

Preservação Digital na Universidade de Aveiro: Estratégias e Desafios

Ana Cristina Fernandes Cortês Santana Justino

Universidade de Aveiro

Rafael Direito

Universidade de Aveiro, rafael.neves.direito@ua.pt

Luísa Falcão

Universidade de Aveiro, lfalcao@ua.pt

Isaac Raimundo

Universidade de Aveiro, isaac.raimundo@ua.pt

Pedro Lobo

Universidade de Aveiro, pedro.lobo@ua.pt

Eduardo Santos

Universidade de Aveiro, eduardosantoshf@ua.pt

Filipe Trancho

Universidade de Aveiro, ftrancho@ua.pt

Filipe Bento

Universidade de Aveiro, fsb@ua.pt

José Vieira

Universidade de Aveiro, jnvieira@ua.pt

Plano de Preservação Digital

Código aberto

Inteligência artificial

Acesso à informação

Num cenário de proliferação de conteúdos digitais, a preservação digital apresenta desafios significativos para garantir o acesso e usabilidade a longo prazo (Digital Preservation Coalition, 2015). Neste sentido, a Universidade de Aveiro (UA) tem vindo a desenvolver ao longo dos últimos dois anos um Plano de Preservação Digital (PPD) visando assegurar a longevidade, autenticidade e acessibilidade dos seus recursos digitais. Este artigo apresenta uma visão geral das estratégias e desafios da preservação digital, com foco no uso de tecnologias de código aberto, como o Archivematica e o Atom. Destaca-se a implementação de fluxos de trabalho para a ingestão de objetos digitais, desde a sua aquisição até ao acesso, em conformidade com o modelo funcional ISO-OAIS. O PPD da UA identifica e caracteriza a informação digital produzida, categorizando-a de acordo com necessidades operacionais e atribuindo prazos de conservação⁴. Determina procedimentos para evitar a obsolescência tecnológica e clarifica responsabilidades dentro da organização. Uma inovação significativa é a integração de inteligência artificial (IA) na transcrição de arquivos audiovisuais, otimizando o acesso e a descoberta de informação. Esta abordagem facilita a pesquisa e análise de grandes volumes de conteúdo audiovisual, tornando-o mais acessível e utilizável. O plano também aborda a gestão da mudança organizacional, atribuindo novas responsabilidades e imputando-as aos vários intervenientes no processo de preservação digital. Conclui-se que a integração de tecnologias de código aberto e a adoção de boas práticas são fundamentais para que o PPD possa assegurar a transmissão do conhecimento produzido, tanto no presente como no futuro. Este esforço garante o acesso através da curadoria, cumpre responsabilidades legais e contribui para a construção da memória coletiva e social da instituição.

Digital Preservation at the University of Aveiro: Strategies and Challenges

Digital Preservation Plan

Open source

Artificial intelligence

Access to information

In a scenario of proliferating digital content, digital preservation presents significant challenges to ensure long-term access and usability (Digital Preservation Coalition, 2015). In this regard, the University of Aveiro (UA) has been developing a Digital Preservation Plan (DPP) over the last two years to ensure the longevity, authenticity and accessibility of its digital resources. This article presents an overview of digital preservation strategies and challenges, focusing on the use of open-source technologies such as Archivematica and Atom. It highlights the implementation of workflows for

ingesting digital objects, from acquisition to access, in accordance with the ISO-OAIS functional model. The UA DPP identifies and characterizes the digital information produced, categorizing it according to operational needs and assigning retention periods. It determines procedures to avoid technological obsolescence and clarifies responsibilities within the organization. A significant innovation is the integration of artificial intelligence (AI) in the transcription of audiovisual archives, optimizing access to and discovery of information. This approach makes it easier to search and analyze large volumes of audiovisual content, making it more accessible and usable. The plan also addresses the management of organizational change, assigning new responsibilities and allocating them to the various players in the digital preservation process. It concludes that the integration of open-source technologies and the adoption of good practices are fundamental if the PPD is to ensure the transmission of the knowledge produced, both now and in the future. This effort guarantees access through curation, fulfils legal responsibilities and contributes to building the institution's collective and social memory.

