

Assessing LLMs and Batch-Processing for MARC21 Bibliographic Metadata Generation

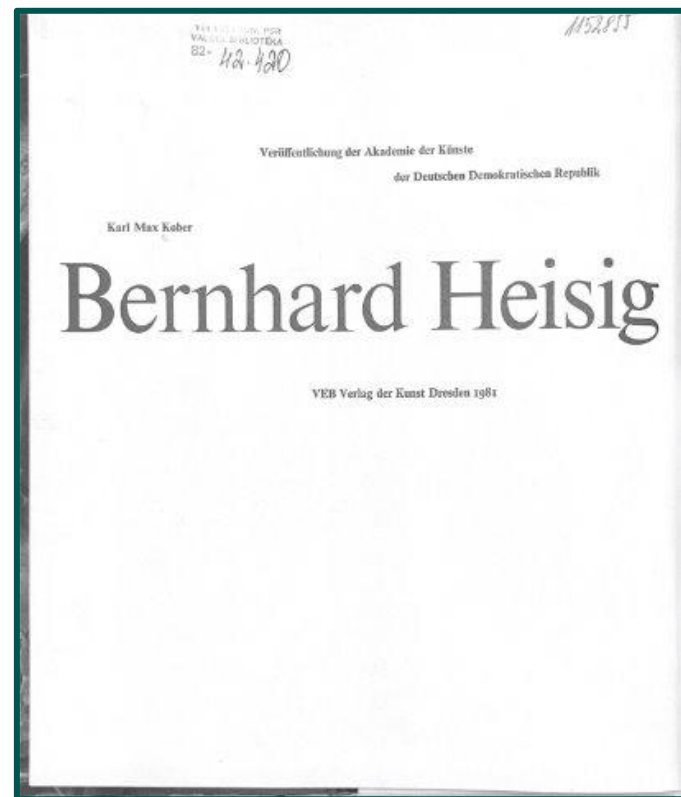
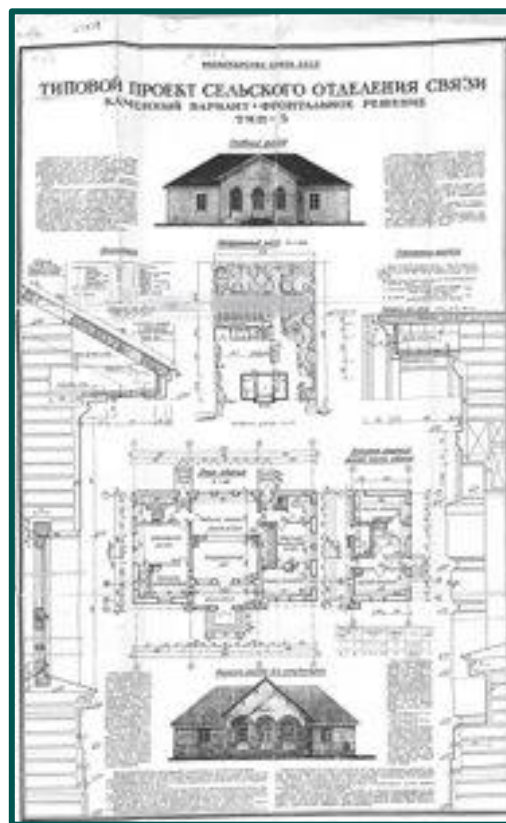
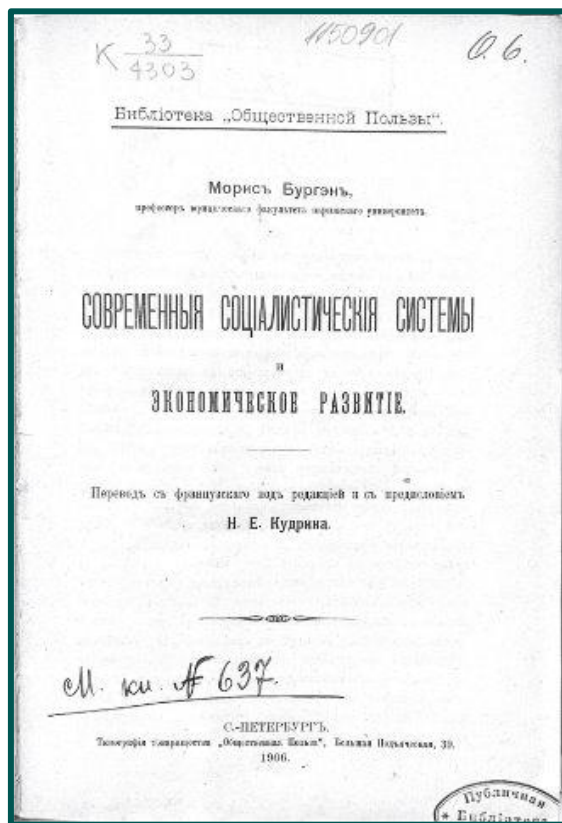
Matīss Bolšteins,

