Alma Al Metadata Assistant testing at the University of Leeds

Alison Hazelaar, Access & Acquisitions Manager: Metadata & Discovery

Ceilan Hunter-Green, Metadata & Discovery Manager

Alice Jones, Metadata & Discovery Coordinator

Zoe Marshall, Metadata & Discovery Coordinator



Outline

- Who we are
- How we got involved
- How it works
- How it progressed
- Findings and thoughts
- What's coming next



Alma Al Metadata Assistant testing at the University of Leeds

University of Leeds

- In Leeds, West Yorkshire
- Established as Leeds School of Medicine in 1831 and Yorkshire College of Science in 1874 which merged in 1884. Formally established as University of Leeds in 1904
- Eleventh largest university in the UK
- Part of the Russell Group of researchintensive UK universities
- Over 39,000 students from 170 different countries and over 9,000 staff of 100 different nationalities





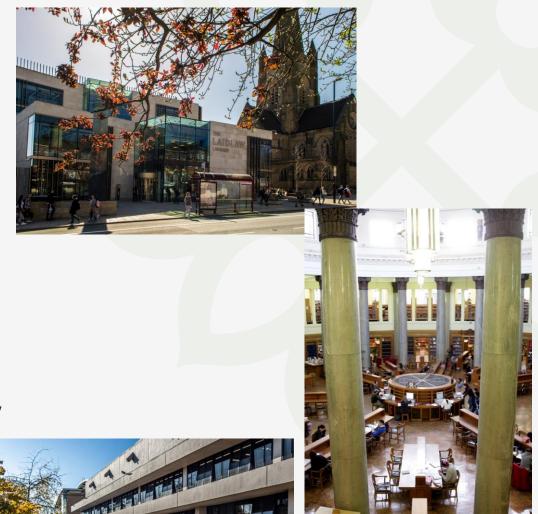


University of Leeds Libraries

- Five libraries with 2.7 million items
- Over 2.5 million annual visits
- White Rose Research open access repositories for papers, theses and data from WR universities (Leeds, York, and Sheffield)
- Five designated collections deemed internationally

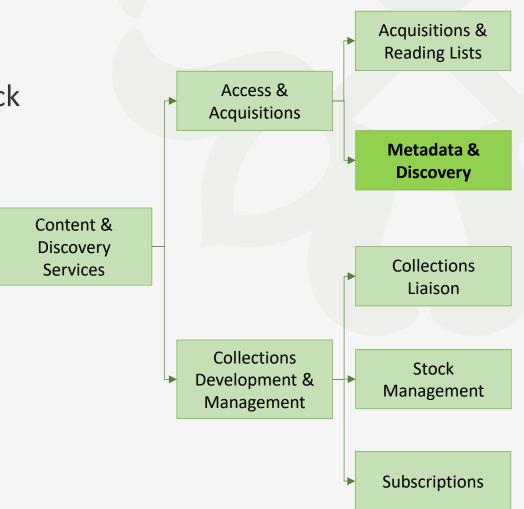
important

- Migrated to Alma in 2019
- Primo VE and a Premium Sandbox



University of Leeds Metadata & Discovery team

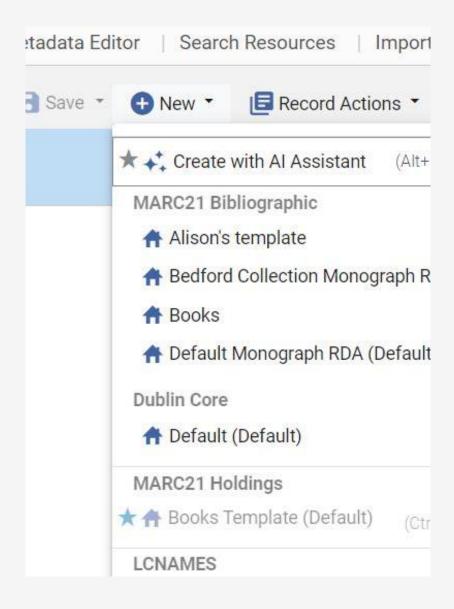
- Fourteen staff members
- Work alongside Acquisitions & Reading Lists, Stock Management, Collections Liaisons and Subscriptions teams under Content & Discovery Services banner
- Remit includes:
 - Print and e-book cataloguing
 - Special Collections monograph cataloguing
 - E-resource access, authentication and maintenance
 - Discovery experience development



