From PDF to Prompt

Toward Universal Bibliographic Control Through Machine-Readable Cataloging Rules

ILFA WLIC 2025 Satellite Meeting

15 August 2025

Hannes Lowagie

KBR - Belgium

KBR Koester de tijd
Protégeons le temps

Universal Bibliographic Control - Principles

- Distribute responsibility globally, each national bibliographic agency taking responsibility for metadata describing its national published output
- Create and maintain metadata at the source, in the territory of origin
- Ensure timeliness of production and distribution of quality metadata
- Make metadata widely and openly available for reuse
- **Use internationally agreed standards to make metadata consistent and interoperable**

Universal Bibliographic Control – Call for Action

> "*Modern technologies offer **powerful tools** that should be actively utilized to **enhance the global exchange**, integration, and accessibility of bibliographic and authority metadata* […] *Rapid technological change and proliferation of digital content require reliable, **standardized information** in order to ensure trustworthy access to global knowledge which is vital for the advancement of knowledge and the functioning of informed societies.*"

- Today: focus on
  - Powerful tools
  - Enhance global exchange
  - Standardized information

Universal Bibliographic Control – Standardization

→Standards about data output

- MARC (MARC21, UNIMARC or other)
  - MARC (.mrc) = ISO 2709 : https://www.iso.org/standard/41319.html
  - MARC (.xml) = ISO 25577 : https://www.iso.org/standard/62878.html
  - MARC (.json) : https://pkiraly.github.io/2018/01/28/marc21-in-json/
- BIBFRAME (.rdf) : https://www.loc.gov/bibframe/
- Dublin Core / OAI-DC : https://www.dublincore.org/

→Standards about data input

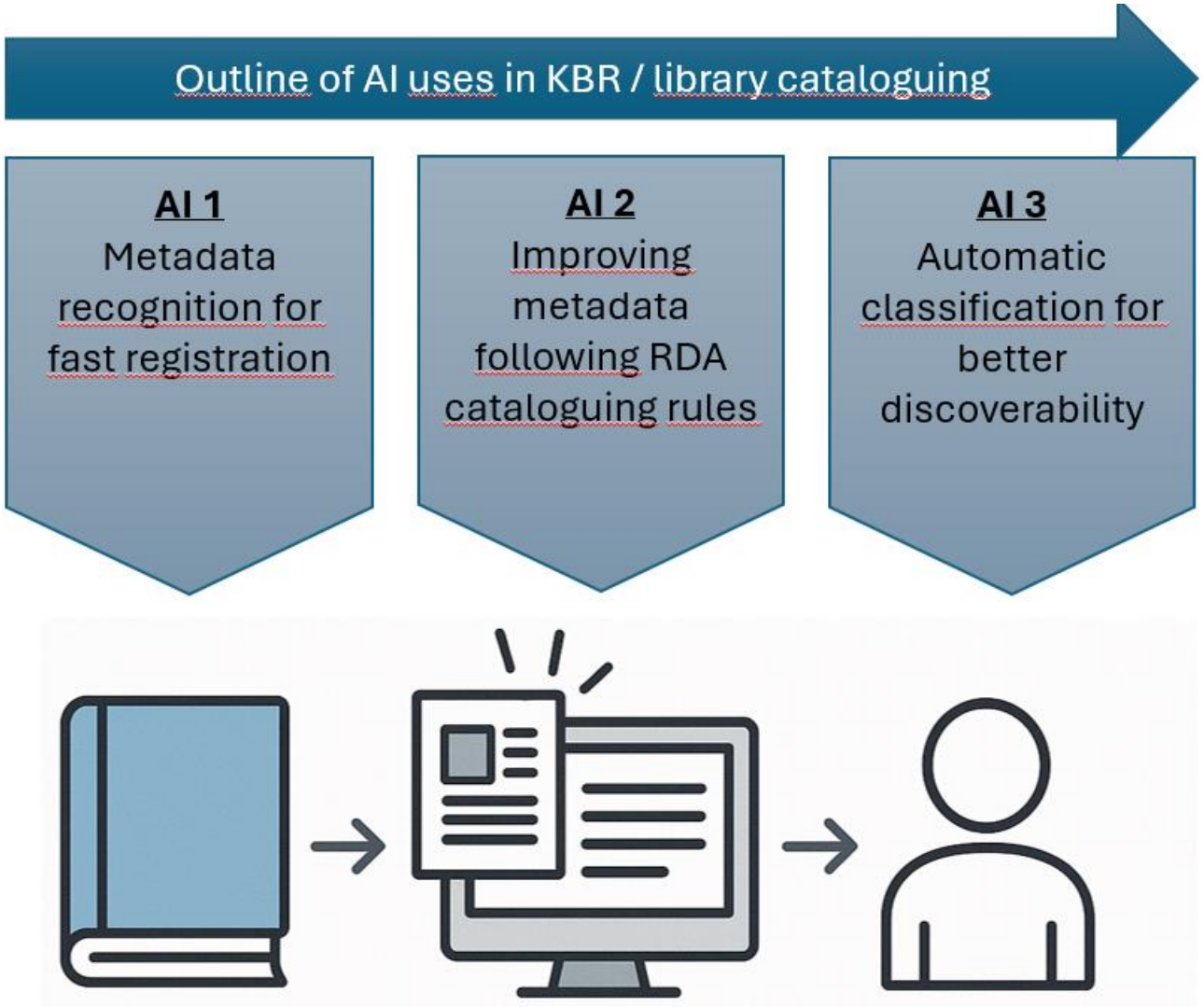
- ISBD
- AACR and RDA
- But every cataloging standard allows for local variations, language, options or adaptations. Every institution has its **own guidelines**.