Head of Data Management Development,

Bibliography Institute, National Library of Latvia

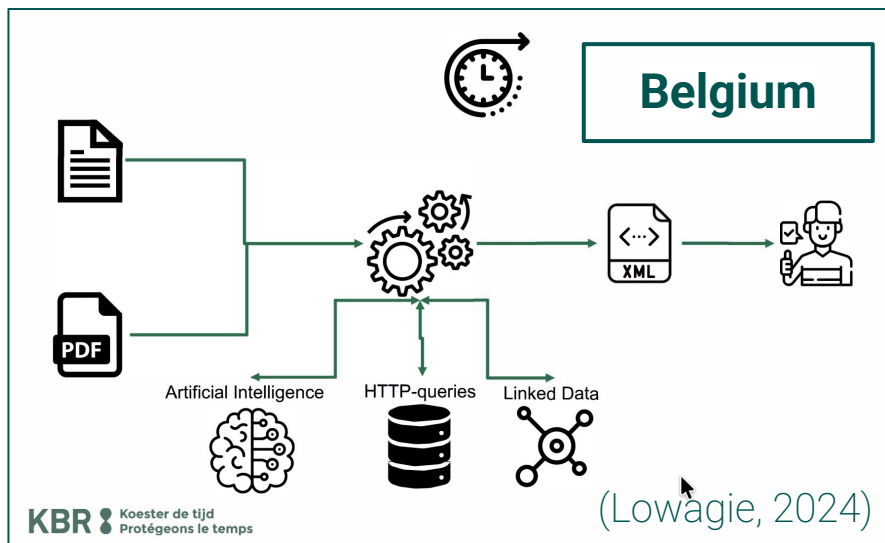
15.06.2026.

Topicality



- rapid adoption of AI-based tools for cataloguing

ALMA:AI Metadata Assistant



meteor Public

A python module and REST API for automatic extraction of metadata from PDF files

● Python ☆ 18 🔗 3

Norway

The screenshot shows a web-based cataloging interface. It features a grid of search results or metadata entries. A text box at the bottom of the interface contains the following text: "De catalogiseerder ziet in één oogopslag mogelijke matches met een matchingscore per veld. The cataloguer is shown possible matches with a matching percentage per field." The interface is labeled 'Denmark' in a box.

(Automatic Metadata Generation at the KB, n.d.)

- limited systemic evaluation of MARC21 conformity

Brzustowicz, R. (2023). From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging. *Information Technology and Libraries*, 42(3), Article 3. <https://doi.org/10.5860/ital.v42i3.16295>

Zavalin, V., & Zavalina, O. L. (2025). Are we there yet? Evaluation of AI-generated metadata for online information resources. *Information Research*, 30(iConf), Article iConf. <https://doi.org/10.47989/ir30iConf47215>

Taniguchi, S. (2024). Creating and Evaluating MARC 21 Bibliographic Records Using ChatGPT. *Cataloging & Classification Quarterly*, 62(5), 527–546. <https://doi.org/10.1080/01639374.2024.2394513>

- need for responsible, evidence-based AI implementation (Praktiskās vadlīnijas MI ieviešanai iestāžu darbā un pakalpojumos, 2025).

PRAKTISKĀS VADLĪNIJAS MI IEVIEŠANAI IESTĀŽU DARBĀ UN PAKALPOJUMOS

Vadlīnijas ir veidotas sadarbībā ar Valsts administrācijas skolas projektu "Publiskās pārvaldes digitālā akadēmija" (projekta Nr. 2.3.2.2.i.0/1/23/I/VARAM/001).

