



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

Ana Isabel Roxo, Biblioteca NOVA FCT, airr@fct.unl.pt, ORCID 0000-0003-4513-3336

António Baptista, Biblioteca NOVA FCT, ap.baptista@fct.unl.pt, ORCID 0000-0002-9185-6335

Tiago Simões, Biblioteca NOVA FCT, tmb.simoes@fct.unl.pt, ORCID 0009-0005-2275-0333

Resumo

Apesar da importância da preservação digital ter vindo a ser destacada por várias instituições internacionais e nacionais, esta ainda fica a cargo de cada organização. Constatámos que o Koha não tem uma forma direta para exportar os dados do Módulo de Assinaturas e que a comunidade nacional e internacional não apresenta soluções claras para fazê-lo.

Adotando uma abordagem mista, pesquisámos por soluções e identificámos que a comunidade internacional partilha relatórios que exportam dados do módulo de assinaturas. A partir destes, adotámos uma abordagem mais quantitativa, elaborando várias experiências com recurso ao Microsoft Copilot, para migrar os dados para o formato MARC. A utilização da IA permitiu o tratamento de um grande volume de dados.

Estas ferramentas permitem a exportação e preservação dos dados alimentados no módulo de assinaturas do Koha e, recorrendo somente à linguagem natural, a IA pode dar autonomia às bibliotecas para elaborar um plano de preservação digital.

Palavras-chave

Preservação digital, Inteligência Artificial, Módulo de assinaturas Koha, Migração MARC.

Texto da proposta

Apesar da importância da preservação digital (Ferreira, 2006) ter vindo a ser já destacada por várias instituições internacionais e nacionais¹, ainda se verifica que fica a cargo de cada organização a responsabilidade do desenvolvimento e implementação de sistemas de preservação. Apesar dos avanços nas ontologias e semânticas de metadados e de dados relacionados, o desenvolvimento de trabalhos para “garantia de acesso continuado a longo-prazo” (Ferreira, 2006) nos vários formatos MARC têm sido significativos (Rudolf, 2024) e a migração de dados tem comprovado a maturidade dos sistemas de gestão integrada na exportação e importação de dados MARC.

Avaliando o Módulo de Assinaturas do Koha, identificamos como sendo sua principal vantagem a gestão de assinaturas abertas, para facilitar a geração e receção de números vindouros. Mas a digitalização da informação e a alteração dos modelos de aquisição de

¹ Tendo-lhe sido até conferido o estatuto de Dia Mundial da Preservação Digital, pela *Digital Preservation Coalition*, em 2017 na primeira quinta-feira de novembro (Branco, N. (2024, November 8). *Preservação digital: Um bem necessário*. Secretaria-Geral da Presidência do Conselho de Ministros. <https://www.sg.pcm.gov.pt/a-sgpcm/noticias/2024/novembro/preservacao-digital/>)



publicação científica, tornou relativamente obsoleto este módulo, dada a redução quase na totalidade das assinaturas de periódicos em formato físico, atualmente menos de 5 na Biblioteca da NOVA FCT.

Realizámos também uma breve análise entre 75 OPAC (Online Public Access Catalog) de bibliotecas portuguesas que utilizam o Koha e verificámos que 39 têm dados de descrição de publicações periódicas no Módulo de Assinaturas, sendo que 17 delas não apresentam dados alguns nos exemplares:

Tipo de descrição	N.º de Bibliotecas
Assinaturas	17
Assinaturas e Exemplares	22
Exemplares	17
Sem Assinaturas e Exemplares (Em campos de notas UNIMARC)	12

Figura 1 – Dados de breve prospeção de descrição de periódicos em OPAC Koha nacionais

Com estes dados, confirmamos a relevância deste trabalho, pois sendo o Koha um sistema de gestão documental com uma base de dados em SQL, não há um mapeamento predefinido das tabelas de dados do Módulo de Assinaturas (*Subscription* e *Serial*) com os códigos MARC.

