# PRESERVAÇÃO ÉTICA DE DADOS SENSÍVEIS EM ARQUIVOS DIGITAIS: MEMÓRIA COLETIVA, DIREITOS HUMANOS E INTELIGÊNCIA ARTIFICIAL

**Catarina Gonçalves[1], Francisca Carvalho[2], Jorge Martins[3], Pedro Oliveira[4], Ana Paula Camarinha[5,6]**

*1,2,3,4,5 Instituto Superior de Contabilidade e Administração do Porto (Portugal)*
*6 CEOS-PP*

## Resumo

Este artigo analisa os desafios éticos, legais e tecnológicos na preservação de dados sensíveis em arquivos digitais, com foco em justiça transicional e direitos humanos. O objetivo é investigar como equilibrar o acesso à memória histórica com a proteção da privacidade, considerando o RGPD e o impacto da Inteligência Artificial (IA). A pesquisa combina revisão de literatura e estudo de caso do Bundesarchiv, que evidencia riscos à privacidade e desafios da digitalização, destacando as soluções adaptadas durante a pandemia de COVID-19. Conclui-se que políticas claras de segurança e anonimização são essenciais para garantir a preservação ética. O artigo contribui com diretrizes úteis para entidades que gerem e preservam dados sensíveis em formato digital. Adicionalmente, enfatiza-se a necessidade de abordagens arquivísticas sensíveis ao trauma e baseadas na ética do cuidado, conforme discutido por Caswell e Cifor (2019), destacando que processos de preservação que ignoram o impacto emocional e social dos arquivos sobre comunidades afetadas correm o risco de reproduzir danos históricos e institucionais.

**Palavras-chave**: Preservação digital, dados sensíveis, memória coletiva, direitos humanos, RGPD, inteligência artificial, privacidade, arquivos digitais.

## Abstract

**This article analyzes the ethical, legal, and technological challenges in preserving sensitive data in digital archives, focusing on transitional justice and human rights. The objective is to investigate how to balance access to historical memory with privacy protection, considering the GDPR and the impact of Artificial Intelligence (AI). The research combines a literature review and a case study of the Bundesarchiv, highlighting privacy risks and challenges of digitization, emphasizing the solutions adapted during the COVID-19 pandemic. It concludes that clear security and anonymization policies are essential to guarantee ethical preservation. The article contributes useful guidelines for entities that manage and preserve sensitive data in digital format. Additionally, it emphasizes the need for trauma-sensitive archival approaches based on an ethics of care, as discussed by Caswell and Cifor (2019), highlighting that preservation processes that ignore the emotional and social impact of archives on affected communities risk reproducing historical and institutional harm.**

**Keywords**: Digital preservation, sensitive data, collective memory, human rights, GDPR, artificial intelligence, privacy, digital archives.

# 1 INTRODUÇÃO

A digitalização de acervos históricos tem ampliado o acesso à memória coletiva, mas simultaneamente colocado em evidência desafios éticos e legais associados ao tratamento de dados sensíveis, particularmente em contextos de justiça transicional e direitos humanos. A incorporação de tecnologias como a Inteligência Artificial (IA) nesses processos intensifica tais desafios, ao automatizar operações de organização e acesso, aumentando tanto a eficiência quanto os riscos de exposição indevida, viés algorítmico e redução da transparência. Assim, a preservação digital exige um equilíbrio rigoroso entre o direito ao acesso à memória histórica e a salvaguarda da privacidade, em conformidade com enquadramentos normativos como o Regulamento Geral de Proteção de Dados (RGPD).

A literatura recente (Gilliland, 2022; Shilton, 2023) evidencia que tal equilíbrio depende também de uma compreensão crítica dos contextos sociais e políticos onde a preservação ocorre, defendendo que decisões arquivísticas têm impacto direto na justiça social, especialmente em arquivos relacionados a minorias, vítimas de violência estatal e comunidades historicamente marginalizadas.

# 2 QUESTÃO DE INVESTIGAÇÃO

Este estudo examina os desafios éticos, legais e tecnológicos inerentes à preservação de dados sensíveis em arquivos digitais, com particular atenção a contextos de justiça transicional e direitos humanos. Analisa-se a forma como as instituições arquivísticas podem conciliar o acesso à memória coletiva com a proteção da privacidade, considerando simultaneamente as exigências do Regulamento Geral de Proteção de Dados (RGPD) e o impacto de tecnologias emergentes, nomeadamente a Inteligência Artificial.

A questão de investigação que orienta este trabalho é: Como podem as instituições arquivísticas assegurar uma preservação ética e segura de dados sensíveis em ambientes digitais?