Finansē Eiropas Savienība – NextGenerationEU. Tomēr paustie uzskati un viedokļi ir tikai autora(-u) uzskati un viedokļi un ne vienmēr atspoguļo Eiropas Savienības vai Eiropas Komisijas uzskatus un viedokļus. Par tiem nav atbildīga ne Eiropas Savienība, ne Eiropas Komisija.

2025. gada oktobris



Finansē
Eiropas Savienība
NextGenerationEU



Nacionālais
attīstības plāns



Digitālā
akadēmija



Viedās administrācijas un
reģionālās attīstības
ministrija



MĀKSLĪGĀ
INTELEKTA
LATVIJAS
ASOCIĀCIJA



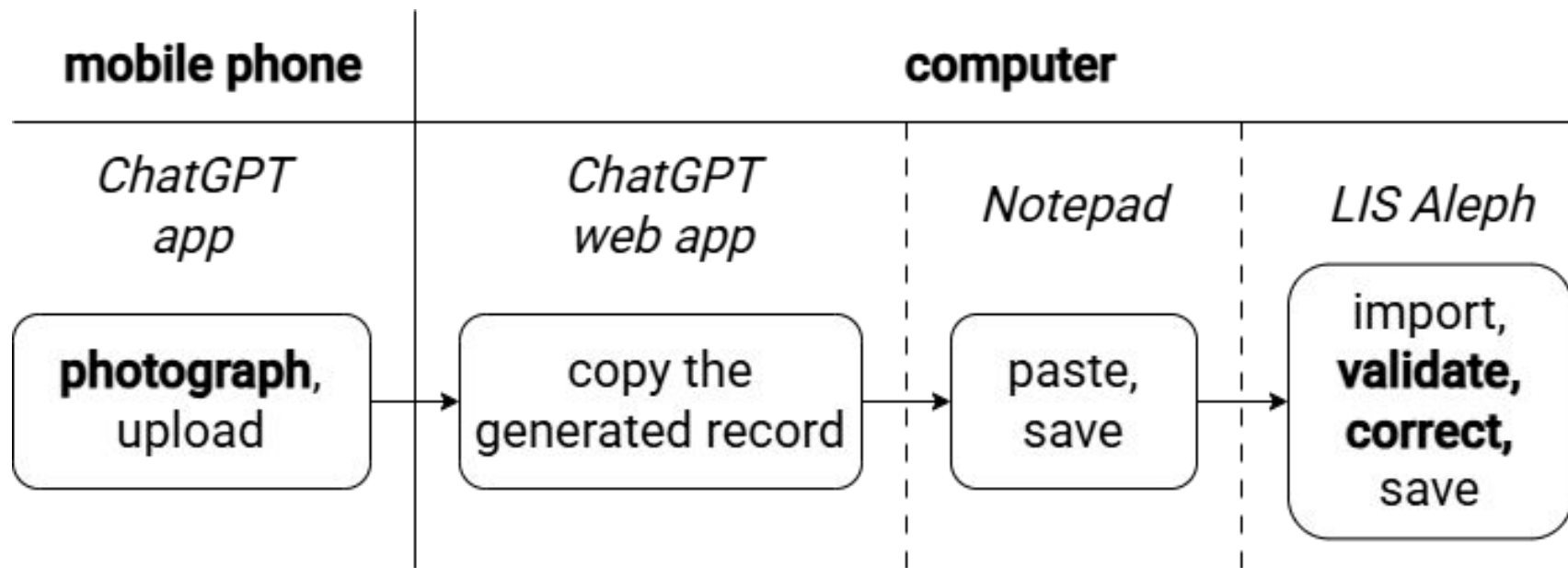
Valsts administrācijas
skola

Our Questions

1. Can AI help cataloguers create bibliographic metadata ***faster*** while maintaining sufficient quality and consistency?
2. What are the main format and structural conformity issues that may arise in LLM-generated MARC21 bibliographic records?

