DOI: 10.5433/1981-8920.2024v29n4p122

ARQUIVOLOGIA COMPUTACIONAL E O FUTURO DOS REGISTROS DIGITAIS

COMPUTATIONAL ARCHIVAL SCIENCE AND THE FUTURE OF DIGITAL RECORDS

Pedro Felipy Cunha da Silva[a]
Wagner Junqueira de Araújo[b]

RESUMO
Objetivo: Este estudo visa explorar como a arquivologia computacional tem sido abordada em publicações revisadas por pares, analisando a integração de tecnologias digitais na gestão de acervos arquivísticos digitais e identificando as principais tendências, ferramentas e desafios associados. **Metodologia:** Realizou-se uma revisão sistemática de literatura utilizando o protocolo PRISMA, com busca nas bases de dados *Web of Science, Scopus, Emerald, LISTA, Science Direct* e *Springer Link*. Foram selecionados 18 artigos após o processo de triagem, análise de títulos, resumos e textos completos. A análise bibliométrica foi realizada com o auxílio do *software VOSviewer*, que mapeou as coocorrências de palavras-chave e relações de coautoria. **Resultados:** A revisão evidenciou que a arquivologia computacional é uma área interdisciplinar em crescimento, caracterizada pelo uso de inteligência artificial, aprendizado de máquina, processamento de linguagem natural e mineração de dados para gerir e preservar registros digitais em grande escala. Identificou-se a crescente demanda por profissionais capacitados e a necessidade de integrar essas tecnologias aos currículos educacionais. Além disso, observou-se que Estados Unidos e Reino Unido lideram as publicações na área. **Conclusões:** A arquivologia computacional representa uma evolução necessária para enfrentar os desafios do ambiente digital, favorecendo a acessibilidade e o processamento eficiente de grandes volumes de dados. A integração entre arquivologia e computação permite novas possibilidades de exploração e uso de acervos digitais, ampliando o impacto da ciência arquivística na era digital. Pesquisas futuras podem focar no desenvolvimento de ferramentas específicas e no aprofundamento epistemológico do campo.

Descritores: Arquivologia computacional. Registro digital. Gestão da informação e do conhecimento. Tecnologias digitais da informação.

[a] Doutorando no Programa de Pós-graduação em Ciência da Informação (PPGCI) da Universidade Federal da Paraíba (UFPB). Arquivista na Universidade Federal da Paraíba (UFPB). Email: pedrofelipy@hotmail.com,

[b] Doutor em Ciência da Informação pela Universidade de Brasília (UNB). Professor Associado III no Departamento de Ciência da Informação da Universidade Federal da Paraíba (UFPB). Email: wagnerjunqueira.araujo@gmail.com

1 INTRODUÇÃO

Com o advento das inteligências artificiais e a crescente diversificação das ferramentas digitais para a produção de documentos e registros arquivísticos, administrar os requisitos de segurança e garantir a confiabilidade da informação contida nestes objetos torna-se um desafio cada vez maior. Por isso, pesquisadores de diferentes áreas do conhecimento (Feng *et al*., 2021) tem estudado e descrito um novo campo que está sendo denominado de C*omputational Archival Science* ou arquivologia computacional.

A arquivologia computacional é descrita como um campo de conhecimento emergente que reúne contribuições interdisciplinares da arquivologia e ciência da computação (Chen *et al*., 2024; Cushing; Osti, 2023; Hou; Kenderdine; Seydou, 2024; Yang *et al*., 2021). Tem sua origem na necessidade do desenvolvimento de ferramentas específicas para o tratamento adequado do crescente volume de material digital dos arquivos e instituições de memória.

Fenômenos como a transformação digital e a dataficação impulsionaram a demanda por sistemas capazes de criar, gerenciar, e prover acesso a grandes volumes de dados, preservando a integridade, acessibilidade e autenticidade das informações. É nesse contexto que a arquivologia computacional se torna relevante, como resposta aos desafios de organizar, descobrir, preservar no longo prazo e garantir o acesso de acervos digitais.

O desenvolvimento de ferramentas tecnológicas, como algoritmos de preservação digital e sistemas de mineração de dados em arquivos, favorece a automatização de processos de trabalho arquivísticos tradicionalmente manuais. Essas soluções podem oferecer maior eficiência na identificação, classificação, avaliação e recuperação de informações, além de possibilitar a segurança de dados sensíveis. Porém, a implementação dessas tecnologias enfrenta desafios, com destaque para a interoperabilidade entre sistemas, a padronização de formatos de arquivos e a necessidade de formar profissionais capacitados para lidar com esses recursos.