Esta questão relaciona-se ainda com debates contemporâneos sobre governança de dados, particularmente no que diz respeito à capacidade das instituições em adotar modelos de gestão orientados por princípios de accountability, privacy-by-design e de avaliação contínua de risco (Voigt & von dem Bussche, 2021; Rinehart, 2022).

# 3 METODOLOGIA

A investigação adotou uma abordagem qualitativa, articulando revisão de literatura e estudo de caso, de modo a analisar de forma integrada os desafios associados à preservação ética de dados sensíveis em arquivos digitais.

A revisão de literatura baseou-se em fontes académicas e documentos normativos selecionados pela sua relevância, atualidade e contributo para os debates sobre preservação digital, proteção de dados sensíveis e impacto da IA. As obras analisadas permitiram identificar enquadramentos conceptuais, requisitos legais e abordagens tecnológicas centrais ao tema.

O estudo de caso incidiu sobre o Bundesarchiv, escolhido pela sua relevância institucional, pela gestão de acervos historicamente sensíveis e pela implementação de práticas alinhadas com o RGPD. A análise fundamentou-se em documentos institucionais, relatórios públicos e estudos especializados, permitindo observar soluções aplicadas à preservação ética de dados sensíveis.

A análise foi conduzida por interpretação qualitativa e triangulação entre literatura científica e documentação institucional, identificando padrões, desafios e princípios orientadores aplicáveis à gestão ética de dados sensíveis em ambientes digitais.

A revisão de literatura foi complementada com fontes recentes sobre riscos de reidentificação em datasets supostamente anonimizados, um tema central quando se tratam de dados sensíveis históricos (Rocher, Hendrickx & de Montjoye, 2019). A integração metodológica beneficiou igualmente de referenciais de análise oriundos dos Critical Archival Studies, que enfatizam a necessidade de considerar o papel do poder e da desigualdade na construção, preservação e acesso aos arquivos (Evans, Wallace & Rawson, 2023).

# 4 RESULTADOS

A articulação entre a literatura analisada e as evidências observadas no estudo de caso do Bundesarchiv indica que a preservação de dados sensíveis em arquivos digitais continua marcada por desafios éticos, normativos e tecnológicos de elevada complexidade. Embora os processos de digitalização e os mecanismos de acesso à memória coletiva tenham evoluído de forma significativa, persistem riscos relevantes para a proteção da privacidade e dúvidas quanto ao equilíbrio adequado entre transparência histórica e salvaguarda de informação pessoal. Os resultados obtidos podem ser organizados em quatro áreas centrais: (1) questões éticas, (2) exigências decorrentes do RGPD, (3) influência da Inteligência Artificial (IA) e (4) práticas institucionais de preservação.

## 4.1 Desafios Éticos na Preservação de Dados Sensíveis

Os desafios éticos associados à preservação de dados sensíveis resultam do constante equilíbrio entre o dever de garantir o acesso à memória histórica e a necessidade de proteger a privacidade individual. Embora a digitalização amplie significativamente a circulação e utilização de arquivos culturais e históricos, também aumenta a exposição de indivíduos e comunidades,

sobretudo em contextos ligados à justiça transicional e a registos de abusos de direitos humanos.

A literatura destaca três dilemas centrais:

• **Risco de revitimização**: relatos de vítimas de violência ou discriminação podem causar danos se acessados indiscriminadamente (Bantin & Zhao, 2022).

• **Proteção da privacidade**: é essencial definir critérios claros para anonimização e controle de acesso (Yakel & Ketelaar, 2023).

• **Equilíbrio entre memória e ética**: políticas institucionais devem garantir preservação histórica sem comprometer direitos individuais, conforme reforçado por diretrizes internacionais como as da UNESCO.

O estudo de caso do Bundesarchiv confirma a relevância destes desafios, evidenciando que decisões relativas à seleção documental, aos modelos de controlo de acesso e aos procedimentos de anonimização são fundamentais para assegurar uma preservação eticamente responsável.

Autores como Caswell e Cifor (2019) defendem que a ética arquivística deve incluir explicitamente práticas de cuidado e empatia, salientando que decisões técnicas ou legais têm

consequências humanas profundas. Essa perspetiva reforça a ideia de que a preservação ética exige não só conformidade normativa, mas também sensibilidade social e histórica.

## 4.2 Implicações do Regulamento Geral de Proteção de Dados (RGPD) na Preservação Digital

O Regulamento Geral de Proteção de Dados (RGPD) introduz exigências rigorosas no tratamento de informação pessoal e sensível, condicionando diretamente os processos de preservação digital em arquivos históricos. Estas restrições afetam tanto a disponibilização pública como a integridade dos registos preservados.

