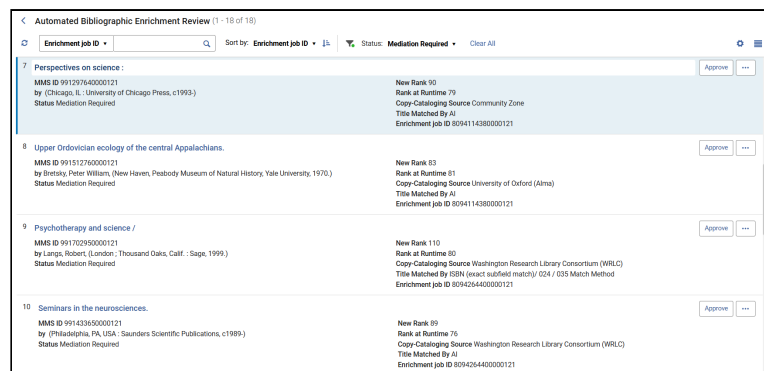
For more information, see Alma's roadmap plans:

- [AI Metadata Assistant - expand to additional languages - early access](#)
- [AI Metadata Assistant - expand to additional subject vocabularies - early access](#)

Streamlined Copy-Cataloging Workflows Using Automation and AI

Today, catalogers can search in search profiles such as partner libraries' catalogs, publicly available catalogs (e.g. Library of Congress), national catalogs (e.g. Libraries Australia) and other cataloging sources (e.g. WorldCat) for bibliographic records similar to their library's resource, review the returned results to locate a matching record and copy it or merge it with their existing record. This is manual process, where the librarian generates or fills in a form including information such as title, author, identifiers and publication year, selects the search profile to use, goes over the returned results and selects a record to copy.

Using filtering and AI title matching processes, a new Alma job will allow catalogers to copy-catalog a set of brief bibliographic records, and Alma will use the search results of the selected profile to locate matching records. The cataloger will be able to review a suggested matching record and decide whether to merge it with their local record or not.



The cataloger reviews the suggested match and approves or rejects the merge

For more information, see Alma's roadmap plans:

- [Bulk Bibliographic Records Enrichment with Automated Copy-Cataloging and AI Title Matching - General Availability](#)

Easily Generate Brief Records for Uncataloged Backlog Items

Many libraries struggle with a backlog of uncataloged physical resources - resources the library owns, but cannot easily expose to patrons or provide services for. Alma will provide AI-based tools that will allow libraries to easily create brief

records for such uncataloged resources, with the option to review them before displaying in discovery. These tools will help libraries improve the awareness and use of their collections.

For more information, see Alma's roadmap plans:

- [AI Metadata import - early access](#)