→ Machine-Readable Cataloging Rules

Universal Bibliographic Control – Machine-readable cataloging rules

- The goal is to make the interpretation of these local guidelines and implementations (*application profiles)* interoperable.
- This ensures that not only the data can be shared in a standardized format, but also that the underlying cataloguing rules themselves become shareable.

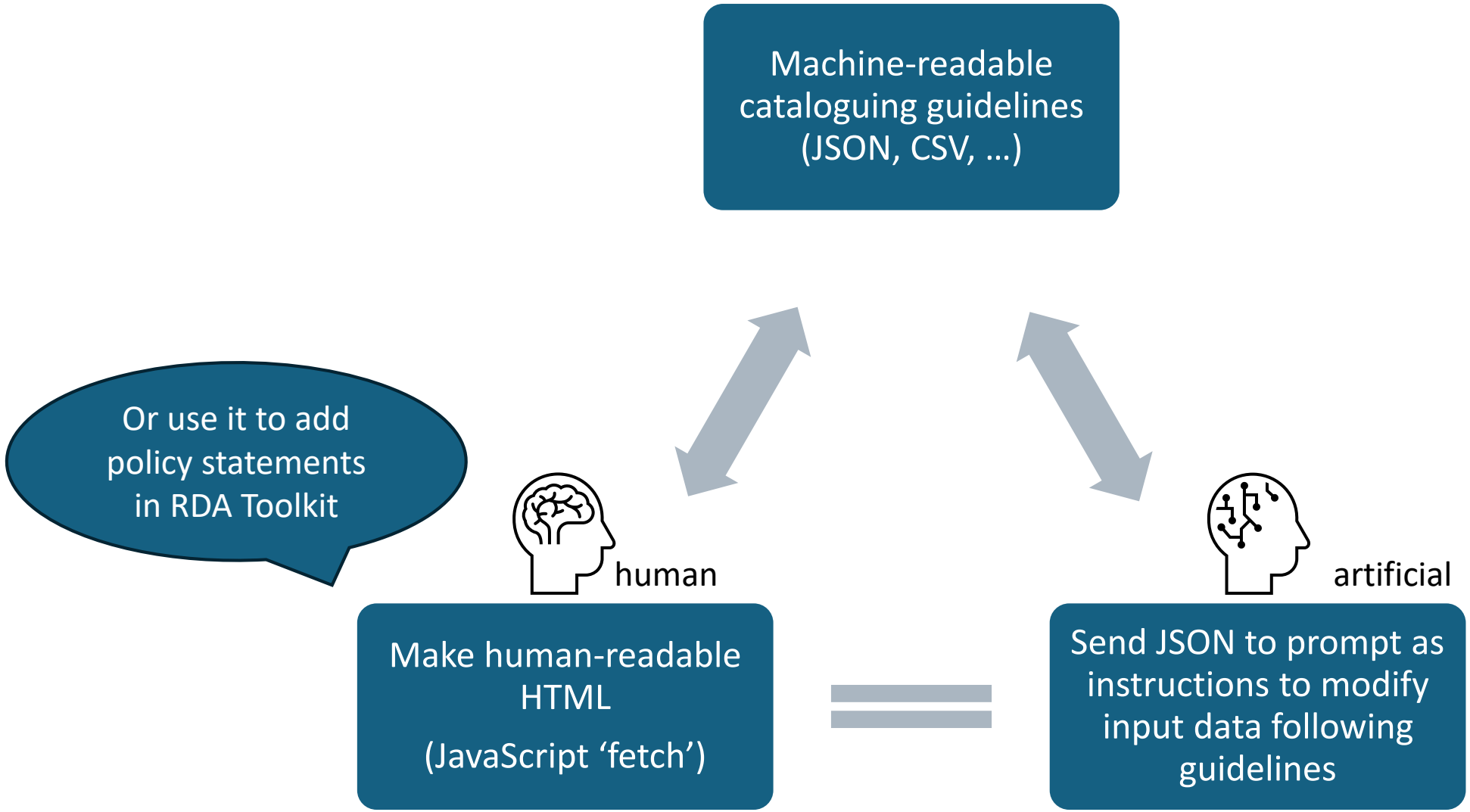
AI as a catalyst for this approach



The need for structured cataloging rules

- Cataloging instructions were first printed, then published in PDF (quickly outdated), then written in HTML (one central, up to date version).
- 2025: Reuse the same instructions as prompts for AI, enabling it to process and transform data as a human would.
- Avoid duplicate work and fragmented files by using a structured implementation.
  - Externalize the instructions into a structured file (JSON)
  - Format them for both human readability and machine actionability