Por sua vez, a ciência da informação dedica-se a estudar os processos

de criação, armazenamento e acesso à informação ao longo do tempo, concentrando-se na estruturação e gestão do conhecimento em diversos contextos. Em paralelo, a arquivologia computacional se apropria desses fundamentos e os expande, aplicando ferramentas tecnológicas para desenvolver métodos computacionais que permitam não só a organização e preservação, mas também a descoberta, descrição e acessibilidade de grandes volumes de dados e informações em arquivos digitais.

Esse campo emergente busca garantir que a informação arquivística seja acessível tanto para usuários humanos quanto para sistemas automatizados, promovendo uma interação mais dinâmica e eficiente com acervos digitais complexos.

Dessa forma, este artigo busca responder à seguinte questão de pesquisa: como a arquivologia computacional tem sido abordada em artigos revisados por pares nas principais bases de dados científicas da área de ciência da informação?

Por meio de uma revisão sistemática de literatura, buscou-se verificar os conceitos apresentados, identificar as principais tecnologias aplicadas e compreender como a integração da computação na gestão de acervos arquivísticos relaciona-se as práticas clássicas da arquivologia.

2 PROCEDIMENTOS METODOLÓGICOS

Para a análise, foram escolhidas seis bases de dados relevantes para a ciência da informação: *Web of Science, Scopus, Emerald, Library Information Science & Technology Abstracts, Science Direct* e *Springer Link*.

A pesquisa foi realizada com o intermédio do portal de periódicos CAPES pelo termo de busca *"Computational Archival Science"* no mês de outubro de 2024. Os resultados podem ser observados na tabela a seguir:

Tabela 1 – Resultados da busca pelo termo *"computational archival science"*

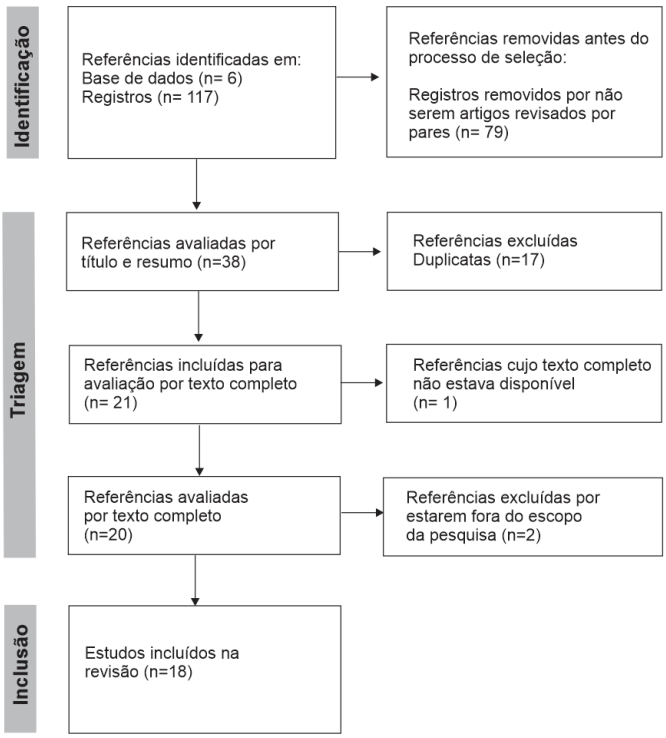
Base	*Web of science*	*Scopus*	*Emerald*	*LISTA (EBSCO)*	*Science Direct*	*Springer Link*

Tipo de documento						
Proceeding paper	**28**	**-**	**-**	**-**	**-**	**-**
Conference paper	**-**	**41**	**-**	**-**	**-**	**-**
Editorial material	**2**	**2**	**-**	**-**	**-**	**-**
Article	**6**	**10**	**8**	**7**	**1**	**6**
Book chapter	**-**	**1**	**-**	**-**	**-**	**2**
Book part	**-**	**-**	**2**	**-**	**-**	**-**
Encyclopedia	**-**	**-**	**-**	**-**	**1**	**-**
Total por base	**36**	**54**	**10**	**7**	**2**	**8**
Total acumulado	**36**	**90**	**100**	**107**	**109**	**117**

Fonte: dados da pesquisa (2024)

Para a revisão sistemática de literatura foi aplicado o protocolo PRISMA, em inglês *Preferred Reporting Itens for Systematic reviews and Meta-Analyses* (Moher *et al*., 2009). As fases de identificação, triagem e inclusão desta revisão podem ser visualizadas na figura a seguir:

Figura 1 – Aplicação do protocolo PRISMA de revisão sistemática de literatura



Fonte: dados da pesquisa (2024)

A etapa de identificação, constituída da pesquisa nas seis bases escolhidas, obteve 117 resultados de 7 tipos de documentos diferentes, conforme a figura 1. O primeiro critério estabelecido foi o de restringir a seleção a apenas um tipo de documento, que estivesse alinhado com o objetivo da pesquisa, de conhecer a produção científica revisada por pares sobre a arquivologia computacional. Para isso, os 79 documentos dos tipos *procedding paper, conference paper, editorial material, book chapter, book part* e *encyclopedia* foram excluídos; e os 38 documentos do tipo *article* foram selecionados para a etapa seguinte.

A etapa de triagem foi dividida em dois momentos: a avaliação dos títulos e dos resumos e a avaliação dos textos completos. Na avaliação dos títulos e dos resumos, foram excluídos 17 artigos duplicados, restando 21 selecionados para a leitura do texto completo. Destes, 1 artigo não estava disponível para acesso aberto, de modo que 20 documentos estavam disponíveis para leitura.

Na avaliação dos textos completos foram excluídos 2 artigos considerados como fora do escopo da pesquisa, restando 18 documentos que

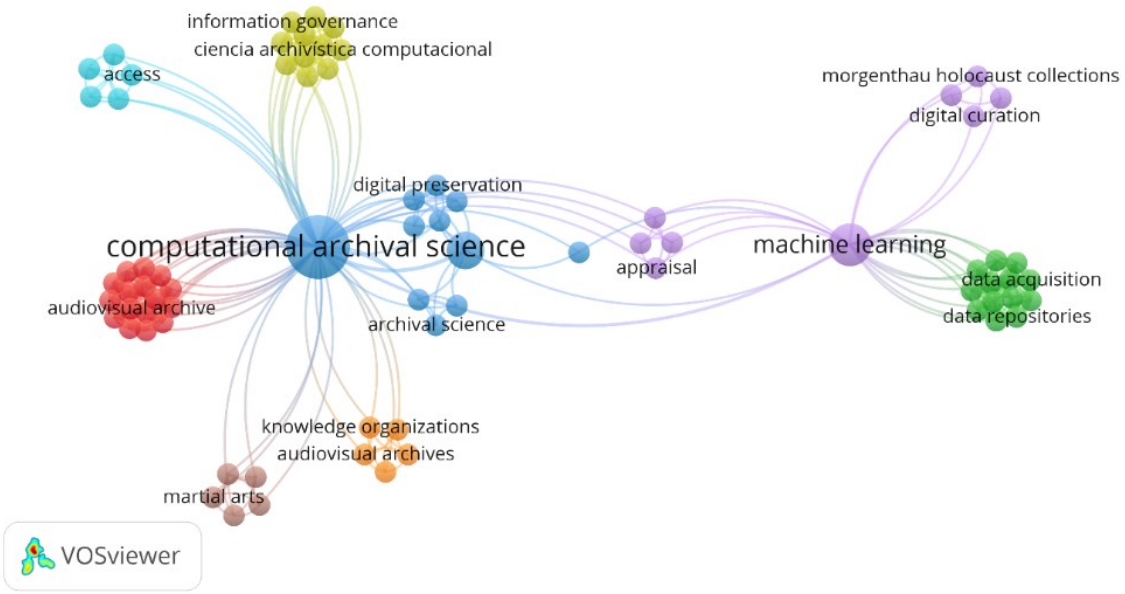
foram incluídos na revisão. A análise bibliométrica e a discussão do conteúdo dos artigos incluídos serão apresentadas na seção seguinte.

2.1 ANÁLISE BIBLIOMÉTRICA

Os periódicos escolhidos foram selecionados por serem considerados relevantes para a área de ciência da informação, conforme indicadores de citações. Nesse sentido, nesta seção serão analisados os dados bibliométricos dos artigos incluídos.

Após a conclusão das fases de identificação e triagem, os artigos foram organizados com o auxílio do *software Mendeley*. Na figura a seguir é possível visualizar uma nuvem produzida a partir das noventa e quatro palavras-chave dos dezoito artigos.