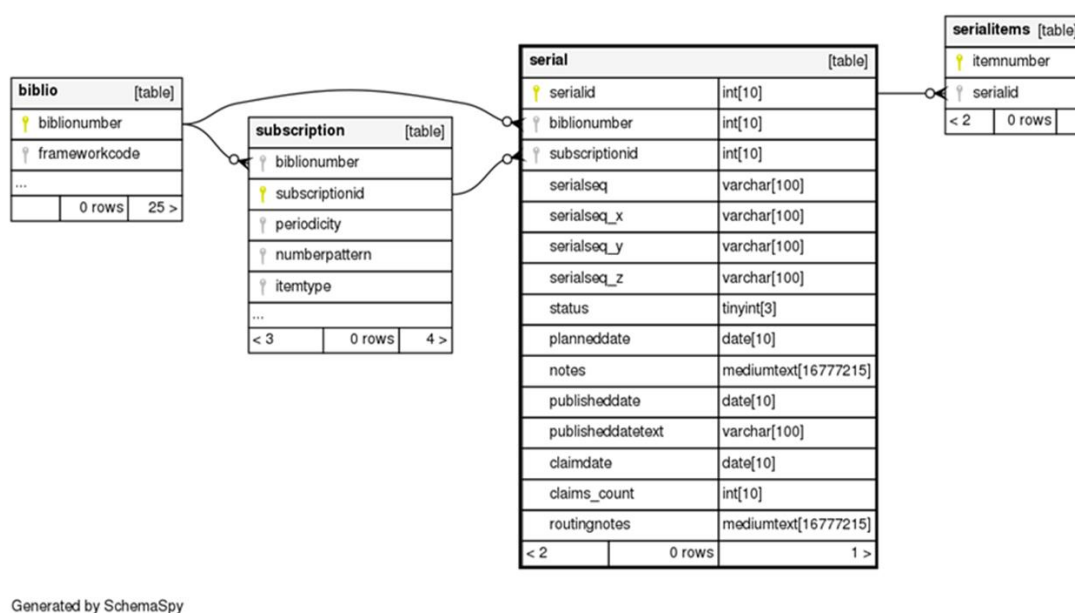


Figura 2 – Schema da Base de dados do Koha – *Serial* table (Koha Community, s.d.)

Por esse motivo, ele não tem uma forma direta para exportar os dados do módulo de assinaturas a partir do Interface dos técnicos: os métodos de exportação de dados e de guardar registos baseiam-se nos campos MARC e a exportação e dados em CSV apenas exporta campos mapeados para MARC. Apesar das opções das Assinaturas permitirem que se crie um exemplar MARC a partir da página da edição de uma existência (*holding*), isto não representa um mapeamento. Na verdade, a opção permite somente o arranque do



HTML da criação de um novo exemplar, com cópia dos dados nessa página HTML de edição, o que implica também a impossibilidade de o fazer em lote. Verificámos ainda que a comunidade nacional e internacional não apresenta soluções claras para que os profissionais de Informação e Documentação possam fazer a exportação dos registos das publicações periódicas com as existências alimentadas no módulo das assinaturas (Fischer, 2024).

Adotando uma abordagem mista, pesquisámos no Google por propostas de solução² e identificámos que a comunidade internacional partilha relatórios que extraem listas de todas as existências das assinaturas abertas e das assinaturas fechadas, com a informação da numeração dos exemplares atribuída na coluna do *Schema serial.serialseq* (Alway, s.d.). Por não ser suficiente para uma descrição completa, editámos o SQL e acrescentámos a informação da data de publicação (*serial.publisheddate*) e o estado de cada uma das existências (*serial.status*), o que permite preservar toda a informação necessária de uma existência de periódico.

A partir desses relatórios, que exportam em bruto a informação, desenvolvemos através de um processo de *prompt engineering* (Schulhoff, 2024, 7 p.) instruções para, com a Inteligência Artificial (IA) Microsoft Copilot, editar os dados migrando-os para o formato UNIMARC, em MARCXML. A utilização da IA permitiu o tratamento em lote de um grande volume de dados, cerca de 12000 registos e 85000 exemplares, migrando com sucesso a informação para o campo dos exemplares UNIMARC (995); a edição em lote dos subcampos que haviam sido gerados com informação genérica (Cota, Localização na prateleira, Restrições de utilização), a partir de um exemplar modelo por título; e a intercalação dos campos 995 criados com os registos dos periódicos extraídos do Koha.