• **Consentimento explícito**: dificulta digitalização de documentos históricos contendo dados de vítimas de violações de direitos humanos (Kosta & Borgesius, 2024).

• **Anonimização e integridade histórica**: medidas de proteção podem enfraquecer a completude das informações preservadas (Ferreira & Martins, 2020; Borges, 2014).

• **Conflito regulatório**: normas de privacidade podem limitar o acesso público, exigindo soluções institucionais para conciliar ambos os direitos.

O exemplo do Bundesarchiv evidencia estas dinâmicas, ao demonstrar como políticas internas de anonimização, controlo de acessos e avaliação de risco são ajustadas para garantir conformidade com o RGPD sem comprometer a função histórica e documental dos arquivos.

Investigadores como Schermer (2021) e Yakel & Ketelaar (2023) chamam a atenção para a insuficiência das exceções arquivísticas do RGPD, que continuam juridicamente ambíguas e aplicadas de modo desigual pelos Estados-membros, o que cria desigualdades no acesso à memória histórica dentro da União Europeia.

## 4.3 Impacto da Inteligência Artificial na Organização e Acesso aos Arquivos Digitais

O uso crescente da Inteligência Artificial (IA) na organização e gestão de arquivos digitais tem o potencial de melhorar significativamente a eficiência na indexação e no acesso a grandes volumes de dados, tornando os arquivos históricos mais rápidos e eficazes para pesquisa e análise. No entanto, a utilização da IA também levanta riscos consideráveis, especialmente em relação à privacidade e à segurança dos dados sensíveis.

Estudos de Silva Júnior e Borges (2016) e Kosta e Borgesius (2024) alertam para os perigos da automação do acesso a dados sem a supervisão adequada. A IA, ao automatizar o processo de organização e indexação, pode inadvertidamente expor dados sensíveis sem a devida proteção ou controle de acesso, aumentando o risco de violações de privacidade. Embora a IA seja uma ferramenta poderosa para aumentar a acessibilidade, ela também pode ampliar os riscos se os sistemas não forem projetados para garantir a segurança e a conformidade com os direitos de privacidade.

Além disso, a falta de transparência algorítmica é um problema crítico identificado na literatura. Silva Júnior e Borges (2016) argumentam que a opacidade nos algoritmos pode introduzir viéses no tratamento dos dados, prejudicando a precisão das informações preservadas e comprometendo a equidade no acesso aos arquivos. Kosta e Borgesius (2024) ressaltam que, para mitigar esses riscos, é essencial que os sistemas baseados em IA sejam transparentes, auditáveis e verificáveis, com mecanismos de controle contínuo para garantir que os processos automatizados não resultem em discriminação ou falhas na preservação de dados sensíveis.

Este debate sobre os riscos e benefícios da IA na preservação digital também está presente no estudo de caso do Bundesarchiv, que enfrentou desafios ao integrar IA em seus processos de digitalização e organização de dados sensíveis, equilibrando a automação com a necessidade de supervisão humana para garantir a proteção da privacidade.

### *4.3.1 Perspectivas Éticas Adicionais: Inteligência Artificial e Algoritmos*

Embora a Inteligência Artificial (IA) ofereça benefícios significativos na organização e acessibilidade de arquivos digitais, ela também apresenta desafios éticos que exigem uma reflexão mais aprofundada. O uso de algoritmos automatizados na organização, categorização e pesquisa de dados sensíveis levanta questões cruciais sobre a transparência e a imparcialidade dos processos. A falta de transparência nos algoritmos pode resultar em decisões não auditáveis, afetando a privacidade e a autenticidade dos dados sensíveis.

Além disso, o viés algorítmico é uma preocupação crescente. Sistemas de IA podem refletir e amplificar preconceitos presentes nos dados de treinamento, o que, em contextos históricos, pode levar à sub-representação ou categorização incorreta de certos dados ou memórias. Isso distorce a memória coletiva, e pode ser especialmente problemático ao lidar com dados sensíveis, como os relacionados a vítimas de violações de direitos humanos, onde a imprecisão pode resultar em revitimização.

Esses desafios éticos exigem que as instituições de preservação adotem práticas rigorosas de supervisão humana e transparência algorítmica. Os sistemas de IA precisam ser auditáveis, permitindo que decisões automatizadas possam ser revisadas quando necessário. Além disso, a anonimização de dados e o controle de acesso devem ser integrados aos protocolos de segurança, a fim de mitigar os riscos de exposição não autorizada de informações sensíveis, como ilustrado na experiência do Bundesarchiv. Durante a digitalização de arquivos sensíveis, a instituição enfrentou o desafio de garantir que os sistemas automatizados de IA respeitassem a privacidade e não distorcessem a memória histórica.