Workflows

Semi-Automated Cataloguing with ChatGPT



ChatGPT Setup and DEMO

The screenshot displays the REKAT interface. At the top left, there is a folder icon and the text "REKAT". To its right, a box contains the text "Project (shared between cataloguers)". Below this is a search bar with a plus sign, the text "Jauna tēzētava REKAT", a microphone icon, and a voice recording icon. Underneath the search bar, there are tabs for "Tēzētavas" and "Avoti", with "Avoti" being the active tab. To the right of these tabs are dropdown menus for "Jaunākie" and "Visi". Below the tabs, there is a plus sign and the text "Pievienot avotus". At the bottom, there is a document icon and the text "REK_šablons_ar_653.txt" followed by "Dokuments · 2026. g. 22. apr.". To the right of this document entry, a box contains the text "Source (general prompt – MARC21 record template supplemented with cataloguing instructions)".

Žoans Sfārs

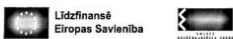
MAZĀIS VAMPĪRIŅŠ

No franču valodas tulkojusi
Elīna Brasliņa



Originaliziedevuma nosaukums "Petit vampire"

"Petit vampire"
Originally published in French under the following title:
Petit vampire, volumes 1 to 3, by Joann Sfar



Grāmata izdota ar Valsts kultūrkapitāla un
Latvijas Republikas Kultūras ministrijas atbalstu

Redaktore Inese Zandere
Korektore Guna Kalniņa
Dizaina adaptācija - Iva Zenne

Iešmas.lv

© Petit vampire, volumes 1 to 3, by Joann Sfar
Editions Delcourt - 2021/2023
First edition 1999
© Joann Sfar: teksts un ilustrācijas, 1999, 2021/2023
© Elīna Brasliņa: tulkojums, 2025
© Ieļa un mazs, izdevums latviešu valodā, 2025

ISBN 978-9934-633-32-4

Iespiesta Jelgavas tipogrāfijā uz papīra Munkens Polar, 150 g/m²
Pērkot šo grāmata, jūs atbalstāt ilgtspējīgu metu apsaimniekošanu.



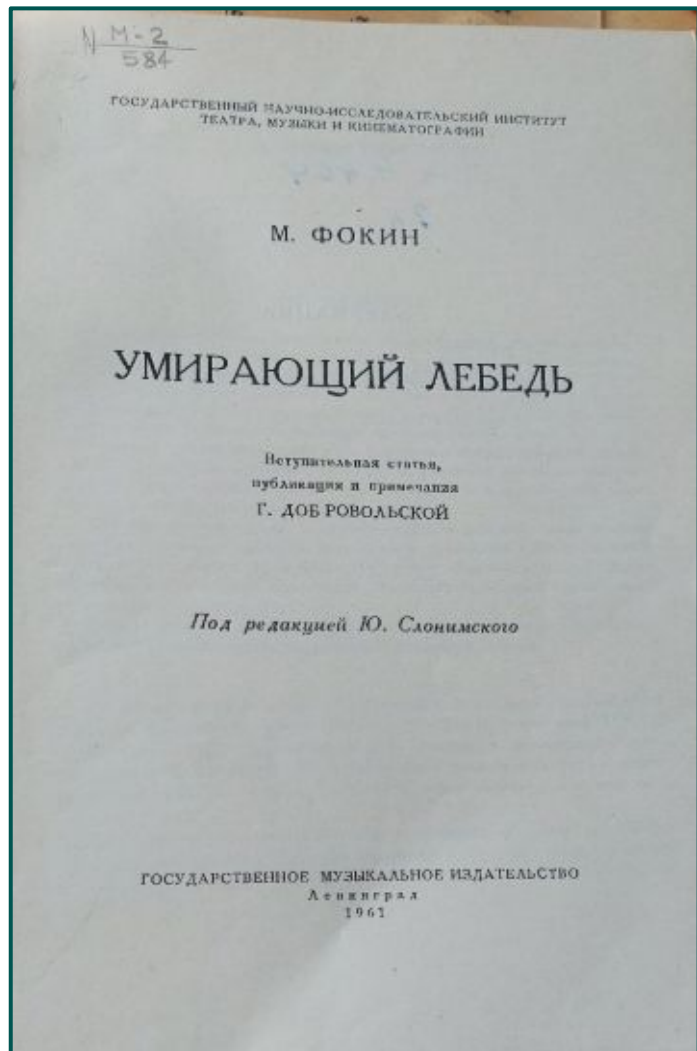
Šeit ir MARC21 ieraksts grāmatai "Mazais vampīriņš":