How we got involved



The AI Metadata Assistant



The Alma AI Metadata Assistant in the Metadata Editor will reduce work needed to research and search for information, allowing the cataloguer to tackle more complex tasks, by suggesting metadata they can choose to use when cataloguing a resource. It will assist in record creation and enrichment.

Project scope

Phase 1 Development was to look at English Language MARC 21 records for books using LCSH vocabulary.

Testing phase to begin on August 18th and last approximately one month.

We had to state how many cataloguers from our institution would be taking part in the testing.

Testing done in Premium Sandbox.

How it works

The LLM behind the AI Assistant is currently GPT-40

The AI Metadata
Assistant works from prompts and is not involved in any AI training.



How it works.. New record



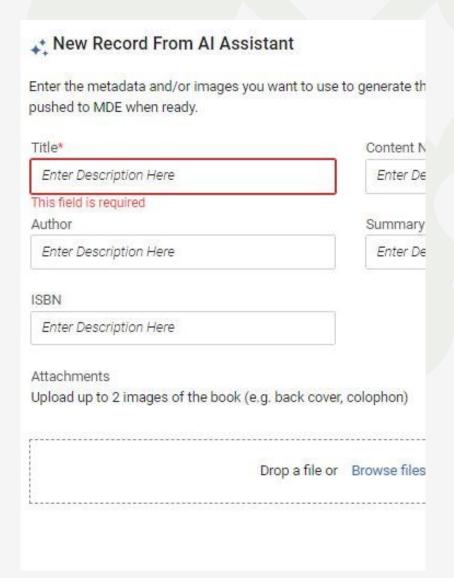
"AI ASSISTED CATALOGING"
ROLE



CONTINUE WORKING IN ALMA WHILE THE AI IS PROCESSING

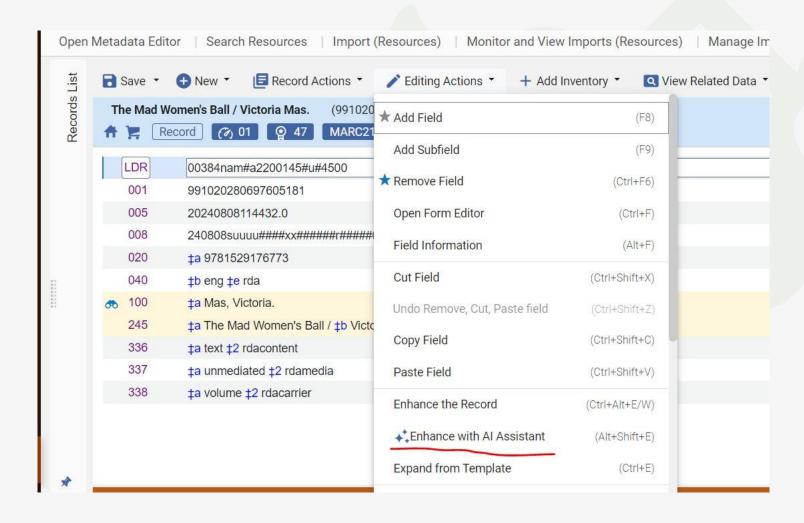


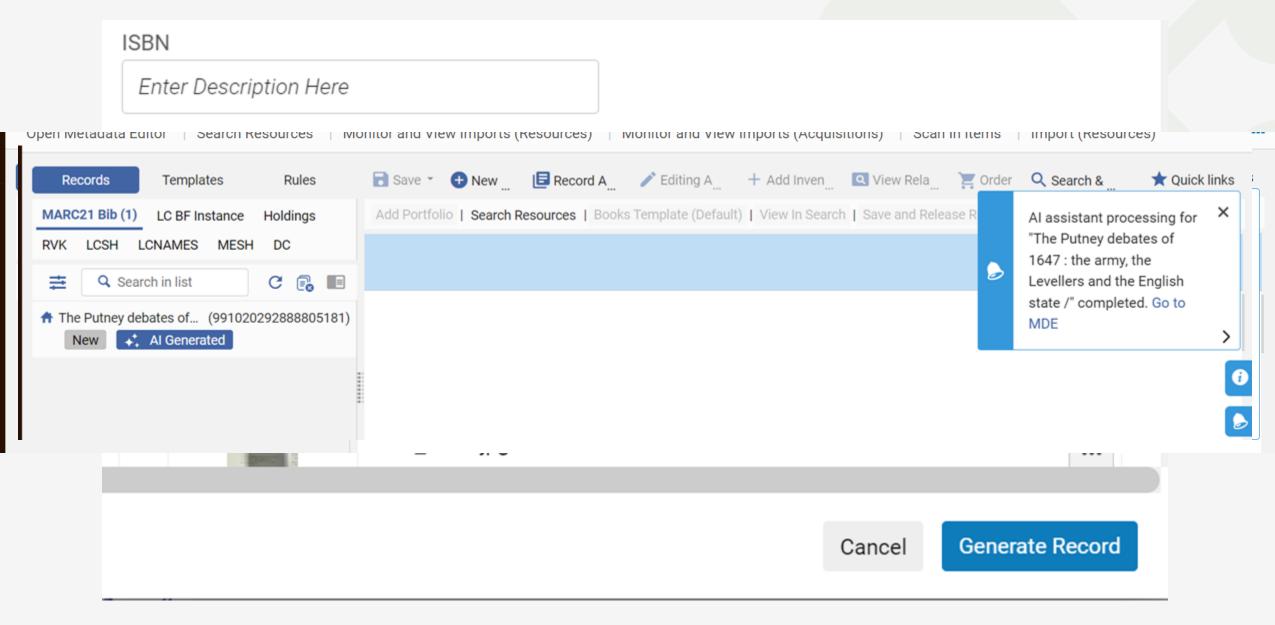
TWO BOOK IMAGES MAX



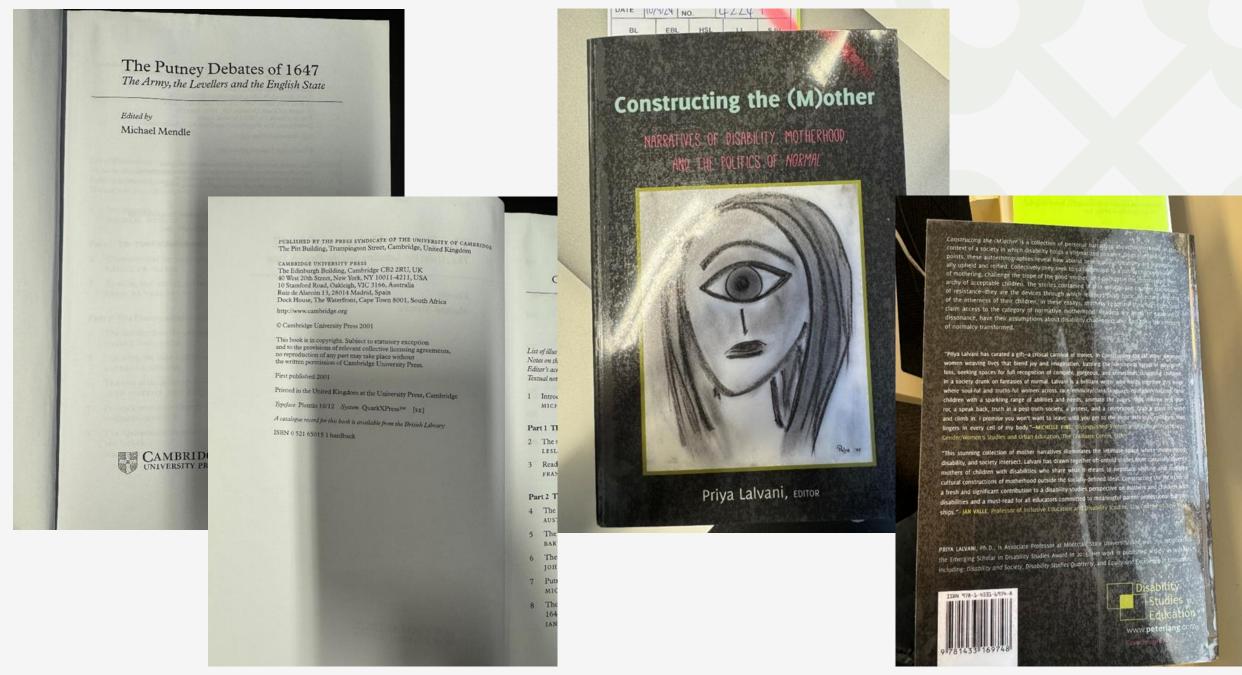
How it works.. Enhancing brief record

- Metadata taken from brief record. Add up to 2 images.
- Authorities get linked automatically.
- LCSH subjects added





Alma Al Metadata Assistant testing at the University of Leeds



Alma Al Metadata Assistant testing at the University of Leeds

Initial prompt

As a cataloger, you receive a book titled: {title} {authored by author} {with ISBN XXX} {and content or summary note: content of notes}. {One or more images of the book are included, with information on the book. It is very important if an image is attached and you cannot read the text on the image, return "could not process image".}

Your task is to use this information to create a Marc 21 record for this book. This task consists of several subtasks:

- Task 1: Include the title in field 245.
- Task 2: Include 5 Library of Congress Subject Headings that fit the topics described, including subdivision subfields if relevant.
- Task 3: A short summary of the book in at most 200 words (in field 520).
- Task 4: Any other relevant bibliographic data that is available should be included in its appropriate field. For example:
- If you have information about the ISBN of the book, return it in field 020 if the book has multiple ISBNs, return each one in a separate 020 field.
- If you know the language the book is in, add 041 field with the Language code in 3 alphabet letters.
- If you have information about the author of the work, include field 100 with this information (if you don't, don't include this field with generic data such as 'author name').
- If you have information on the edition, include field 250 with this information.
- If you have publishing data (such as publisher, publication year) include field 264.
- If you have information about another person that contributed to the work, include field 700 with this information.
- If you can tell that the work is part of a series, include field 830 with the series name.
- If you're not sure which field to include certain information in, use 500.
- Don't return filler data, such as "unknown" or "not defined".