A discussão internacional sobre fairness e ética computacional demonstra que modelos de IA treinados sobre acervos incompletos ou historicamente enviesados tendem a reproduzir esses enviesamentos nos processos de classificação, descrição e recuperação da informação (Jo & Gebru, 2020). Tal risco torna essencial a adoção de modelos explicáveis, documentação técnica transparente (“model cards”), auditorias algorítmicas e supervisão humana constante (Mitchell et al., 2019).

## 4.4 Práticas Institucionais e Políticas de Gestão de Dados Sensíveis

A literatura demonstra que a preservação ética de dados sensíveis depende de políticas institucionais sólidas e suficientemente flexíveis para responder a novas exigências tecnológicas e normativas.

• **Segurança e anonimização**: controlo rigoroso sobre quem acede a dados sensíveis.

• **Flexibilidade e atualização contínua**: políticas dinâmicas capazes de integrar novas tecnologias e melhores práticas.

• **Conformidade legal e ética**: alinhamento com o RGPD e normas internacionais, garantindo proteção de direitos individuais sem comprometer a memória histórica.

O estudo de caso do Bundesarchiv confirma essa tendência ao evidenciar que, durante a pandemia de COVID-19, a eficácia dessas políticas foi determinante para assegurar acesso

remoto seguro, proteger dados sensíveis e manter a continuidade dos serviços, reforçando a importância de orientações institucionais claras, operacionais e adaptáveis.

Comparações internacionais mostram que instituições que adotam frameworks de governance baseados em risco e transparência apresentam maior consistência na proteção de dados sensíveis (Rinehart, 2022). Estes modelos são particularmente úteis em ecossistemas híbridos onde convivem sistemas analógicos, digitais e algoritmos de IA.

## 4.5 Estudo de Caso: O Impacto da Pandemia de COVID-19 na Digitalização e Preservação de Dados Sensíveis

O contexto pandémico acelerou processos de digitalização de acervos sensíveis, acentuando dificuldades relativas à proteção e preservação responsável dessas informações.

De seguida identificam-se os principais desafios enfrentados:

• **Aceleração da digitalização**: necessidade de tornar documentos acessíveis remotamente sem comprometer a privacidade.

• **Protocolos de segurança**: implementação de anonimização, criptografia e controlo de acesso.

• **Equilíbrio entre acesso e proteção**: garantir que a memória coletiva fosse acessível enquanto se preservavam os direitos de indivíduos e descendentes.

As soluções adotadas listam-se de seguida:

• Criação de portais digitais seguros, com restrição de acesso a investigadores qualificados.

• Anonimização e redação de dados sensíveis antes da disponibilização online.

• Formação dos funcionários para lidar com RGPD e novas tecnologias de preservação digital.

A relevância deste exemplo reside no facto de evidenciar o dilema central da investigação: preservar dados históricos sensíveis sem comprometer a privacidade, reforçando a necessidade de políticas institucionais robustas e mecanismos de supervisão ética.

Relatórios independentes (Kleinsteuber, 2023) destacam ainda desafios persistentes, como a integração entre sistemas legados e plataformas digitais mais recentes, bem como limitações na escalabilidade das ferramentas de anonimização quando aplicadas a grandes acervos históricos. Essas limitações ilustram a dificuldade estrutural em adaptar infraestruturas tradicionais às exigências tecnológicas atuais.

# 5 CONCLUSÕES

A análise realizada demonstra que a preservação de dados sensíveis em arquivos digitais envolve desafios éticos, legais e tecnológicos que exigem um equilíbrio rigoroso entre a proteção da privacidade e a salvaguarda da memória histórica. Este estudo oferece três contributos essenciais: (1) a identificação de diretrizes éticas fundamentais para o tratamento e acesso a dados sensíveis; (2) uma análise crítica das implicações do RGPD na preservação digital; e (3) a problematização do uso de Inteligência Artificial, evidenciando tanto o seu potencial como os riscos associados à falta de transparência e supervisão.

Todavia, o estudo apresenta limitações. O âmbito empírico restringe-se ao Bundesarchiv, o que condiciona a generalização dos resultados para outras realidades institucionais. Adicionalmente, a análise baseou-se predominantemente em documentação institucional e literatura especializada, não integrando entrevistas ou observação direta que poderiam aprofundar o entendimento dos processos analisados.

