

A preservação da informação do módulo de assinaturas do Koha com ferramentas de Inteligência Artificial

NOVA FCT, Divisão de Biblioteca, Arquivo e Cultura

Ana Isabel Roxo, airr@fct.unl.pt

António Baptista, ap.baptista@fct.unl.pt

Tiago Simões, tmb.simoese@fct.unl.pt



Dia Mundial da Preservação Digital

Primeira quinta-feira de novembro (06 novembro 2025)

“Designa-se, assim, por preservação digital o conjunto de actividades ou processos responsáveis por garantir o acesso continuado a longo-prazo à informação e restante património cultural existente em formatos digitais” (Ferreira, 2006, 20 p.).

“Apesar [...] das ferramentas de software que auxiliam o processo de arquivo e preservação [...], ainda se verifica uma escassez assinalável de produtos comerciais com capacidades de preservação. Isto faz com que cada organização se sinta de certa forma responsável pelo desenvolvimento e implementação do seu próprio sistema de preservação, bem como pela definição de políticas de preservação adequadas” (Ferreira, 2006, 66 p.).

Preservação de dados MARC

- O processo de migração das monografias e a exportação dos dados dos exemplares, garantindo a sua conversão em vários formatos de código MARC e alimentação e eventuais futuras bases de dados, está já em prática e tem tido resultados satisfatórios.
- Apesar dos avanços nas ontologias e semânticas de metadados e de dados relacionados, também tem sido provada por vários autores e mantida a preservação dos formatos MARC, pela sua grande disseminação:
 - Rudolf, U. (2024). A MARC-formátum két szintje és magyar fordításának jelentősége. *Central European Library and Information Science Review Közép-Európai Könyvtár- És Információtudományi Szemle*, 1(3). <https://doi.org/10.3311/CELISR.37588>

Módulo de Assinaturas

- Principal vantagem: facilitar a receção de números vindouros.
- Grande número de exemplares = registo pesado
 - O tempo de carregamento de um registo, tanto no OPAC como no *Interface dos técnicos*, aumenta significativamente e para gestão de trabalho em *frontend* é relevante o tempo de carregamento de registos.

Módulo de Assinaturas

- Principal vantagem: facilitar a receção de números vindouros.
- Grande número de exemplares = registo pesado
 - O tempo de carregamento de um registo, tanto no OPAC como no *Interface dos técnicos*, aumenta significativamente e para gestão de trabalho em *frontend* é relevante o tempo de carregamento de registos.
- Digitalização da produção científica = Periódicos científicos online
 - Menos de 5 assinaturas abertas de periódicos em formato físico
- A possibilidade de estabelecer um limite a partir do qual o Koha oculta os exemplares e só os carrega se o utilizador clicar num link (opção da Administração do Koha: `OpacMaxItemsToDisplay`).

Módulo de Assinaturas

- Breve avaliação a 75 OPAC Koha de bibliotecas portuguesas

Tipo de descrição	N.º de Bibliotecas
Assinaturas	17
Assinaturas e Exemplares	22
Exemplares	17
Sem Assinaturas e Exemplares (Campos de Notas)	12

} 39 Bibliotecas têm dados no Módulo de Assinaturas

Módulo de Assinaturas

Não há mapeamento predefinido com os campos MARC

- O Koha é um sistema de gestão documental que funciona com uma base de dados em SQL
- Nossa versão: **Versão 22.11.20**, o Database Type é o MariaDB, versão 10.11.6 e corre no Debian 12
- Tabelas de dados
 - Correspondência com linguagens diferentes, conforme o **MA**chine-**R**eadable **C**ataloging (MARC) que se escolha.