First prompt update

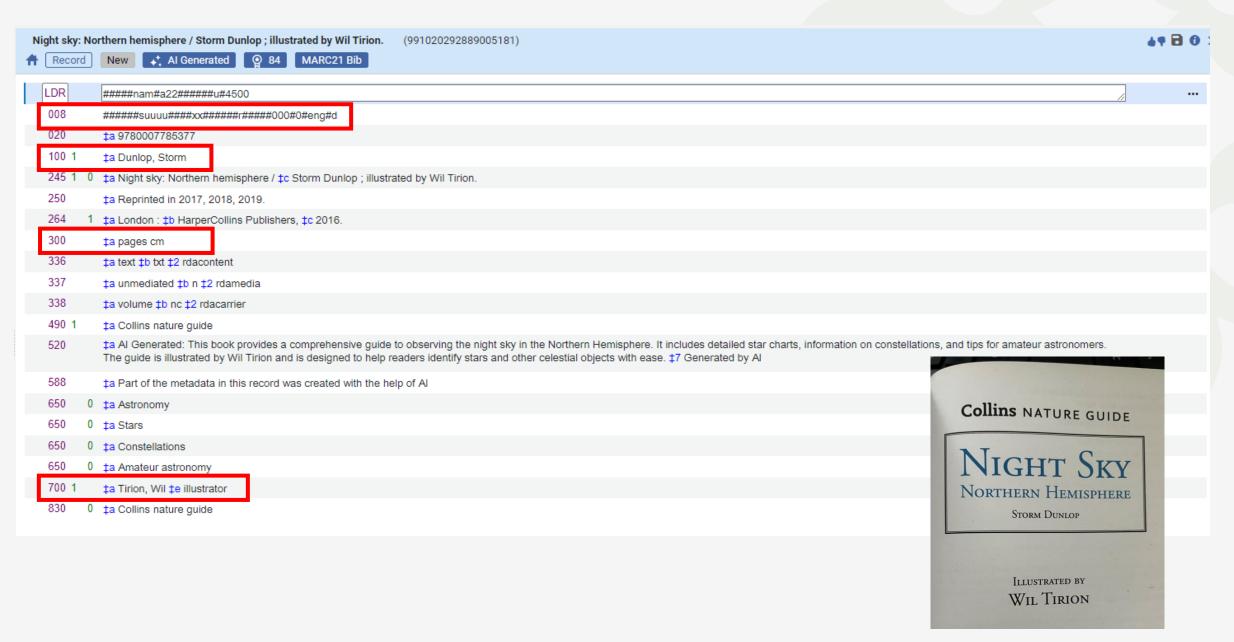
- Your task is to use this information to create a Marc 21 record for this book, with correct ISBD punctuation. This task consists of several subtasks:
- Task 3: A short summary of the book in field 520. If you know the abstract provided by the
 publisher, include the full abstract as a quote with double quotation marks, prefixed with
 "Publisher abstract:". If not, generate a summary of at most 200 words.
- If you know the LC call number of the book, include 050. {*}
- If you know the RDA content type, include field 336.
- If you know the RDA media type, include field 337.
- If you know the RDA carrier type, include field 338.
- If you have information about the table to contents, include field 505.
- If you can tell that the work is part of a series, include the series name in fields 490 and 830. If you know the series ISSN, include it in subfield x.