pgsql

Copy Edit

```
00000001 FMT L BK
00000001 LDR L ^^^^nam^a22^^^^i^4500
00000001 008 L 250613s2025^^^lv^^^^^^^^000^0^lav^^
00000001 020 L $$a9789934633324$$qprošēts
00000001 040 L $$aLV-RivB$$blav$Serda
00000001 0411 L $$alav$$hfre
00000001 1001 L $$aSfārs, Žoans.$$4aut
00000001 2401 L $$aPetit vampire.$1Latviešu valodā
00000001 24510 L $$aMazais vampīriņš /$$cŽoans Sfārs ; no franču valodas tulkojusi Elīna Brasliņa
00000001 264 1 L $$a[Rīga] :$$bLiels un mazs,$$c[2025]
00000001 264 3 L $$a[Jelgava] :$$bJelgavas tipogrāfija
00000001 300 L $$a101 lpp. :$$bilustrācijas ;$$c30 cm
00000001 336 L $$ateksts$$btxt
00000001 336 L $$aattēls$$bsti
00000001 337 L $$atiešuztvere$$bn
00000001 338 L $$asējums$$bnc
00000001 590 L $$aIeraksta pamatlīmenis izveidots ar mākslīgā intelekta rīka palīdzību.
00000001 7201 L $$aBrasliņa, Elīna.$$4trl
00000001 7201 L $$aZandere, Inese.$$4edt
00000001 7201 L $$aKalniņa, Guna.$$4edc
00000001 7201 L $$aZenne, Iva.$$4bkd
00000001 7102 L $$aLiels un mazs.$$4pbl
00000001 91500 L $$a2025.NLL
00000001 91503 L $$ak20250613.03
00000001 91503 L $$a202506.rgb
00000001 919 L $$anb
```

[NB KAT - Mazais vampīriņš ieraksts](#)



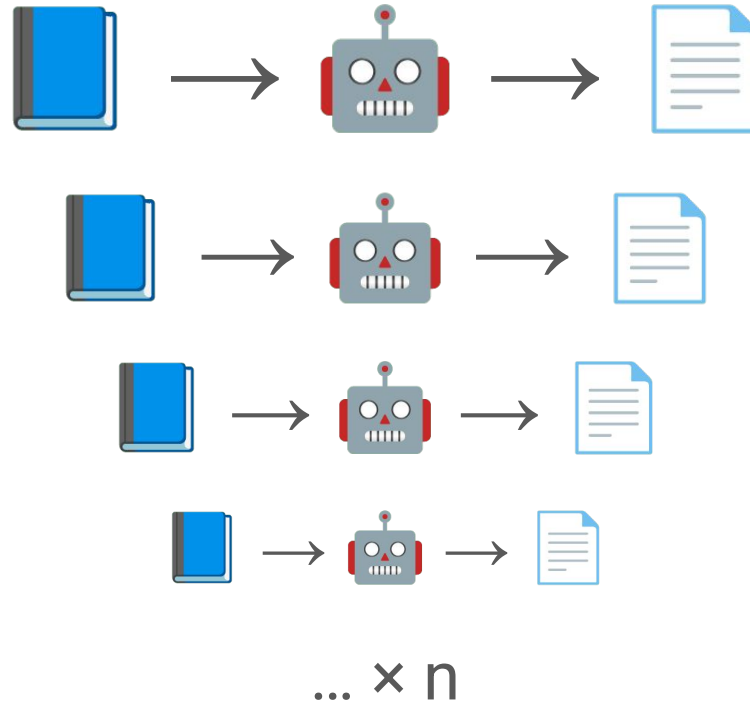
Pamatojoties uz attēliem un šablonu, zemāk ir sagatavots MARC21 ieraksts izdevumam:

```
bash

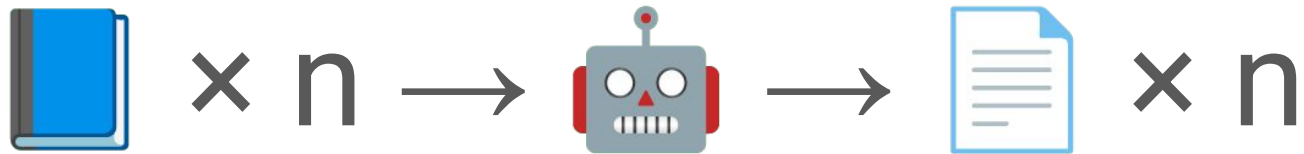
00000001 FMT L BK
00000001 LDR L ^^^^^nam^a22^^^^^^i^4500
00000001 008 L 250606s1961^^^^ru^^^^^^^^^^^^000^0^rus^^
00000001 099 L $$aM-2/584
00000001 1001 L $$aФокин, Михаил Михайлович.$$4aut
00000001 24510 L $$aУмирающий лебедь :$$bвступительная статья, публикация и примечания Г. Добров
00000001 2463 L $$aУмирающий лебедь
00000001 264 1 L $$aЛенинград :$$bГосударственное музыкальное издательство,$$c1961.
00000001 300 L $$a31 lpp. :$$bilustrācijas ;$$c22 cm
00000001 336 L $$ateksts$$btxt
00000001 336 L $$aattēls$$bsti
00000001 337 L $$atiešuztvere$$bn
00000001 338 L $$asējums$$bnc
00000001 500 L $$avecajā ortogrāfijā.
00000001 590 L $$aIeraksta pamatlīmenis izveidots ar mākslīgā intelekta rīka palīdzību.
00000001 7201 L $$aДобровольская, Г.$$4aui
00000001 7201 L $$aСлонимский, Ю.$$4edc
00000001 7201 L $$aГолубовский, И. В.$$4edc
00000001 91500 L $$a2025.NLL
00000001 91503 L $$arek20250606.03
```