General structure of this approach



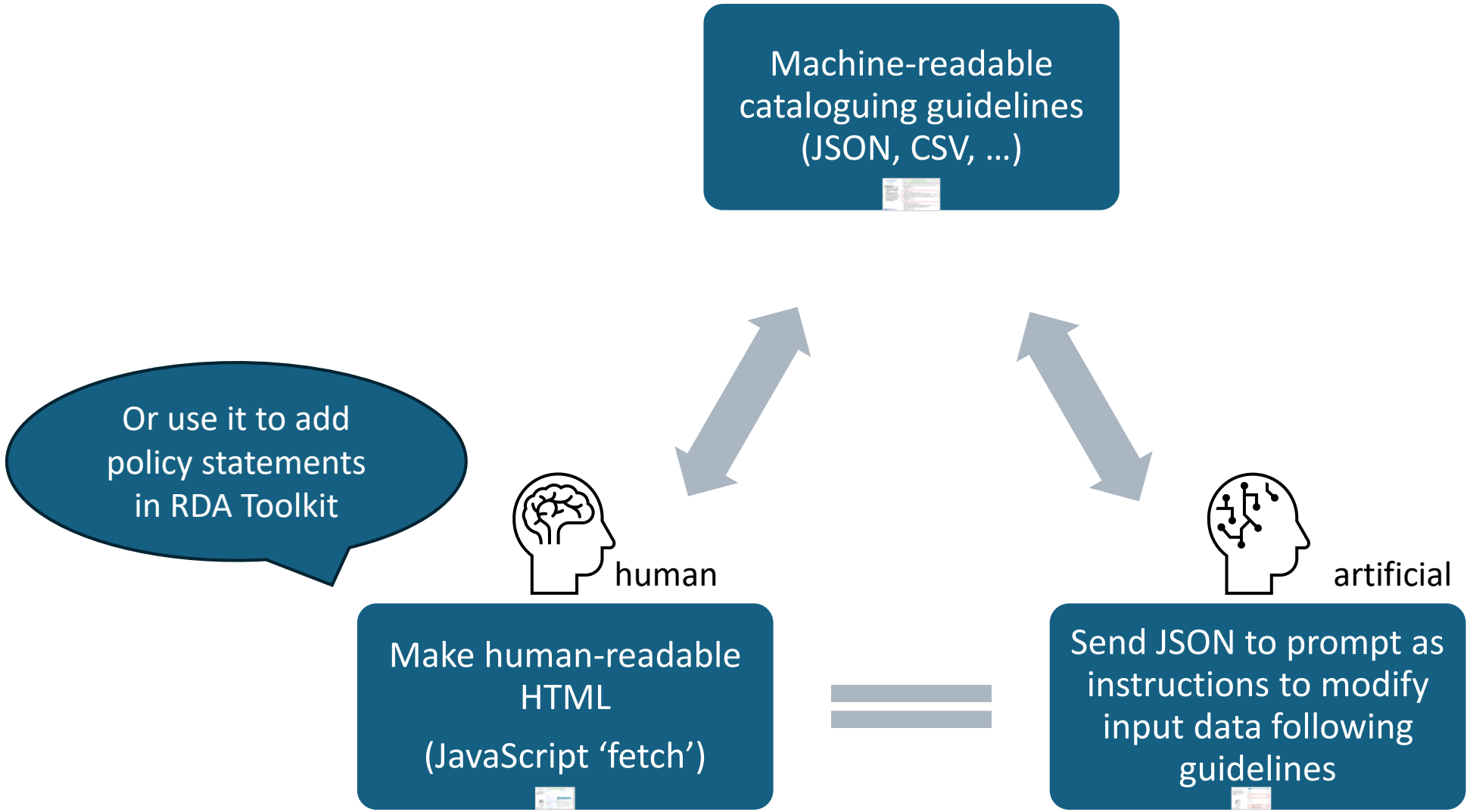
Why JSON?

- Machine readable and easy to parse
- Flexible and extensible
- Ideal for defining rule sets and mappings
- Published via GitHub https://kbrbe.github.io/cataloging-guidelines/guidelines.json
- But : still looking for best JSON structure and collaboration.

JSON Elements

- **ID**: Unique identifier for the rule.
- **TitleFR / TitleNL**: French and Dutch labels.
- **MARC21**: Link to the official MARC21 documentation.
- **RDA_ELEMENT**: Reference to the corresponding RDA element.
- **RDA_TOOLKIT**: URL of the corresponding RDA Toolkit entry.
- **RulesFR / RulesNL**: Cataloguing instructions in French and Dutch.
- **Entity**: The related IFLA-LRM entity
- **FirstInstance**: Indicates the field is mandatory for the first occurrence.
- **AdditionalInstances**: Is recording of other instances mandatory.
- **DataType**: String, Number, ….
- **VES** (Value Encoding Scheme): Specifies the vocabulary (e.g. MARC language code or RDA Binding Type).
- **RecordingMethod**: Unstructured/Structured/Identifier.