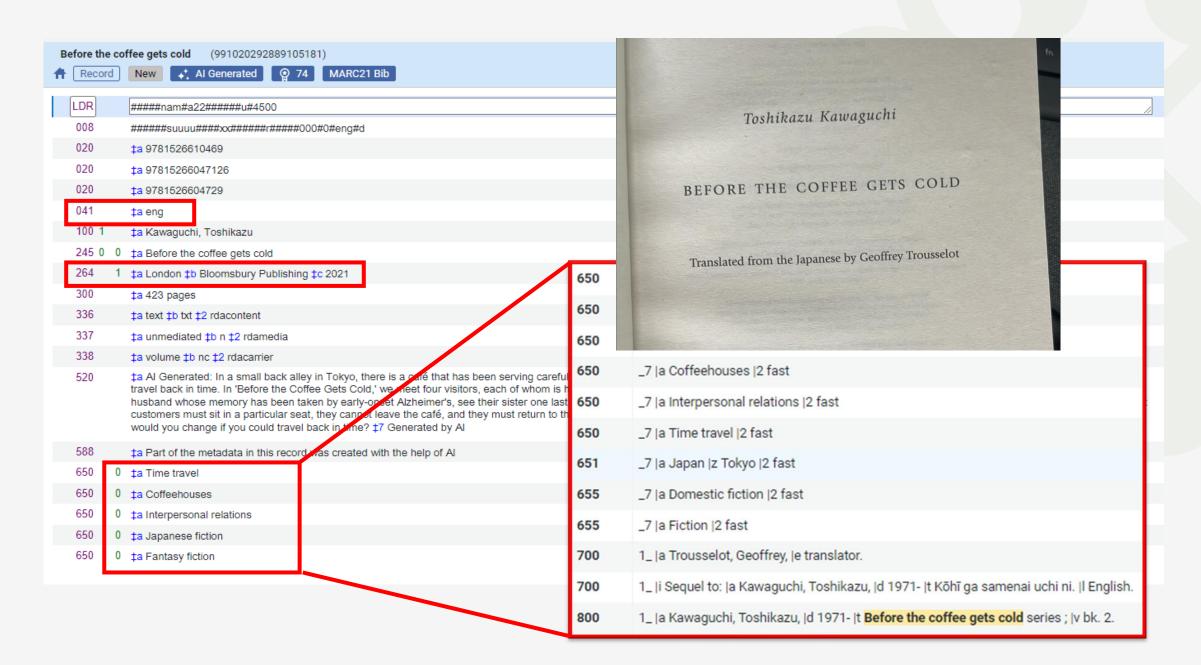
Second prompt update

- Cataloging in English
- a full Marc 21 record for this resource
- If the resource language is different than the cataloging language, the following fields, if included, should be in the resource language (make sure to include a linked 880 field containing the content converted to Latin script for any field not in the cataloging language): 245,250,264,490,830,any field starting with '1' or '7'.
- Task 1: Include the title statement in field 245, including subfields if relevant.
- Task 2: Include 5 subjects that fit the topics described, including subdivision subfields (form information in v, chronological in y, geographic in z and general in x) if relevant. The subjects should be from the list of Library of Congress Authorities.
- If you have other information about the physical description of the resource, return it in the appropriate field.

Third prompt update

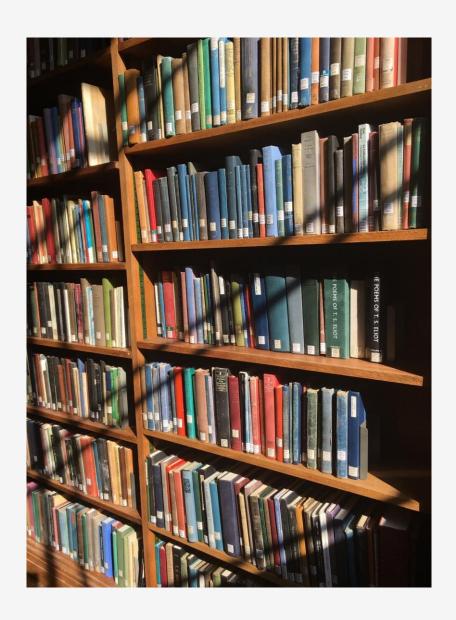
- Use this information to provide a full Marc 21 record for the resource, with correct ISBD punctuation. If the resource language is different than the cataloging language, only fields 245,250,264,490,830 and fields starting with '1' or '7' should be in the resource language (using UTF-8 characters), if included. If resource language is not in latin alphabet, include a linked 880 field after each non-latin field, containing a romanized transliteration of its contents.