[REKAT - MARC21 ieraksts izdevumam](#)

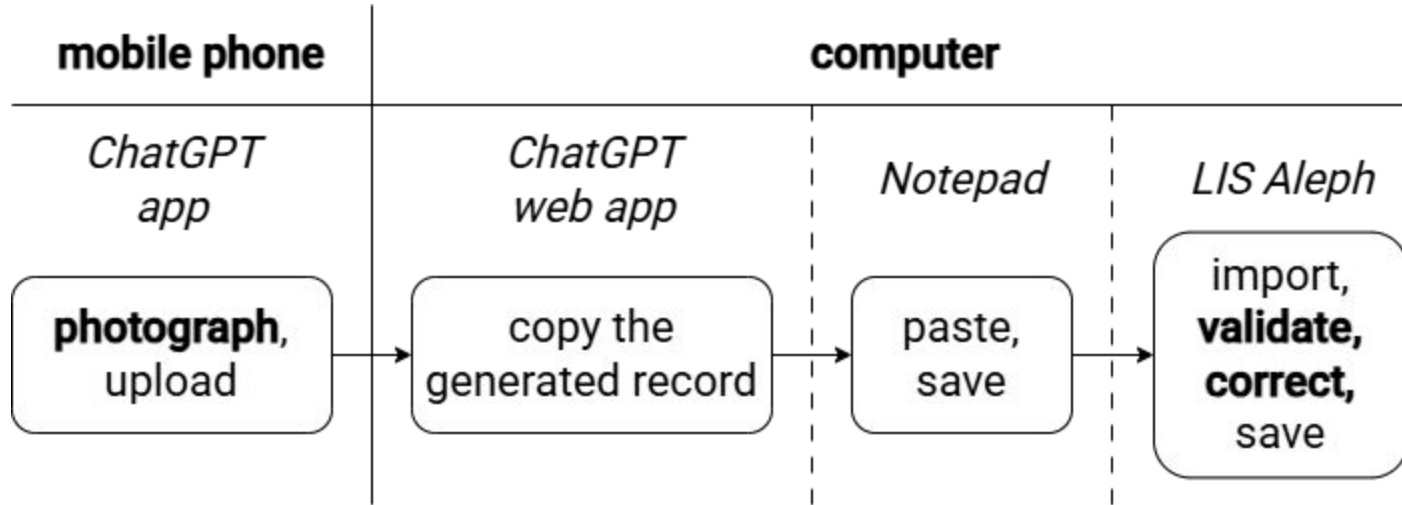
Individual Processing – *Before*



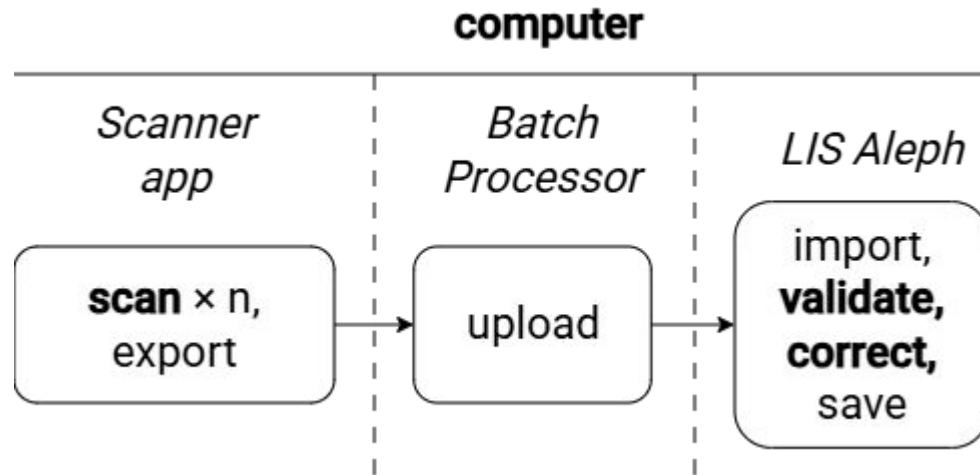
Batch Processing – Now



Semi-Automated Cataloguing with ChatGPT



Semi-Automated Cataloguing with Batch Processor



Batch Processor Setup and DEMO

Batch Processor →

API Key:

System Prompt:

Load from static/prompts/marc.txt

```
Return only Aleph Sequential format record strictly following the template.
00000001 FMT   L BK
00000001 LDR   L ^^^^^nam^a22^^^^^^i^4500
00000001 008   L ^^^^^s[publication year, if there is no publication year, use
"uuuu"]^^^^[publication MARC country code 3 positions, corresponds to field
264]^^^^^^^^^^^^000^0^[publication language code]
00000001 040   L $$aLV-RiVB$$blav$$erda
```

This tool is based on a notebook developed at the National Library of Latvia:

https://github.com/LNB-DH/PublicTools/blob/main/notebooks/llm/batch_processing.ipynb

The MARC21 version of the notebook has not yet been published. If you would like additional information, such as a sample prompt, please contact matiss.bolsteins@lnb.lv.

Challenges...

Observations...

Challenges in AI-assisted cataloguing

LLM-related challenges:

- prompt sensitivity,
- hallucinations,
- *“black-box” nature*

Cataloguing-related challenges:

- multilingual resources and script variation,
- outlier resources,