Módulo de Assinaturas

Não há mapeamento predefinido com os campos MARC

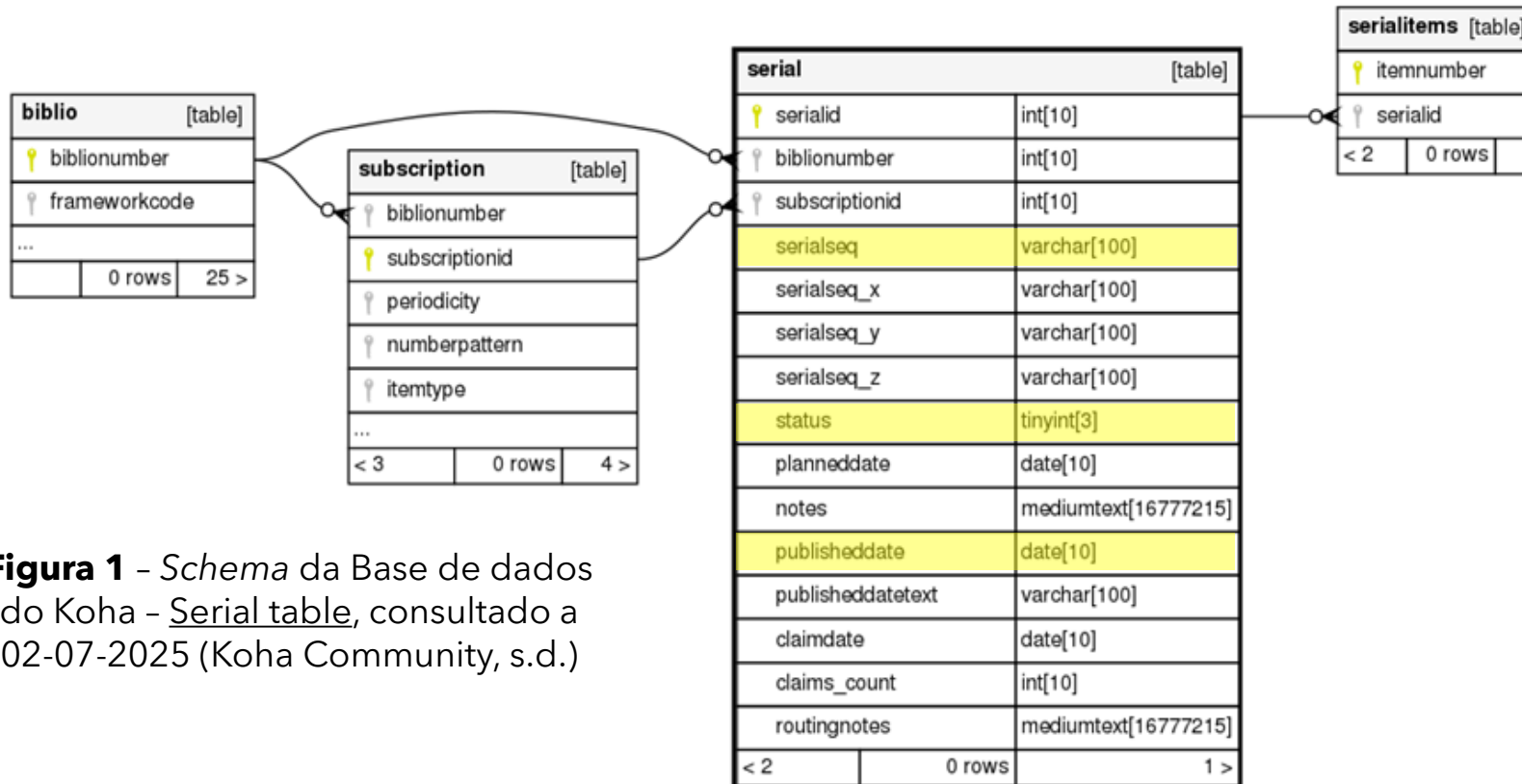


Figura 1 - Schema da Base de dados do Koha - Serial table, consultado a 02-07-2025 (Koha Community, s.d.)

