PERPETUA PROJECT: USING PYTHON LIBRARIES IN A WORKFLOW FOR PRESERVING LEGACY ARCHIVAL COLLECTIONS

PROJETO PERPETUA: USO DE BIBLIOTECAS PYTHON EM FLUXO DE TRABALHO PARA PRESERVAÇÃO DE ACERVOS ARQUIVÍSTICOS LEGADOS

Murilo Souza dos Santos[1][0000-0002-6856-4561] and Andressa Cristiani Piconi[1][0000-0001-6734-4770]

[1] State University of Campinas, Brazil

murilos@unicamp.com, andressa@unicamp.com

Abstract. The project aims to promote the digital preservation of legacy document collections within an archival institution facing challenges common to many small and medium-sized entities. The typical scenario involves collections comprising both paper records and previously digitized materials, often lacking adequate control over the correspondence between digital files and their descriptive metadata. This leads to duplication and redundant versions of digital records, hindering effective information retrieval.The proposed solution centers on reusing existing digitizations, regardless of their lack of standardization or prior organization, thus avoiding costly re-digitization processes. The adopted methodology is adaptable and semi-automated, using open-source libraries, allowing institutions to structure and manage their digital collections without placing a heavy burden on their resources.Moreover, the model contributes to the standardization of archival workflows for digital preservation, supporting institutional collaboration and the adoption of accessible, cost-effective strategies. It aligns with institutional preservation policies and complies with both national and international standards.Preliminary results confirm the feasibility of the approach, with the main advantage being the reuse of prior digitizations without the need to repeat processing steps. This supports the creation of standardized records, enhances traceability, and strengthens archival control. The conclusion underscores the critical role of preservation policies and standards beginning at the point of record creation. It emphasizes that it is indeed possible to apply technical criteria that ensure the long-term integrity, authenticity, and accessibility of digital collections without the need for costly disposal or reprocessing efforts.

Keywords: Digital Preservation, Record Management, Task Automation. Python Library, Information Package.

Resumo. O projeto visa promover a preservação digital de conjuntos documentais legados em uma instituição arquivística que enfrenta desafios comuns a muitas outras entidades de pequeno e médio porte. O cenário típico envolve acervos com documentos em papel e digitalizações realizadas sem o devido controle sobre a correspondência entre os arquivos digitais e suas informações descritivas. Esse problema gera duplicações e versões redundantes dos documentos digitais, dificultando a recuperação da informação. A solução proposta é reaproveitar digitalizações existentes, mesmo sem padronização ou organização prévia, evitando a re-digitalização e seus altos custos. A metodologia adotada é adaptável e semi-automatizada, utilizando bibliotecas open source, permitindo que as instituições estruturem seus acervos digitais sem comprometer recursos. O modelo também contribui para a padronização de práticas em um fluxo de trabalho de gestão arquivística e preservação digital, facilitando a colaboração entre instituições e a adoção de estratégias acessíveis e eficazes, seguindo a política de preservação da instituição, de acordo com padrões nacionais e internacionais. Os resultados, até o momento, mostram a viabilidade da abordagem cuja principal vantagem é o reaproveitamento das digitalizações anteriores, sem a necessidade de refazer os processos, mantendo registros padronizados e melhorando a rastreabilidade e controle arquivístico. A conclusão reafirma a importância de políticas de preservação e de padrões desde a produção do documento, destacando que é possível aplicar critérios técnicos que assegurem a integridade, autenticidade e acessibilidade dos acervos digitais a longo prazo, sem a necessidade de descartes ou reprocessamentos dispendiosos.

Palavras-chave: Preservação Digital, Gestão Arquivística, Automação de Processos, Bibliotecas Python, Pacote de informação.

References

1. ISO. Space data system practices — Reference model for an open archival information system (OAIS). ISO 14721:2025(en). Geneva: International Organization for Standardization (2025).

2. BRASIL. Conselho Nacional de Arquivos (CONARQ). Resolução Conarq nº48, de 10 de novembro de 2021. Estabelece diretrizes e orientações aos órgãos e entidades integrantes do Sistema Nacional de Arquivos quanto aos procedimentos técnicos a serem observados no processo de digitalização de documentos públicos ou privados. Disponível em: <https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/Diretrizes_digitalizacao__2021.pdf>. Acesso em: 03 jun. 2024.

3. BRASIL. Conselho Nacional de Arquivos (CONARQ). Resolução nº 51, de 25 de agosto de 2023. Dispõe sobre as "Diretrizes para a Implementação de Repositórios Arquivísticos Digitais Confiáveis", Versão 2. Disponível em: <https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-c onarq-no-51-de-25-de-agosto-de-2023>. Acesso em: 03 jun. 2024.

4. CHIN. Canadian Heritage Information Network. Digital Preservation Toolkit. 2021-11-25. Disponível em: <https://www.canada.ca/en/heritage-information-network/services/digital-preservation/tool kit.html>. Acesso em: 23 de maio de 2025.