- local cataloguing rules.

Additional Text Generation + Aleph Sequential Format Conformance

When additional text alongside Aleph Sequential format is generated, Aleph splits the resulting file into multiple records and marks invalid text as “???”



>> this adds an extra manual step – the cataloguer must select the valid record.

MARC21 Conformance I

Table 1

Generated examples of field 008 for one resource

Model	Generated output
anthropic-claude-sonnet-4.5	<u>251111</u> s1907^^^^ru^^^^^^^^^^^^000^0^rus^^
openai-gpt-5	<u>251111</u> s1907^^^^ru^^^^^^^^^^^^000^0^rus^^
google-gemini-2.5-flash	1907 uuusru^^^^^^^^^^^^000^0^rus^^

MARC21 Conformance II

Automātisks datums (00-05)	<input type="text" value="1907uu"/>
Datuma veids (06)	<input type="text" value="u"/>
Datums 1 (07-10)	<input type="text" value="sru"/>
Datums 2 (11-14)	<input type="text"/>
Publicēšanas vieta (15-17)	<input type="text"/> ...
Ilustrācijas (18-21)	<input type="text"/> ... <input type="text"/> ... <input type="text"/> ... <input type="text" value="0"/> ...
Mērķauditorija (22)	<input type="text" value="0"/>
Fiziskā forma (23)	<input type="text" value="0"/>
Satura kodēts apraksts (24-27)	<input type="text"/> ... <input type="text" value="0"/> ... <input type="text"/> ... <input type="text" value="r"/> ...
Valdības publikācija (28)	<input type="text" value="u"/>
Konferences publ. (29)	<input type="text" value="s"/>