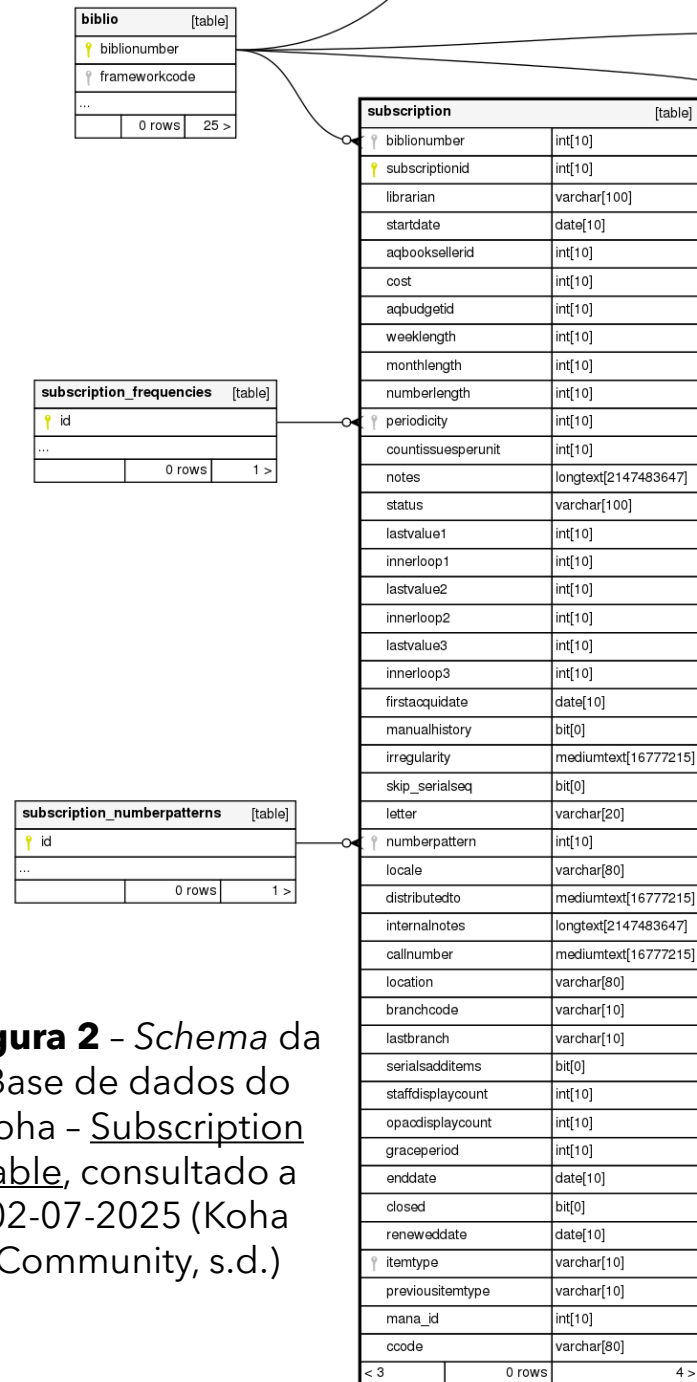


Figura 2 - Schema da Base de dados do Koha - Subscription table, consultado a 02-07-2025 (Koha Community, s.d.)

Módulo de Assinaturas

- Exportação dados das existências via Interface dos técnicos
 - Guardar registo? ✗
 - Exportar dados do catálogo? ✗
 - Exportar dados em CSV? ✗ Para registos, apenas exporta o mapeamento MARC
- Assinatura com criação de exemplar? ⚠ Retroativamente, criar um a um
 - Não cria uma correlação, apenas copia dados na página html

Relatório SQL

- O Koha Wiki disponibiliza o SQL para os relatórios, por Joseph Alway (2025):
 - List of **Current Periodicals** and their holdings -> Assinaturas **abertas**
 - List of **Non Current Periodicals** and their holdings -> Assinaturas **fechadas**
- Exporta dados da assinatura e de todas as existências

Edição de periódico *Metals forum* cota: TA401M110

Numeração	Publicado em	Publicado em (texto)	Esperado em	Estado	Notas
Número Vol 2, No 11				Esperado	
Número Vol 5, No 2	 1982-09-13		 1982-09-13	Recebido	
Número suplementar				----	