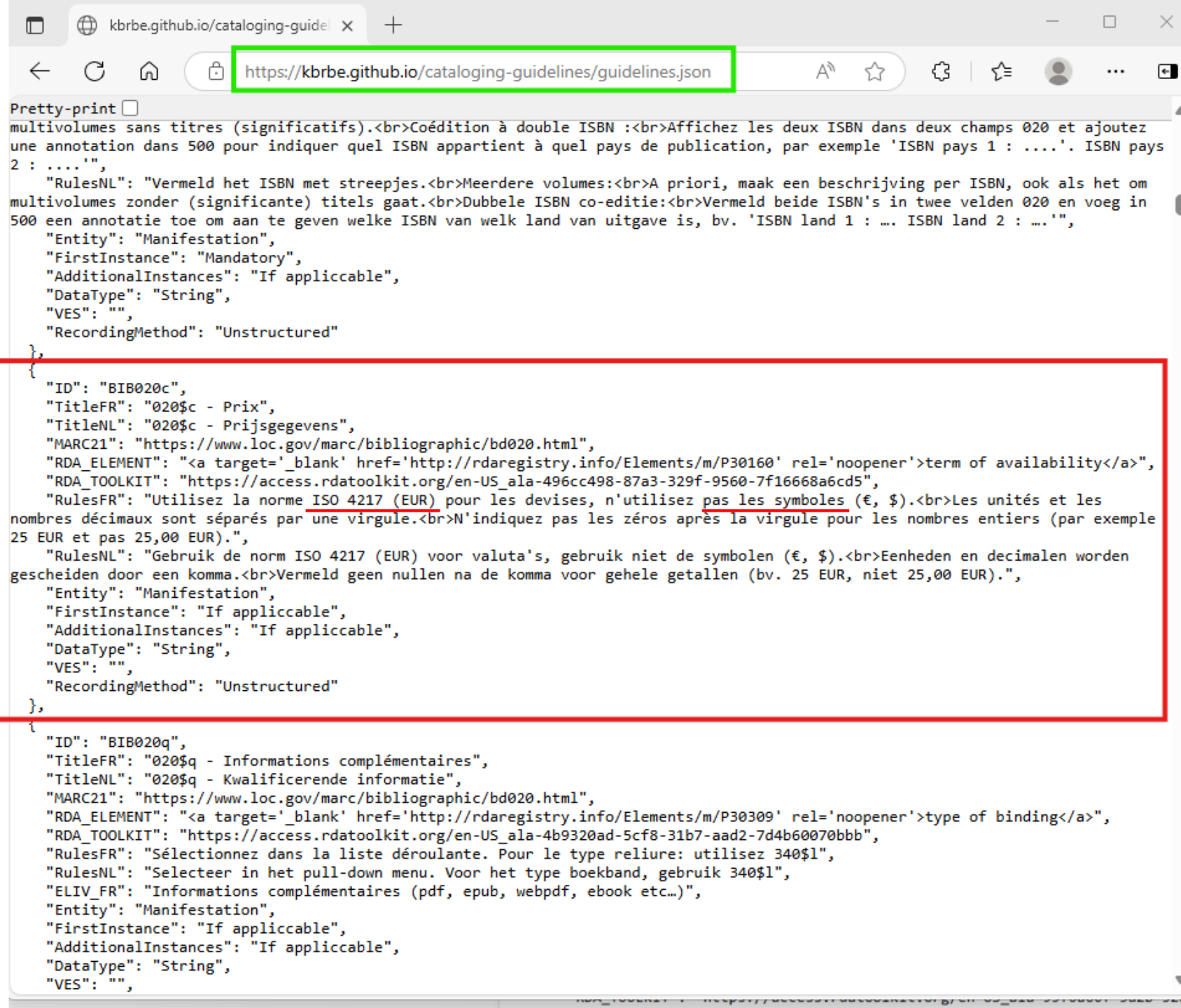
General structure of this approach - Example



Machine-readable guidelines

Example: price

- Use the ISO 4217 (EUR) standard for currencies, do not use the symbols (€, $).
- Units and decimals are separated by a comma.
- Do not include zeros after the comma for whole numbers (e.g. 25 EUR, not 25.00 EUR).



KBR Koester de tijd Protégeons le temps

Human-readable interface

“I see a book with in it ‘15,00 €’, so I put in 020$ 15 EUR”

human

KBR Koester de tijd Protégeons le temps

KBR Cataloguing guidelines page

https://kbrbe.github.io/cataloging-guidelines/cataloging-guidelines.html

Documents

Comment créer une notice bibliographique ?