Figura 2 – Nuvem de palavras-chave



Fonte: dados da pesquisa (2024)

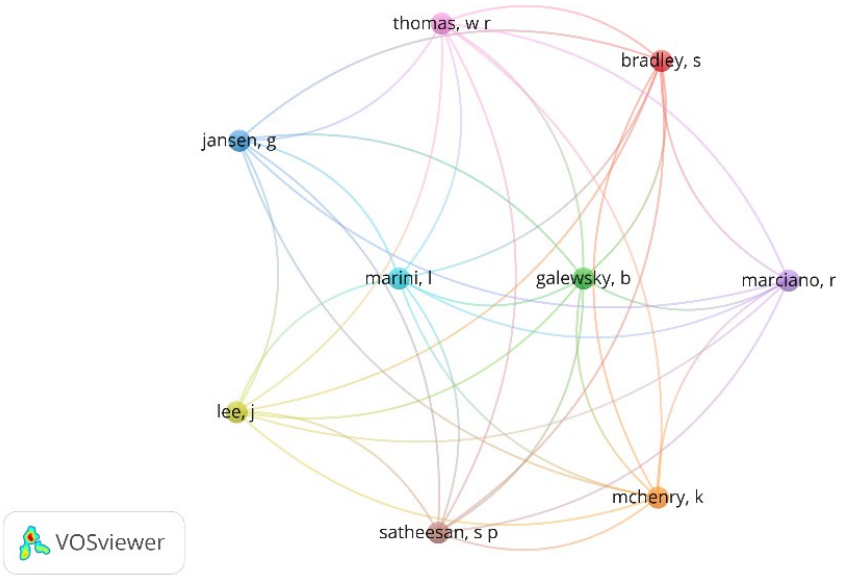
Os dados bibliométricos foram exportados do *Mendeley* no formato. RIS e, por meio do *software VOSviewer*, foi realizada a análise de coocorrência de palavras-chave, com o método de contagem *full conting* e o número mínimo de ocorrências de uma palavra-chave. Das 94 palavras-chave existentes, foram identificadas relações entre 75 termos, reunidos em 8 clusters, separados por

cores.

Limitações de visualização do *software VOSviewer* impedem que as palavras com menor relação sejam exibidas na figura, contudo, termos como patrimônio cultural imaterial, modelo de dataficação, ciência da computação, processamento de linguagem natural e arquivamento digital também se destacaram.

No mesmo *software* foi possível verificar a existência de relações de coautoria entre os artigos incluídos nesta revisão. O resultado pode ser visto na figura a seguir.

Figura 3 – Análise de coautoria



Fonte: dados da pesquisa, 2024

Dos 55 autores identificados, apenas 9 possuem relações de coautoria, são eles: Bradley, S; Galewsjy, B.; Jansen, G.; Lee, J.; Marciano, R.; Marini, L.; Mchenry, K.; Satheesam, S. P. e Thomas, W. R.

Foram identificadas as instituições nas quais os autores estão afiliados, e o que se observou é o que se pode verificar na tabela abaixo:

Tabela 2 – Afiliações institucionais dos autores

Tipo de instituição	Nome da instituição
Arquivos	• Arquivos Nacionais, Kew, Reino Unido

	• Arquivos Nacionais, Londres, Reino Unido • Arquivos Nacionais, Richmond, Reino Unido • Arquivos Universitários e Coleções Especiais, Universidade de Saskatchewan, Saskatoon, Canadá
Centros de Documentação	• Biblioteca e Museu Presidencial Franklin D. Roosevelt, Hyde Park, NY, EUA
Universidades	• Departamento de Ciência da Informação, Universidade do Norte do Texas • Departamento de Ciências da Computação e da Informação, Universidade de Northumbria, Reino Unido • Departamento de Preservação, Digitalização, Budapeste, Arquivos Nacionais da Hungria • Ecole Polytechnique Federale de Lausanne • Escola de Estudos de Informação e Comunicação, University College Dublin, Dublin, Irlanda • Escola de Gestão de Recursos de Informação, Universidade Renmin, Pequim, China • Escola de Informação, Universidade Estadual de San Jose, San Jose, Califórnia, EUA • Faculdade de Humanidades e Ciências Sociais, Universidade de Zagreb, Croácia • Laboratório Principal de Engenharia de Dados e Engenharia do Conhecimento, Ministério da Educação, Pequim • Universidade da África do Sul • Universidade de Ghent, Ghent, Bélgica • Universidade de Sussex, Brighton, Reino Unido • Universidade Eötvös Loránd, Budapeste, Hungria • University College London, Londres, Reino Unido • University of Guadalajara • University of Maryland, College Park, MD,

Fonte: dados da pesquisa (2024)

A maior parte dos artigos tem seus autores vinculados às universidades. Seguidos pelos autores vinculados às instituições arquivísticas, com predominância do Arquivo Nacional do Reino Unido. E apenas um autor é vinculado a um centro de documentação.