Guardar

+ Multi receber

- Relatório:

bibliumber	subscriptionid	title	ISSN	Holdings
50512 ▲	663	Metals forum		Vol 1, No 1; Vol 1, No 2; Vol 1, No 3; Vol 1, No 4; Vol 2, No 1; Vol 2, No 2; Vol 2, No 3; Vol 2, No 4; Vol 3, No 1; Vol 3, No 2; Vol 3, No 3; Vol 3, No 4; Vol 4, No 1; Vol 4, No 2; Vol 4, No 3; Vol 4, No 4; Vol 5, No 5; Vol 5, No 2; Vol 5, No 4; Vol 2, No 11

O que alterámos

- Incluimos Data de publicação e Estado de todas as existências
- Permite preservar toda a informação

Informação de coleção de periódico para *Metals forum* cota: TA401M110

Resumo da assinatura

Número da assinatura	Periodicidade	Modelo de numeração	Biblioteca	Cota	Notas	
# 663	Trimestral	Vol., N.º		TA401M110		+ Criar uma lista de circulação Renovar

Manage	1982	1981	1980	1979	1978
--------	------	------	------	------	------

[Editar periódicos](#) [Gerar próximo](#) [Multi receber](#) [Eliminar números seleccionados](#)

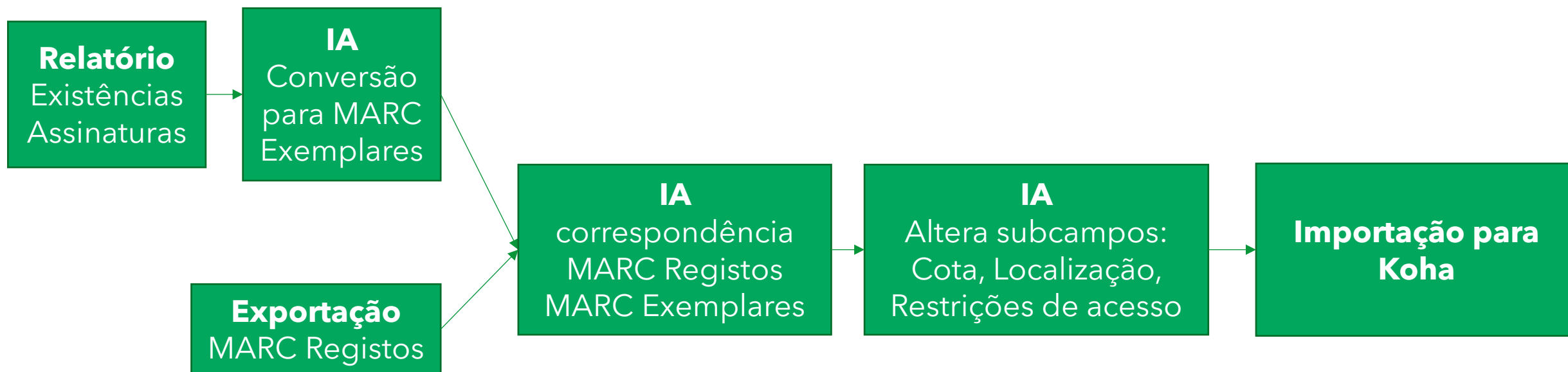
[Seleccionar todos](#) | [Limpar todos](#)

Alterar	Data de publicação	Data de publicação (texto)	Data de recepção	Número	Estado	Notas	Biblioteca	Circulação
☐	1982-09-13		1982-09-13	Vol 5, No 2	Recebido			Imprimir lista
☐	1982-06-13		1982-06-13	Vol 5, No 5	Recebido			Imprimir lista
☐	1982-03-15		1982-03-15	Vol 5, No 4	Recebido			Imprimir lista
	serial.publisheddate			serial.serialseq	serial.status			