1. positions 00-05 cannot be edited in Aleph,
2. other positions must be corrected manually in this interface

Prompt Adherence I

There have been cases where a model generated content in the 300 field, **ignoring the instruction to leave it blank**, such as:

- 300 L \$\$a [2], 134, [2] lpp.
- 300 L \$\$a32 lpp.
- 300 L \$\$a [1], 102 c.

Values in human created records:

- *[2], 216, ii, 8 lpp.*
- *16 lpp.*
- *93, [1] lpp.*

Conclusion = hallucination

Prompt Adherence II

The prompt intentionally uses field 720 so that added entries can be verified and linked in Aleph.

Table 2

Use of field 720 or 700 and subfields \$\$d and \$\$4 in generated records

Model	720	700	\$\$d	\$\$4
openai-gpt-5	16	0	0	2
google-gemini-2.5-flash	9	7	1	12
anthropic-claude-sonnet-4.5	6	6	2	12

Prompt Adherence III

- 7201 L \$\$aA. Г. Конюс **(перевод)** ; В. В. Цубербиллер **(редактор, предисловие)** .
- 7201 L \$\$aСаутин, И. В., **отв. редактор**.
- 7201 L \$\$aКохн, М. П.; Рыбкинд, С. В.; Гринько, Г. Ф.
- 7201 L \$\$aКокс, Харольд; Муратов, Д. Л.

Other Oddities I

In Local Fields (9XX):

- omitted dates,
- omitted punctuation,
- unrelated information generated.

In 337, 338 (already in the prompt):

- omitted diacritical marks,
- new values *invented*.

Other Oddities II - Russian Language *Transcribed* with Latvian Characters

...

000000001 24510 L **\$\$aOčerki finansovoj nauki :\$\$bVypusk vtoroj /\$\$cI. M. Kulišer.**

000000001 2463 L **\$\$aОчерки финансовой науки :\$\$bВыпуск второй**

000000001 250 L **\$\$aVypusk vtoroj.**

000000001 264 1 L **\$\$aPetrograd :\$\$bKnigoizdatel'stvo "Nauka i Škola",\$\$c1920.**

...

Other Oddities III - Russian Language *Transcribed* with Latvian Characters

...

000000001 24510 L **\$\$aZemelnyĭ kodeks USSR :\$\$btekst i
postateĭnyi komentariĭ /\$\$cM. Gershonova, E. Keliemana.**

000000001 250 L **\$\$a3-e izdanie.**

000000001 264 1 L **\$\$aHar'kov :\$\$bIŪridicheskoe
izdatelstvo NKUIe USSR,\$\$c1929.**

...

Conclusions

Conclusions I

- Useful **assistant** for cataloguing resources – saves time, especially with old orthography (e.g., Russian, German Fraktur),
- **Human verification remains essential** – every record must still be checked,
- Hallucinations occur, **consistency is not guaranteed**. Results indicate the need for prompt revisions. However, models may still fail to follow instructions in unexpected ways.

List of References

- *Automatic metadata generation at the KB.* (n.d.). KB, National Library of the Netherlands. Retrieved 18 October 2025, from <https://www.kb.nl/en/over-ons/projecten/automatisch-metadateren>
- *Country Report: National Library of Norway.* (2024). National Library of Norway. <https://cms.sikt.no/sites/default/files/2024-06/NNG2024-Oslo.pdf>
- Lowagie, H. (2024, January 17). *Power Automate flow with Microsoft AI Builder.* Network Group “AI in Libraries” Webinars Fall/Winter 2023/24, Online. <https://www.cenl.org/network-group-ai-in-libraries-webinars-fall-winter-2023/>
- *Praktiskās vadlīnijas MI ieviešanai iestāžu darbā un pakalpojumos.* (2025). Viedās administrācijas un reģionālās attīstības ministrija. <https://www.varam.gov.lv/lv/maksligais-intelekts-valsts-parvalde>