Outro dado identificado foi o número de citações. Com base nos registros bibliométricos obtidos junto as bases pesquisadas, foram listados os 5 artigos com o maior número de citações, são eles:

Tabela 3 – Citações por artigo

Título do artigo	Citações
Processamento de linguagem natural e aprendizado de máquina como conjuntos de ferramentas práticas para processamento de arquivos (tradução nossa), em inglês, *Natural language processing and machine learning as practical toolsets for archival processing.*	47
“Então, como podemos equilibrar todas essas necessidades? ”: como o conceito de tecnologia de IA impacta a experiência em arquivamento digital (tradução nossa), em inglês, *“So how do we balance all of these*	32

needs?": how the concept of AI technology impacts digital archival expertise.	
Arquivos nascidos digitalmente (tradução nossa), em inglês, *Born-digital archives.*	27
Inclusão de registros digitais nos currículos de arquivo e gestão de registros em um ambiente abrangente de e-learning aberto à distância (tradução nossa), em inglês *Inclusion of digital records in the archives and records management curricula in a comprehensive open distance e-learning environment.*	22
Retrospectiva e prospectiva: o cenário de pesquisa dos estudos arquivísticos (tradução nossa), em inglês, *Retrospect and prospect: the research landscape of archival studies.*	6

Fonte: dados da pesquisa (2024)

Os resultados desta análise oferecem uma base para futuras pesquisas ao mapear tendências de publicação, coautorias e palavras-chave mais frequentes sobre o tema. Esses dados podem ser utilizados para verificar como se expandem as principais redes de colaboração entre autores e instituições, o que facilita a compreensão da influência mútua entre pesquisadores e o desenvolvimento de abordagens conjuntas.

Além disso, a análise de palavras-chave ajuda a destacar temas centrais e emergentes, relacionados aos estudos da arquivologia computacional, orientando novas pesquisas para lacunas de conhecimento na área. Para estudos futuros, esses dados podem embasar análises longitudinais sobre o crescimento e a diversificação da arquivologia computacional, permitindo observar como os tópicos e tecnologias em uso evoluem ao longo do tempo. Isso contribui para que pesquisadores e profissionais compreendam melhor as necessidades de formação interdisciplinar e o impacto das tecnologias emergentes na prática arquivística. Na seção a seguir, será feita a discussão do conteúdo dos artigos inseridos nesta publicação.

2.2 DISCUSSÃO DO CONTEÚDO

O conceito de *Computational Archival Science*, aqui tratado como arquivologia computacional, foi formalmente proposto "em abril de 2016, no simpósio Descobrindo Novos Conhecimentos: Documentos de Arquivo na Era do *Big Data*, realizado na Universidade de Maryland (Yang *et al*., 2021, p. 5, tradução nossa) ".

Definida como uma "transformação da arquivística" e relacionada aos "arquivos digitais de grande escala" e aos "objetos de dados arquivísticos complexos" por Yang *et al*. (2022, p. 5, tradução nossa), a arquivologia computacional é proveniente de uma combinação de diferentes campos do conhecimento para responder aos novos desafios relacionados aos registros digitais arquivísticos.

Nos estudos de Hegedus (2020) o termo arquivologia computacional é definido como "uma disciplina que combina *big data* e direções e métodos de inteligências artificiais e aprendizado de máquina para criar os bancos de dados mais completos e relevantes possíveis a partir dos dados existentes" (Hegedus, 2020, p. 59, tradução nossa).

O *big data* impulsionou o crescimento do volume de dados arquivísticos e os desafios para os processar. "Desbloquear as informações ocultas em grandes estoques desses registros (Carter *et al*., 2022, p. 838, tradução nossa)" é apenas um dos desafios. De acordo com Yang (2021, p. 5, tradução nossa) "se não dermos atenção suficiente aos métodos informacionais, seremos afogados de informação".