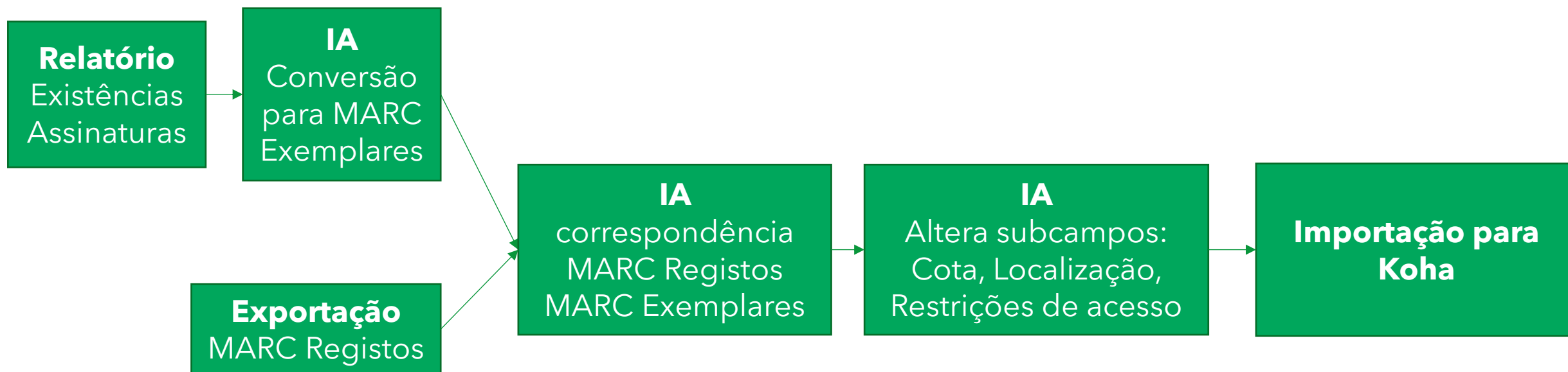
IA para tratamento de dados

- Criada *prompt* para tratamento dos dados em bruto extraídos através do relatório
- Dados foram tratados e convertidos em MARC pela IA
 - Tipo de tratamento: atribuir cotas, localizações, restrições de acesso
- IA usada para intercalar o código MARC dos exemplares com a exportação dos registos

Plano de trabalho



Plano de trabalho



Desafios

- Big data
 - Mais de 12000 títulos
 - Mais de 85000 existências
- Apesar da grande capacidade do Microsoft Excel:
 - Dimensões:

MARCXML vs MARC

MARCXML	MARC
<pre><record> <datafield code="995"> <subfield code="k">COTA</subfield> </datafield> </record></pre>	<pre>=995 \ \$kCOTA</pre>

- Limite na quantidade de dados a tratar
- Lentidão de tratamento de grande volumes de dados

Desafios

- Big data - IA
 - Dividir dados - Vários ficheiros com tabelas de 1 subcampo
 - *Prompt Engineering* para replicar, em mais de 12000 títulos, os dados dos exemplares "modelo" para os exemplares criados (Cota, Localização na prateleira, Restrições de utilização)

Modelo	<subfield code="k">	QA26.3J1	</subfield>
Criado	<subfield code="k">	QA26.3J1	</subfield>
Criado	<subfield code="k">	QA26.3J1	</subfield>
Criado	<subfield code="k">	QA26.3J1	</subfield>
Criado	<subfield code="k">	QA26.3J1	</subfield>
Modelo	<subfield code="k">	QC1P40	</subfield>
Criado	<subfield code="k">	QC1P40	</subfield>
Criado	<subfield code="k">	QC1P40	</subfield>
Criado	<subfield code="k">	QC1P40	</subfield>
Criado	<subfield code="k">	QC1P40	</subfield>

Desafios

- Resiliência
 - Confronto com as limitações da IA
 - Confronto com as alucinações da tecnologia
- Linguagem clara e explícita
 - *Prompts* bem estruturadas para tratamento de tabelas de dados