1. CONTEXTO

A preservação digital, num cenário caracterizado pela crescente proliferação de conteúdos digitais, emerge como um campo complexo e em constante evolução, apresentando desafios significativos para assegurar o acesso e o uso da informação a longo prazo (Digital Preservation Coalition, 2015). Neste sentido, o Plano de Preservação Digital (PPD) deverá incluir linhas estratégicas, nos vários domínios do conhecimento. Nomeadamente, na componente tecnológica, diretivas que incluem a migração e emulação para garantir a longevidade dos objetos digitais (Becker et al., 2009; Rothenberg, 1999); no domínio da Ciência da Informação, a gestão do ciclo de vida da informação desde a sua criação até ao seu arquivo (Constantopoulos et al., 2009), a criação de metadados (ou metainformação) e documentação descritiva para assegurar a acessibilidade e autenticidade dos registos a longo prazo (Giaretta, 2011; Lemieux, 2016); bem como, numa perspetiva de governança, o desenvolvimento de modelos económicos e políticas institucionais que promovam a sustentabilidade dos esforços de preservação (Lavoie, 2008; Walters & Skinner, 2011).

Neste sentido, conscientes das inúmeras dificuldades que o processo de Preservação Digital representa, mas também sensibilizada

para a necessidade e urgência de garantir o acesso à Memória Institucional a Universidade de Aveiro (UA) deu início à elaboração de um PPD, com o objetivo de garantir a transmissão do conhecimento produzido para as gerações futuras. O PPD inclui a implementação de um fluxo de trabalho (workflow) para a ingestão de objetos digitais, utilizando tecnologias de código aberto como o Archivematica e o Atom, tendo como foco os fundos e as coleções especiais da Instituição, tanto as nascidas digitalmente (com prioridade dos recursos digitais que correm o risco de se tornarem obsoletos ou inacessíveis) como as convertidas em formato digital (priorizando os fundos/coleções especiais com a maior avaliação, após classificação numa tabela de avaliação).

2. DESENVOLVIMENTO DO PLANO DE PRESERVAÇÃO DIGITAL

O plano de Preservação Digital apresenta as linhas estratégicas adotadas pela UA para a gestão, preservação e disponibilização dos objetos digitais, previamente selecionados, de acordo com a importância patrimonial, histórica e valor legal, para a preservação a longo prazo, garantindo a sua autenticidade, integridade, usabilidade e fiabilidade (tabela de avaliação).

Para tal, a UA estabeleceu um conjunto de práticas e princípios que refletem as melhores normas internacionais no domínio da preservação digital, nomeadamente esclarece sobre o que entende por recurso digital, os princípios-chave e os requisitos da preservação digital.

2.1. Definição de Recursos Digitais

Entende-se por recursos digitais todos os objetos que nasceram digitalmente, através do uso de software e hardware, bem como aqueles que foram convertidos a partir de recursos físicos tangíveis, por meio de técnicas de digitalização ou fotografia. O processo de preservação digital reflete o ciclo de vida de um objeto digital, desde a sua criação e custódia até ao depósito na plataforma de preservação, integrando-se num contexto mais amplo de património cultural e académico.

Na UA, os recursos digitais abrangem uma variedade de formatos, incluindo texto, gráficos, imagem, áudio, bases de dados, páginas web e emails. Para garantir a sustentabilidade destes formatos, a UA compromete-se a efetuar a sua conversão sempre que necessário, seguindo as boas práticas internacionais, como as definidas pela Library of Congress (<https://www.loc.gov/preservation/digital/formats/>).

2.2. Princípios-Chave da Preservação Digital

A UA baseia as suas atividades de preservação digital em três princípios fundamentais, alinhados com as normas e boas práticas internacionais:

- a. Autenticidade: A UA realiza auditorias regulares aos objetos digitais para garantir a sua autenticidade.
- b. Longevidade: Sempre que são detetadas alterações ou degradações, são

implementadas ações corretivas para assegurar a preservação dos recursos.

- c. Integridade: A UA aplica uma série de ações no sistema de preservação digital para verificar e manter a integridade dos objetos digitais.

2.3. Requisitos para a Preservação Digital

Os requisitos incluem as tarefas de seleção, inventariação, estratégia, planeamento, acesso e uso, bem como os direitos/licenças.

2.3.1. Seleção de Recursos Digitais

A seleção de recursos digitais para preservação a longo prazo baseia-se nos seguintes critérios:

- a. Recursos digitais produzidos no âmbito das funções administrativas da UA, indicados pelo RADA (Regulamento de Arquivo e Documentação Administrativa) como de preservação permanente.
- b. Recursos digitais que refletem a história e memória da UA, incluindo produção académica e científica (teses, artigos, datasets) e eventos de carácter académico e cultural.
- c. Recursos digitais externos à UA, mas integrados nos seus fundos por aquisição, oferta ou doação, desde que relevantes para as áreas de estudo e investigação da academia.
- d. Objetos físicos digitalizados pela UA com os seguintes objetivos:
 - i. Preservação do objeto físico, devido ao seu estado de degradação ou formato obsoleto.
 - ii. Facilitação do acesso para divulgação e investigação.
 - iii. Promoção das coleções e fundos patrimoniais da UA.