- Task 3: Include field 520 in the resource language: Only if you have the publisher's abstract (from the book cover or the publisher itself) inlcude a full quote of it, exactly as the publisher wrote it without rephrasing, in double quotes (add escape characters so the quote characters are properly communicated by API) and prefixed with "Publisher abstract:". If no abstract is available, generate a concise summary of up to 200 words with prefix "AI Generated:". If cataloging language is different than the resource language, add another 520 with a concise summary of up to 200 words in the cataloging language, with prefix "AI Generated:"
- If you have information about the author or creator of the work, include the appropriate 1XX field with this information (100 for a person, 110 for a corporation, etc.). If there is more than one creator, include the first in a 1XX field, and the rest in 7XX fields. If you don't know the creator, don't include this field with generic data such as 'author name').







Looking ahead



- NOT REPLACING EXPERTS!
- Use cases: Foreign languages, Gifts and donations, No other records available

Continuation of the early access project



Dear all,

We've been working for over a month, and it's been an intense - and very useful! - project. As we still need to improve the prompt further, I would like to continue this process, perhaps with more libraries to get more perspectives.

Could you please let me know if your library would like to continue taking part in the testing, and at what capacity, so I know how many new libraries I can add to the project?

Thanks! Lili



Questions?

Alison Hazelaar <u>a.l.hazelaar@leeds.ac.uk</u>

Ceilan Hunter-Green c.hunter-green@leeds.ac.uk

Alice Jones <u>a.jones6@leeds.ac.uk</u>

Zoe Marshall z.r.marshall@leeds.ac.uk



Leeds University Library Leeds, United Kingdom LS2 9JT Library no library.leeds.ac.uk