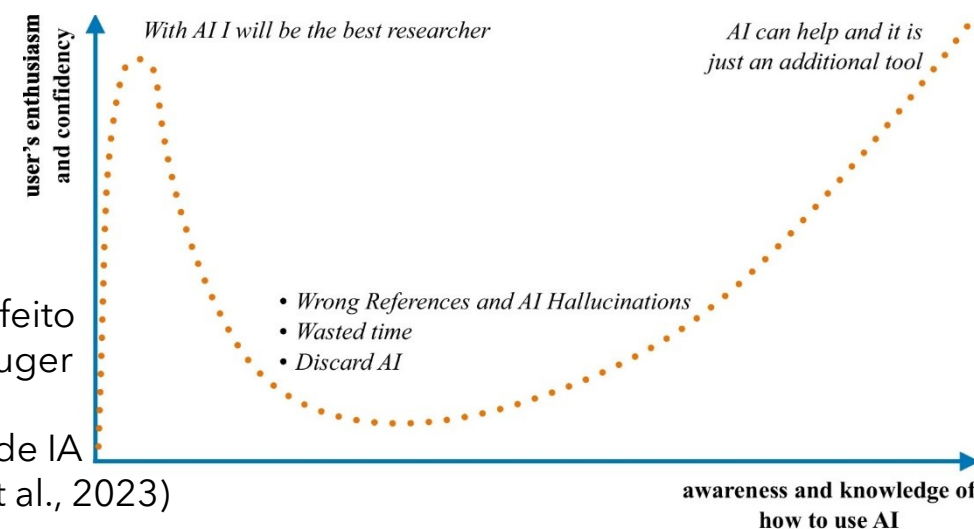


Figura 3 – Efeito Dunning-Kruger adaptado à utilização de IA (Salvagno et al., 2023)

Exemplo da *prompt* usada no slide anterior:

"Identifica que carreguei 1 ficheiro.

Analisa cada Record individualmente e identifica que o primeiro campo 995 tem um subcampo \$k.

Copia esse valor e substitui o valores dos \$k seguintes dentro desse Record.

Repete esta substituição para todos os Record do ficheiro.

Exporta o resultado como um ficheiro .txt."

Conclusões

- **Exportação** da totalidade dos **dados** alimentados no **módulo de assinaturas** do Koha
 - Migração dos dados para MARC
 - **Preservação do MARC** (Rudolf, 2024)
- **Linguagem natural**
 - acessível a qualquer profissional de Informação e Documentação (Garg & Rajendran, 2024)

Conclusões

- **Custo reduzido**
 - A necessidade de uma assinatura de um plano mensal de IA dependerá do volume de dados
- **Autonomia das bibliotecas**
 - Todas estas tarefas podem ser feitas através do Interface dos técnicos

Bibliografia

Alway, J. (2025, February 28). *Serial reports* - Koha Wiki. Acedido a 15-05-2025, de https://wiki.koha-community.org/wiki/Serial_reports

Ferreira, M. (2006). *Introdução à preservação digital - Conceitos, estratégias e actuais consensos*. Escola de Engenharia da Universidade do Minho. Acedido a 15-07-2025, de <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>

Garg, A., & Rajendran, R. (2024). Analyzing the Role of Generative AI in Fostering Self-directed Learning Through Structured Prompt Engineering. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14798 LNCS, 232-243. https://doi.org/10.1007/978-3-031-63028-6_18

Koha Community (s.d.). serial. [schema.koha-community.org](https://schema.koha-community.org/22_11/tables/serial.html). Acedido a 02-07-2025, de https://schema.koha-community.org/22_11/tables/serial.html

Koha Community (s.d.). subscription. [schema.koha-community.org](https://schema.koha-community.org/22_11/tables/serial.html). Acedido a 02-07-2025, de https://schema.koha-community.org/22_11/tables/serial.html

Rudolf, U. (2024). A MARC-formátum két szintje és magyar fordításának jelentősége. *Central European Library and Information Science Review Közép-Európai Könyvtár- És Információtudományi Szemle*, 1(3). <https://doi.org/10.3311/CELISR.37588>

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care*, 27(1), 1-2. <https://doi.org/10.1186/S13054-023-04473-Y/FIGURES/1>