Para instituições com desafios semelhantes, os resultados sublinham a importância de adotar políticas claras de anonimização e controlo de acesso, atualizar infraestruturas tecnológicas e investir na formação especializada das equipas. Recomenda-se igualmente uma implementação cautelosa e eticamente supervisionada de sistemas de IA, assegurando que a modernização tecnológica não compromete os direitos individuais nem a integridade dos registos históricos.

Os resultados obtidos estão alinhados com tendências internacionais que defendem a adoção de infraestruturas arquivísticas capazes de incorporar auditorias algorítmicas, protocolos de explicabilidade e metodologias participativas, especialmente em acervos sensíveis relacionados com direitos humanos (Caswell & Cifor, 2019; Gollins, 2021).

## REFERÊNCIAS BIBLIOGRÁFICAS


Bantin, P. C., & Zhao, L. (2022). Privacy considerations in archival practice and research. International Journal of Information Management, 66, 102534. https://doi.org/10.1016/j.ijinfomgt.2022.102534

Borges, M. M. (2014). Políticas de preservação digital: Estruturação e características. Universidade de Coimbra. https://estudogeral.uc.pt/handle/10316/29995

Bundesarchiv. (2023). Memory of society: The federal archives at a glance. https://www.bundesarchiv.de/assets/bundesarchiv/en/Downloads/Memory_of_Society.pdf

Caswell, M., & Cifor, M. (2019). From human rights to feminist ethics: radical empathy in archives. Archivaria, 81.

Evans, J., Wallace, D., & Rawson, K. (2023). Critical Archival Studies and community-centered memory practices. Archival Science, 23(2).

Ferreira, N. M., & Martins, Ó. D. (2020). À procura do conceito de preservação digital. Cadernos BAD, 2(1), 1–15. https://publicacoes.bad.pt/revistas/index.php/cadernos/article/view/2051

Gilliland, A. (2022). Conceptualizing 21st-century archival pluralism. Archival Science, 22(4).

Gollins, T. (2021). Managing the risk of data re-identification in digital archives. Journal of Documentation, 77(5).

ISCTE-IUL (Gabinete de Proteção de Dados). (2022). Orientações aos investigadores sobre proteção de dados pessoais. https://www.iscteiul.pt/assets/files/2022/04/19/1650384518354_Orientac_o_es_aos_investigadores_sobre_protec_a_o_de_dados_pessoais.pdf

Jo, N. & Gebru, T. (2020). Lessons from archives: Strategies for collecting sociocultural data in ML. ACM FAccT.

Kleinsteuber, H. (2023). Digital preservation infrastructures in German federal archives. Bundesarchiv Research Notes.

Kosta, E., & Borgesius, F. J. Z. (2024). Learnings from EU-GDPR: Organisational challenges for data protection compliance. Computer Law & Security Review, 52, 105957. https://doi.org/10.1016/j.clsr.2024.105957

Mitchell, M. et al. (2019). Model Cards for Model Reporting. Google Research.

Rinehart, A. (2022). Risk governance in digital archives. Archival Futures Review, 14(1).

Rocher, L., Hendrickx, J., & de Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets. Nature Communications, 10.

Rodrigues, M. E. P., & Rodrigues, A. M. (2023). Preservação digital no contexto dos repositórios portugueses – estudo comparativo. https://www.researchgate.net/publication/374072632_Preservacao_digital_no_contexto_dos_repositorios_portugueses_-_estudo_comparativo

Schermer, B. (2021). Archival exceptions in the GDPR: Scope and limitations. European Data Protection Law Review, 7(3).

Shilton, K. (2023). Ethics of care in data governance. New Media & Society, 25(7).

Silva Júnior, L. P., & Borges, M. M. (2016). Preservação digital no Repositório Científico de Acesso Aberto de Portugal (RCAAP). https://www.researchgate.net/publication/305082548_Preservacao_Digital_no_Repositorio_Cientifico_de_Acesso_Aberto_de_Portugal

Travassos de A. Novaes, M. F. A. (2022). Preservação digital à luz do RGPD. Páginas a&b: Arquivos e Bibliotecas, 12(2), 55–68. https://ojs.letras.up.pt/index.php/paginasaeb/article/download/10411/9905/36554

Voigt, P., & von dem Bussche, A. (2021). The EU General Data Protection Regulation (GDPR). Springer.

Yakel, E., & Ketelaar, E. (2023). The impact of GDPR on the access to archives and the creation of finding aids. Archival Practice, 15(2), 145–162. https://doi.org/10.1017/aap.2023.5