Durante todo este processo, sentimos a necessidade de resiliência e insistência no uso da tecnologia de modo a superar as suas limitações e alucinações, como é sugerido na Figura 3.

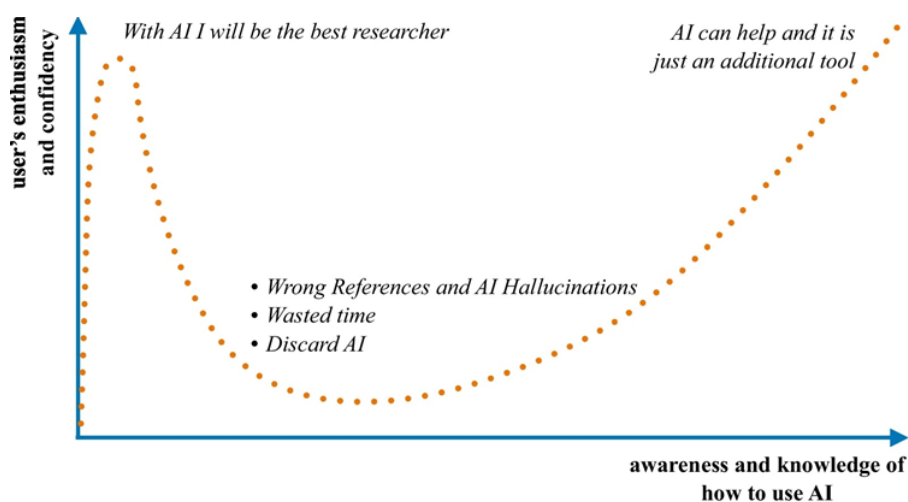


Figura 3 – Efeito Dunning-Kruger adaptado à utilização de IA (Salvagno et al., 2023)

² Pesquisa realizada entre 15/05/2025 e 15/07/2025, com os termos “export” AND “Koha” AND “serials” ou “export” AND “Koha” AND “subscriptions”.



No tratamento deste grande volume de dados, foi necessário dividir alguns dados em diferentes ficheiros, de modo a conseguir tratá-los na IA e, apesar da grande vantagem do uso de linguagem natural, foi necessário usar termos claros e explícitos, bem como *prompts* estruturadas para executar os comandos de tratamento de tabelas de dados.

Apesar dos vários desafios encontrados, estas ferramentas permitem a exportação e preservação dos dados produzidos no módulo de assinaturas do Koha a partir do Interface dos técnicos, com a sua conversão para MARC, garantindo a sua preservação. Através da *prompt engineering*, recorrendo somente à linguagem natural acessível a qualquer profissional de Documentação e Informação, e da vantagem de apresentar custos muito mais reduzidos que os valores de uma empresa externa de informática para realizar o serviço, este trabalho demonstra que a IA pode capacitar as bibliotecas de autonomia para elaborar um plano de preservação digital, incluindo agora os dados alimentados no módulo de assinaturas.

Referências bibliográficas

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

Brzustowicz, R. (2023). From ChatGPT to CatGPT: The implications of artificial intelligence on library cataloging. *Information Technology and Libraries*, 42(3). <https://doi.org/10.5860/ital.v42i3.16295>

Cox, A. M., & Mazumdar, S. (2024). Defining artificial intelligence for librarians. *Journal of Librarianship and Information Science*, 56(2), 330–340. <https://doi.org/10.1177/09610006221142029>

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

Fischer, K. (2024, October 18). Bug 38202 – Add table configuration and export options to table of subscriptions. *Bugzilla - Koha Community*. Acedido a 15-07-2025, de https://bugs.koha-community.org/bugzilla3/show_bug.cgi?id=38202

K. Vrindha & Syamili C (2025) Navigating the AI Landscape in Libraries: A PRISMA-Based Systematic Analysis of AI Applications in Libraries. *Journal of Web Librarianship*, 19:1, 45-61, DOI: 10.1080/19322909.2025.2468697

Koha Community (s.d.). serial. *schema.koha-community.org*. 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>



6 e 7 de novembro de 2025, Oeiras

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). *The prompt report: a systematic survey of prompt engineering techniques*. arXiv preprint arXiv:2406.06608