Livre

008 - Éléments de données de longueur fixe

017$a - Numéro de dépôt légal

017$b - Source d'attribution

017$z - Numéro de dépôt légal annulé/invalide

020$a - Numéro international normalisé des livres (ISBN)

020$c - Prix

020$q - Informations complémentaires

020$z - ISBN annulé ou invalide

040$a - Organisme responsable du catalogage original

040$b - Langue du catalogage

040$d - Organisme responsable des modifications

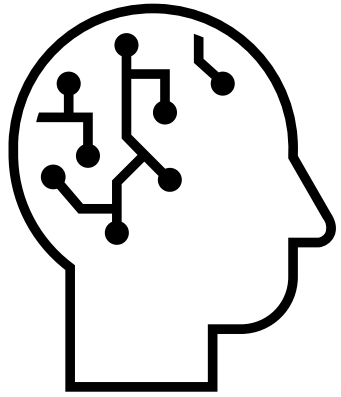
Documenten

Hoe maak ik een bibliografische beschrijving?

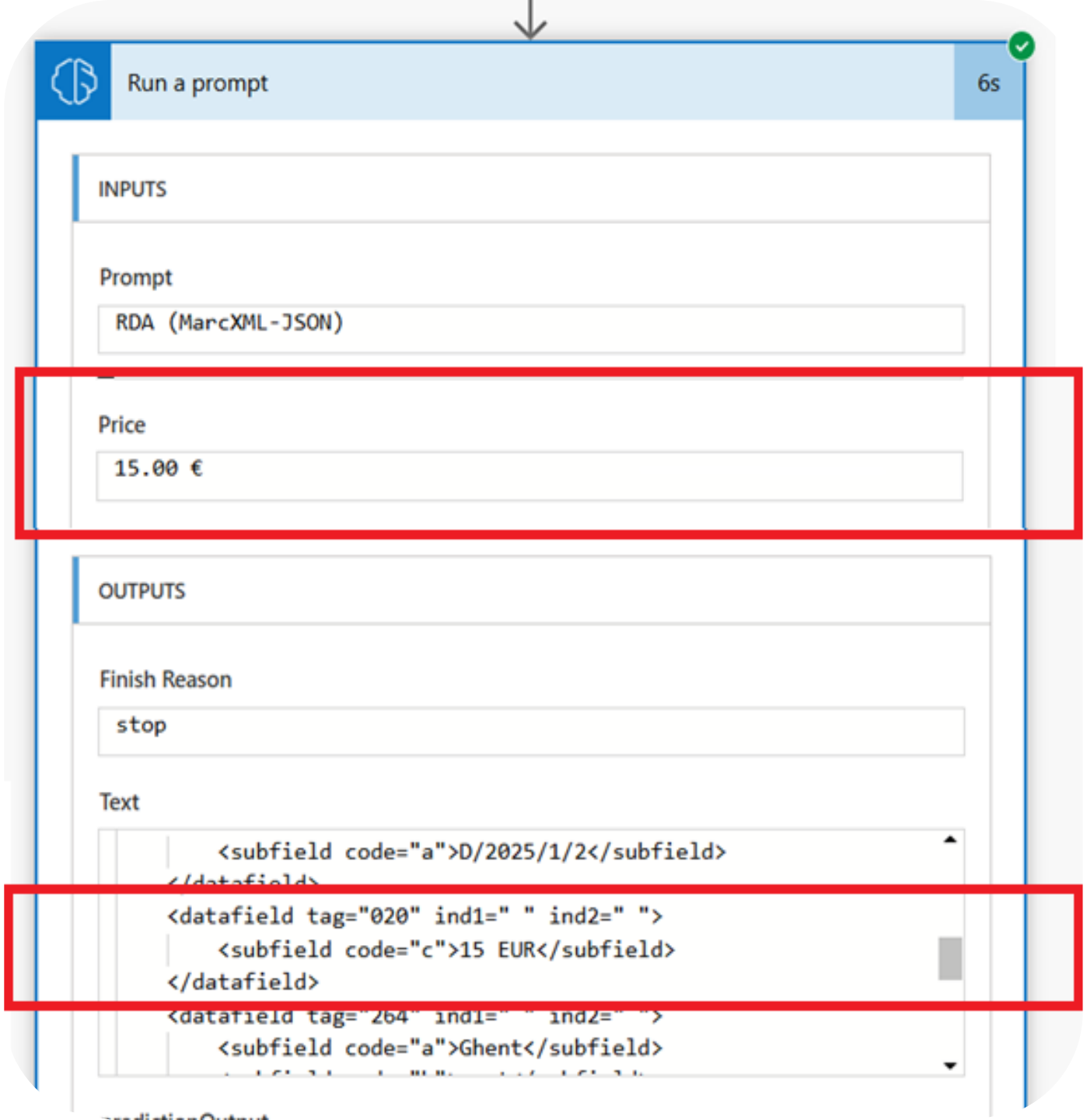
020$c - Prix	020$c - Prijsgegevens
MARC21 : https://www.loc.gov/marc/bibliographic/bd020.html	
RDA element : term of availability	
RDA Toolkit : https://access.rdatoolkit.org/en-US_ala-496cc498-87a3-329f-9560-7f16668a6cd5	
Utilisez la norme ISO 4217 (EUR) pour les devises, n'utilisez pas les symboles (€, $). Les unités et les nombres décimaux sont séparés par une virgule. N'indiquez pas les zéros après la virgule pour les nombres entiers (par exemple 25 EUR et pas 25,00 EUR).	Gebruik de norm ISO 4217 (EUR) voor valuta's, gebruik niet de symbolen (€, $). Eenheden en decimalen worden gescheiden door een komma. Vermeld geen nullen na de komma voor gehele getallen (bv. 25 EUR, niet 25,00 EUR).