2.3.2. Inventariação

No processo de introdução dos recursos digitais na plataforma de preservação, é essencial que estes estejam devidamente documentados,

mantendo a sua proveniência custodial. Os recursos são sujeitos a uma inspeção de antivírus e seguem os seguintes passos:

- a. Identificação, caracterização e validação dos formatos de ficheiro.
- b. Registo de metadados descritivos e administrativos.
- c. Verificação da autenticidade dos recursos.
- d. Produção de cópias de preservação e de acesso.

2.3.3. Estratégia de Preservação

A UA preserva o original em formato matriz (de maior qualidade), bem como outras manifestações resultantes do processo de preservação. Para tal, foram criados workflows de apoio à planificação e ingestão dos recursos digitais, suportados por infraestruturas tecnológicas adequadas.

2.3.4. Planeamento

O planeamento é crucial para garantir a preservação dos recursos digitais. Este processo inclui a consideração de dimensões legais, financeiras, tecnológicas e institucionais, bem como a definição de estratégias para mitigar riscos de acessibilidade. A UA realiza ações periódicas de:

- a. Avaliação de risco (tipo de ficheiro e opções de preservação).
- b. Monitorização de tecnologias (para acompanhar a evolução tecnológica).
- c. Correção (definição de soluções adequadas à preservação, com base na avaliação e evolução tecnológica).

2.3.5. Acesso e Uso

O catálogo de recursos digitais está disponível para pesquisa, com conteúdos organizados em:

- a. Acesso aberto.
- b. Acesso restrito à comunidade UA.
- c. Conteúdos em acesso restrito.
- d. Conteúdos interditos.

2.3.6. Direitos/Licenças

A UA disponibiliza os recursos digitais em formatos de código aberto, como o Dublin Core. O acesso pode estar condicionado por:

- a. Inclusão de dados sensíveis.
- b. Proteção por direitos de autor.
- c. Abrangimento por protocolos ou contratos.
- d. Sempre que possível, a UA disponibiliza os recursos em acesso aberto, sob licenças Creative Commons.

3. FLUXO DE TRABALHO PARA INGESTÃO DE OBJETOS DIGITAIS

Após a seleção do fundo/coleção especial a preservar — correspondente à Pasta 1 da série “**Cerimónias Académicas**” (60 fotografias) —, iniciou-se o processo de descrição dos metadados (metainformação) num ficheiro Excel (Figura 1).

| Tipo de objeto | Descrição | Data | Assunto | Assunto |
|----------------|--------------------------------|------------------------------|------------------------|------------|
| Objeto Risco | PT_04_00004_UA_029_FT_UA_1_001 | Descrição de entrada de 1980 | Universidade de Aveiro | Entrada de |
| Objeto Risco | PT_04_00004_UA_029_FT_UA_1_002 | Descrição de entrada de 1980 | Universidade de Aveiro | Entrada de |
| Objeto Risco | PT_04_00004_UA_029_FT_UA_1_003 | Descrição de entrada de 1980 | Universidade de Aveiro | Entrada de |
| Objeto Risco | PT_04_00004_UA_029_FT_UA_1_004 | Descrição de entrada de 1980 | Universidade de Aveiro | Entrada de |
| Objeto Risco | PT_04_00004_UA_029_FT_UA_1_005 | Descrição de entrada de 1980 | Universidade de Aveiro | Entrada de |

Figura 1 - campos do ficheiro Excel

Para otimizar o acesso à informação e facilitar sua recuperação, recorreremos a ferramentas de **Inteligência Artificial (IA)**. Entre as opções disponíveis, optámos por soluções locais:

- **Jan AI e AnythingLLM** (baseado em modelos *LLaVA*), utilizadas para análise/síntese de imagens e extração de dados.
- Limitação: Embora multimodais (texto + imagens), estas ferramentas têm funcionalidades restritas para análise puramente visual — modelos

mais leves não interpretam imagens sem texto associado.

Após o preenchimento do Excel, preparámos a estrutura arquivística no ATOM para ingestão em bloco. O fluxo de trabalho (Figura 2) inclui:

- Organização, classificação, armazenamento e difusão de objetos digitais na plataforma **Archivematica**, com publicação final no ATOM.