Ao analisar a relação da arquivologia computacional com o conceito de big data, Stančić (2018) pontuou que o termo big data é:

> Geralmente descrito usando a definição 5Vs referindo-se a 1) volume, 2) velocidade, 3) variedade, 4) veracidade e 5) volatilidade. A característica de volume fala sobre o fato de que o arquivo digital está lidando com grande quantidade de registros ou grandes conjuntos de dados para os quais ele precisa de software especializado (arquivamento) que possa lidar com o volume de dados que está sendo ingerido. A característica de velocidade descreve a taxa na qual o arquivo está recebendo novos registros. A característica de variedade descreve o número de formatos de arquivo diferentes que o arquivo digital está preservando. Quanto mais formatos de arquivo, mais difícil será realizar a preservação a longo prazo. A característica de veracidade fala sobre precisão, confiança ou credibilidade nos dados, ou seja, os dados estão corretos. Finalmente, a característica de volatilidade descreve por quanto tempo os registros precisam ser mantidos após o qual se tornam irrelevantes e podem ser excluídos (adaptado de Stančić, 2018, p. 324, tradução nossa).

Segundo o mesmo autor, a arquivologia computacional pode ser definida como:

> [...] um campo interdisciplinar preocupado com a aplicação de métodos e recursos computacionais ao processamento, análise, armazenamento, preservação de longo prazo e acesso de registros/arquivos em larga escala, com o objetivo de melhorar a eficiência, produtividade e precisão em apoio à avaliação, arranjo e descrição, decisões de preservação e acesso, e envolvimento e realização de pesquisas com material de arquivo (Stančić, 2018, p. 325, tradução nossa).

Para Talboom e Bell (2022, p. 393, tradução nossa), foi o aumento dos registros nascidos digitalmente, e a interrupção no modo tradicional de aquisição, preservação e acesso dos arquivos e instituições de memória, que fez surgir novas áreas de estudo, dentre elas a arquivologia computacional.

A arquivologia computacional para Thomas *et al.* (2020, tradução nossa):

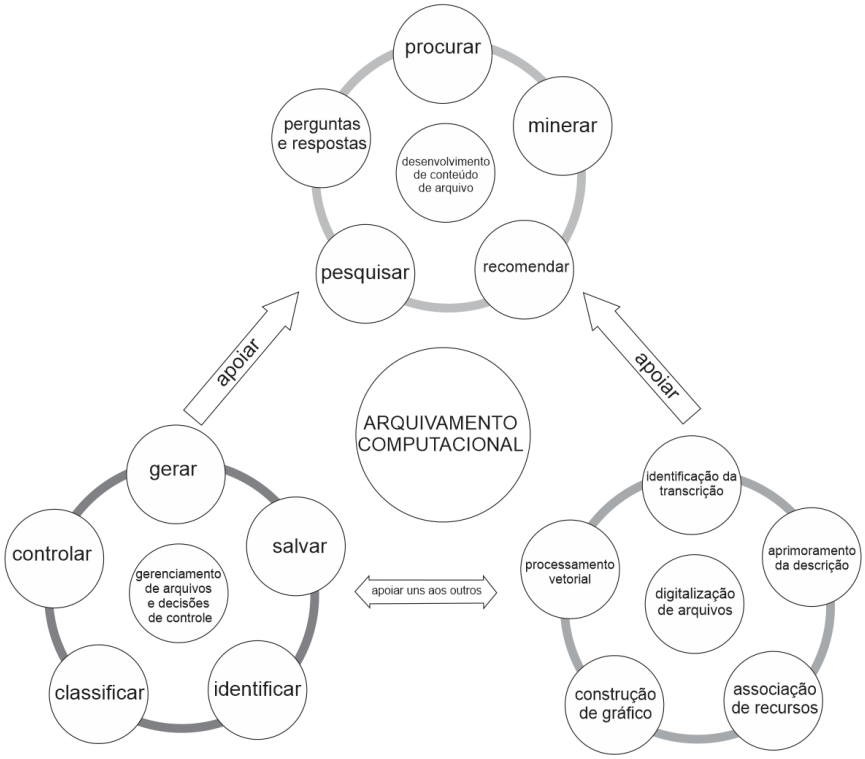
> Pode ajudar a teorizar novos relacionamentos entre comunidades e as instituições de memória que as atendem. [...] como levar a computação aos dados em um registro para construir descrições mais profundas do que seria possível de outra forma, bem como facilitar a recontextualização e a melhoria incremental na descrição do registro, e como podemos dimensionar o armazenamento horizontalmente conforme o número de imagens aumenta (Thomas et. al., 2020, p. 1, tradução nossa).