GPT in Power Automate

Input Data is ‘15,00 €’,
LLM converts it to 15 EUR,
thanks to instruction from JSON file

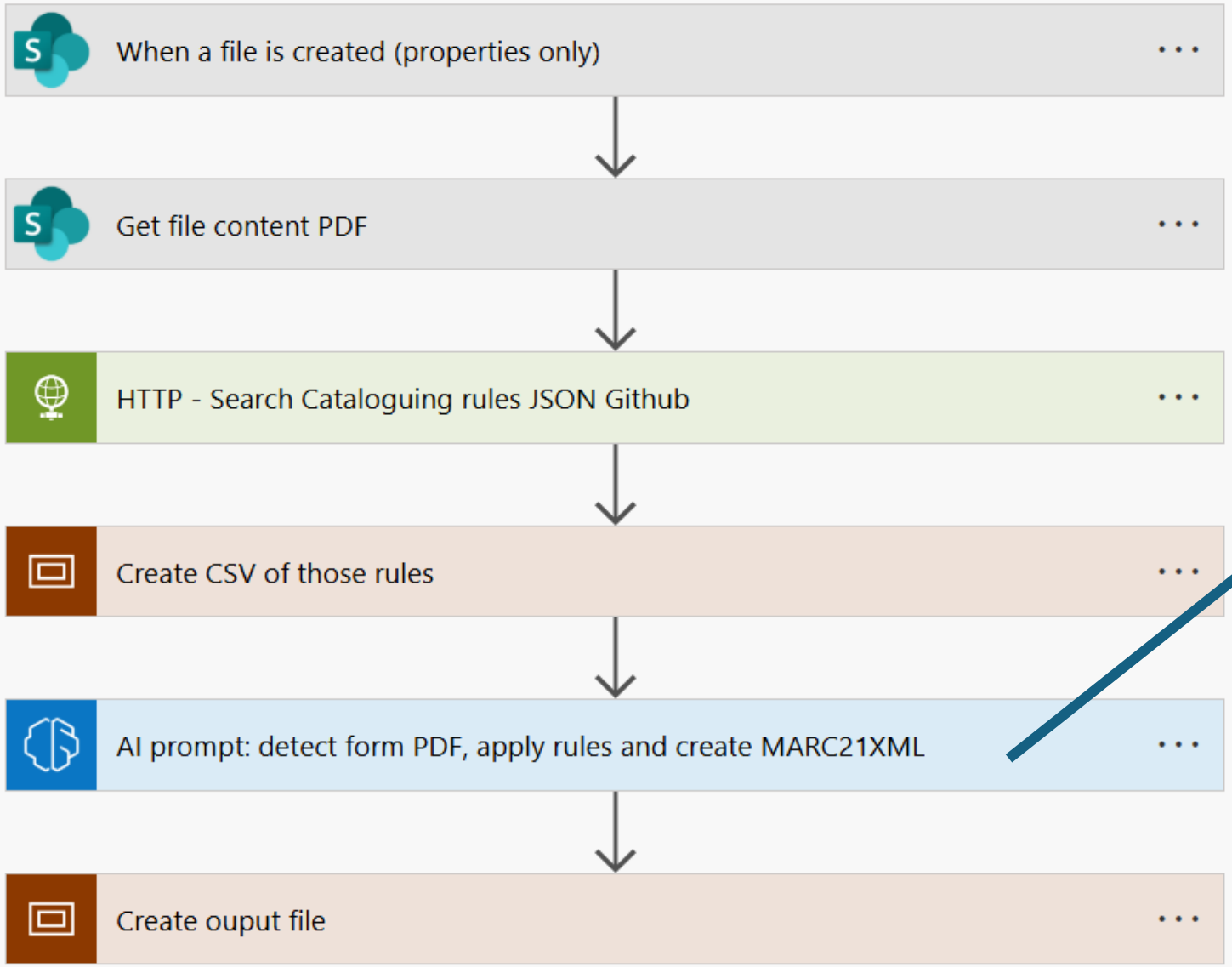


artificial



KBR Koester de tijd
Protégeons le temps

Outline of Power Automate flow



Prompt:

- Extract bibliographic information from the PDF file.
- Apply imported cataloguing rules to modify the metadata accordingly.
- Generate the output in either MARC21 XML or JSON format.

Toward a standardized JSON format for cataloging rules

- Collaborate to define a unified, machine-readable structure (e.g., JSON, URIs) to make cataloging rules interoperable and reusable across systems.
- Such a standardized format can
  1. Support hybrid cataloguing workflows, combining both human cataloging and AI capabilities through shared, standardized guidelines (like explained in previous slides)
  2. Support integration of policy statements in RDA Toolkit
  3. Support data sharing and linking:
     - Metadata are now shared through all kind of protocols (Z39-50, SRU, OAI-PMH, API, …), but it can also be useful to share the cataloguing rules behind the creation of those metadata in machine-readable format in order to ‘understand’ how the data were created.
  4. Support the development of shared applications (one AI tool but fed with local cataloging instructions).
  5. Overall improve and enhance UBC principles

Conclusion I: The limits of standardized metadata for UBC

- UBC depends mainly on:
  - Interoperability and reusability of data
  - Trust in the quality and consistency of the original metadata
- What can undermine UBC in practice?
  - Even when bibliographic data is shared, the output is standardized and a normative descriptive standard (like ISBD) is followed, local refinements and applications are not always shared/known. This causes
    - Inconsistencies in interpretation
    - Doubts about accuracy and applicability in local contexts
    - Re-cataloguing or re-editing when reused

Conclusion II: How machine-readable cataloguing rules can help realize UBC

- Improved transparency
  - Interpret how the data were constructed
  - Evaluate the rules with own local practices (re-change if needed for local use)
  - Better trust and reuse
- Interoperability of cataloguing processes
  - Translate or map one rule set to another
  - Adapt one library's data to another's local profile
  - Hybrid workflows (human and AI) become more consistent
  - → supports UBC goal of 'catalogue once, use everywhere'.
- Scalable, shared tools
  - Shared cataloging tools can ingest local profiles dynamically
  - Support automated validation (cf. IFLA Metadata Standards Matter Symposium / SHACL4BIB)

Thank you

- Hannes.Lowagie@kbr.be or https://www.linkedin.com/in/hanneslowagie/
- More information?

KBR Koester de tijd
Protégeons le temps