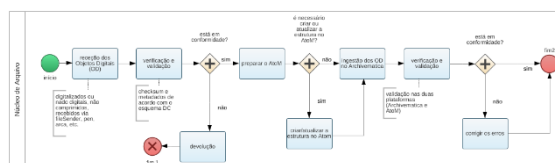


Figura 2 - fluxo de ingestão

À semelhança do processo com a série “Cerimónias Académicas”, utilizámos IA para descrever o arquivo audiovisual:

- **Whisper** (ferramenta offline) para transcrição automática de entrevistas, podcasts e gravações académicas.
- Resultado: Mais de **1.716** horas de conteúdo transcrito, incluindo o programa 3810UA.

4. RESULTADOS E DISCUSSÃO

Da *praxis* realizada verificámos que as ferramentas de IA reduziram em 40% o tempo de descrição manual (métrica comparativa com processos anteriores), bem como observámos que, com o uso do Whisper, obtivemos uma precisão em transcrições de 85% (avaliado por amostragem).

Todavia, identificámos algumas fragilidades, nomeadamente uma precisão variável na análise de imagens sem texto, bem como a necessidade de revisão humana para validação dos conteúdos.

CONCLUSÃO

A preservação digital na UA envolve uma combinação de estratégias para enfrentar desafios tecnológicos, organizacionais e éticos. A implementação de um fluxo de trabalho para a ingestão de objetos digitais e o uso de tecnologias de código aberto, como o Archivematica e o Atom, são essenciais para garantir a longevidade e acessibilidade dos recursos digitais.

A integração de IA na transcrição dos fundos/coleções especiais oferece soluções promissoras, mas também exige uma consideração cuidadosa da precisão, implicações éticas e integração de sistemas. Ao enfrentar estes desafios, a UA pode garantir a preservação e acessibilidade a longo prazo de informação digital valiosa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ahmad, R., & Rafiq, M. (2023). The digital preservation 101 guide for small libraries. *Library Hi Tech News*, 40(9), 12–14.
- Becker, C., Rauber, A., Heydegger, V., Schnasse, J., & Thaller, M. (2008). A generic XML language for characterising objects to support digital preservation. *Proceedings of the ACM Symposium on Applied Computing*, 402–406.
- Bosi, M., Canazza, S., Pretto, N., Russo, A., & Spanio, M. (2024). From Tape to Code: An International AI-Based Standard for Audio Cultural Heritage Preservation. *IEEE Access*, 12, 152544–152558.
- Burda, D., & Teuteberg, F. (2013). Sustaining accessibility of information through digital preservation: A literature review. *Journal of Information Science*, 39(4), 442–458.
- Corrado, E. M. (2022). Digital Preservation Is Not Just a Technology Problem. *Technical Services Quarterly*, 39(2), 143–151.
- El Idrissi, B. (2019). Long-term digital preservation: A preliminary study on software and format obsolescence. *ACM International Conference Proceeding Series*.
- Nguyen, Q. L. (2013). Towards a large access-oriented digital archive. *Archiving 2013 - Final Program and Proceedings*, 70–73.
- Pacheco, A. (2018). The fair accessor as a tool to reinforce the authenticity of digital archival information. *Lecture Notes in Computer Science*, 11057 LNCS, 317–320.
- Subbulakshmi, B., et al. (2023). Efficient AI-Powered Audio-to-Text Transcription: A GUI-Enhanced Stack with EXE Build for Innovation in Communications. *2023 International Conference on Energy, Materials and Communication Engineering*.
- Yue, M., Xueyang, C., & Ziyun, Q. (2025). A conceptual framework for the path of digital preservation of intangible cultural heritage: A thematic review. *Multidisciplinary Reviews*, 8(2).

Justino, Ana Cristina Fernandes Cortês Santana; Falcão, Luísa; Lobo, Pedro; Tranco, Filipe; Vieira, José; Direito, Rafael; Raimundo, Isaac; Santos, Eduardo; Bento, Filipe (2025). "Preservação Digital na Universidade de Aveiro: Estratégias e Desafios". Cadernos BAD, n. 1-2. <https://doi.org/10.48798/bad3093>

Acesso e licença

Artigo em acesso aberto distribuído nos termos da licença Creative Commons Atribuição 4.0 Internacional (CC-by 4.0).

Conflitos de Interesse

Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.

Revisão por Pares

Esta revista usa um sistema de revisão duplamente cega por pares assegurada pelo conselho científico da Cadernos BAD.

Confidencialidade dos Dados

Os autores declaram ter seguido os protocolos de RGPD.

Financiamento, apoio e patrocínios

Recebido

23/02/2025

Aceite

17/03/2025

Publicado

27/07/2025