O trabalho de Yang *et al.* (2021, tradução nossa) dividiu as pesquisas sobre o assunto em antecedentes, evolução de conotação e características de desenvolvimento, sendo um dos que melhor estruturou uma proposta para a arquivologia computacional

As áreas de pesquisa defendidas por Yang *et al*. (2021) são distribuídas em três dimensões. Na dimensão 1, o objetivo é desenvolver conteúdo de arquivo por meio de recursos computacionais, como é o caso da mineração e das recomendações. Na dimensão 2, a proposta é gerir de forma estratégica os recursos informacionais constantes do acervo, por exemplo, através da criação inteligente e controlada de registros que atendam aos critérios arquivísticos já na sua concepção. E na dimensão 3, a ideia é materializar a dataficação dos acervos, isto é, converter os arquivos em coleções de dados arquivísticos, e tornar digitais os materiais de arquivo que ainda não são acessíveis por máquina, através de automatizações, transcrições dinâmicas e outros recursos computacionais.

Essa visão é apresentada na figura 4 a seguir:

Figura 4 – Principais áreas de pesquisa da arquivologia computacional



Fonte: Yang *et al.*, 2021, tradução nossa

De acordo com Chen *et al*. (2024, tradução nossa), com o surgimento e crescimento do princípio da "coleção como dados" no campo das humanidades digitais – que propõe "a integração de métodos computacionais na pesquisa em humanidades", foi quando a arquivologia computacional ganhou força, como "um campo transdisciplinar que integra teorias, métodos e recursos computacionais e arquivísticos" (Hedges *et al*., 2022 *apud* Chen *et al*., 2024, p. 2, tradução nossa).

Essa abordagem se assemelha ao que é proposto por Hou, Kenderdine e Seydou (2024, p. 2, tradução nossa) que parte do contexto dos desafios das instituições de patrimônio cultural para lidar com a escala dos registros digitais, alertando que "o campo emergente da arquivologia computacional sugere a integração do pensamento computacional com o arquivístico". De acordo com os autores, trata-se de uma:

> [...] abordagem promissora a integração da inteligência algorítmica (auto processamento e aprendizado de máquina) com modelos de descrição de conteúdo (tecnologia semântica e engenharia de dados vinculados) para dar suporte a um modo

> produtivo, eficaz a interpretativo de uso arquivístico (Hou; Kenderdine; Seydou, 2024, p. 2, tradução nossa).

O estudo de Cushing e Osti (2023, tradução nossa) complementa que a arquivologia computacional "Tenta combinar habilidades e conhecimento da arquivologia, ciência da informação e ciência da computação para criar um novo campo interdisciplinar" (Cushing; Osti, 2023, p. 15, tradução nossa). E a definem como:

> [...] um campo interdisciplinar preocupado com a aplicação de métodos e recursos computacionais ao processamento, análise, armazenamento, preservação de longo prazo e acesso de registros/arquivos em larga escala, com o objetivo de melhorar a eficiência, produtividade e precisão no suporte à avaliação, arranjo e descrição, decisões de preservação e acesso, e envolver e empreender pesquisas com material de arquivo (Payne, 2018, p. 2743 *apud* Cushing; Osti, 2023, p. 15, tradução nossa).

Outro conceito, sendo mais abrangente que os anteriores sobre a arquivologia computacional nesta revisão a define como:

> [...] um campo interdisciplinar que integra teorias, métodos e recursos de computação e arquivo, que pode não apenas apoiar a geração e preservação de documentos/arquivos autênticos e confiáveis, mas também resolver o processamento, análise, armazenamento e acesso a documentos/arquivos em grande escala, para melhorar a eficiência, produtividade e precisão, apoiar a manutenção relacionada à gestão, identificação, organização e descrição, preservação e acesso de documentos tomada de decisões, bem como participar e conduzir pesquisas usando materiais de arquivo (Yang *et al.*, 2021, p. 6, tradução nossa).

Contudo, apesar dos diversos conceitos identificados, na perspectiva empírica, os trabalhos analisados não apontam ferramentas ou características exclusivas da arquivologia computacional, mas indicam possibilidades tecnológicas que vêm sendo utilizadas e podem ser vinculadas a este campo do conhecimento emergente. Algumas dessas possibilidades foram agrupadas e são apresentadas tabela a seguir:

Tabela 4 – Tecnologias relacionadas à arquivologia computacional

Tecnologia	Fonte
Softwares que combinam computação do movimento com preservação de arquivos multimodais para preservação de coleções de patrimônio cultural imaterial	Hou; Kenderdine; Seydou (2024),

Modelos de descrição para arquivos audiovisuais	Yang (2024)
Processamento de linguagem natural	Chen *et al.* (2024) Yang (2023) Hutchinson (2020) Stančić (2018)
Análise de dados para compreensão de coleções de arquivos audiovisuais de história oral como ferramenta para preservação de memórias coletivas	Chen *et al.* (2024)
Visão computacional	Cushing; Osti (2023) Yang (2023) Bell e Bunn (2022)
Modelos para recuperação da informação de vídeos a partir de textos	Yang (2023)
Blockchain	Cushing; Osti (2023) Yang *et al.* (2021)
Aprendizagem de máquina	Cushing; Osti (2023) Carter *et al.* (2022) Hutchinson (2020)
Blocos de notas computacionais como ferramentas para acesso colaborativo de material nato digital	Talboom e Bell (2022) Carter *et al.* (2022)
Inteligência artificial	Carter *et al.* (2022) Yang *et al.* (2021)
Mineração de texto e dados	Yang *et al.* (2021) Hutchinson (2020) Stančić, 2018
Recomendação de conteúdo e criação de FAQs	Yang *et al.* (2021)
Redes neurais, aprendizado profundo e gráficos de conhecimento	Yanh *et al. (2021)*
Computação em nuvem	Feng *et al*. (2021) Stančić (2018)
Datificação de arquivos	Yang (2024) Yanh *et al. (2021)*
Brown dog e DRAS-TIC	Thomas *et al.* (2019)
Analise de rede	Stančić (2018) Bell (2020)

Fonte: dados da pesquisa (2024)

São diversas as tecnologias encontradas por esta revisão para tratamento de dados estruturados e não estruturados em grandes volumes, conforme aponta Thomas *et al.* (2020), ao afirmar que "[...] a transdisciplinaridade emergente da arquivologia computacional vincula estruturas como *Brown Dog* e software de repositório como o *Digital Repository At Scale To Invite Computation* (DRAS-TIC) para produzir uma compreensão do trabalho com coleções digitais

em escala [...]" (Thomas *et al*., 2020, p. 1, tradução nossa).

O *Brown Dog* é um serviço de transformação de dados para compreensão de dados não estruturados por meio de autocuradoria e indexação. Ele foi desenvolvido para ajudar a dar sentido a dados que exigem acesso a um conjunto diversificado de software para processamento. É um software que tenta alavancar todas as demais ferramentas de software disponíveis para curadoria automática (Padhy *et al*., 2015, tradução nossa). Também "Auxilia no compartilhamento de ferramentas de transformação de dados, pois as ferramentas de software são expostas como serviços e são disponibilizadas para atender as necessidades de transformação de seus usuários". (Thomas *et al.*, 2020, p. 3, tradução nossa).

Na perspectiva teórica, a arquivologia computacional se concentra "na natureza do registro e na aplicação de ferramentas computacionais a esse registro, em vez de focar no processamento de dados, o que é típico da ciência da computação" (Marciano *et al*., 2018 *apud* Cushing; Osti, 2023, tradução nossa).

Segundo McLeod e Lomas (2023) a arquivologia computacional desenvolve-se junto com as humanidades digitais "[...] da necessidade de conjuntos de habilidades específicas para entender as evidências digitais, registros e considerações arquivísticas" (McLeod; Lomas, 2023, p. 419, tradução nossa). Para elas, os registros digitais são definidos pela política e não pela teoria arquivística, porém, esclarecem que:

> No mundo digital dinâmico, líquido e flexível, isso levanta a questão de se devemos nos concentrar em gerenciar os componentes do DNA (do registro) para que a "registricidade" possa ser julgada pelo usuário para qualquer propósito, em qualquer contexto, tempo e espaço, em paralelo ao hiperobjeto "autoautenticador" de Smit (2017). Se for assim, devemos reconhecer os componentes do registro (seu DNA) e nos concentrar em permitir o gerenciamento desses bits granulares ao longo do tempo, aproveitando o conhecimento e as habilidades de muitas disciplinas, reconhecendo na arquivologia computacional a necessidade de atrair outras (McLeod; Lomas, 2023, p. 419, tradução nossa).

Ou seja, embora esta área ainda esteja em processo de desenvolvimento em conjunto com outras, há uma vinculação evidente entre a arquivologia computacional e a compreensão dos registros digitais arquivísticos.

Especificamente quanto a natureza destes registros, seus componentes e a necessidade de construir ou incorporar ferramentas computacionais que favoreçam o processamento, gerenciamento e criação de sentido para os usos e usuários existentes e potenciais.

A arquivologia computacional também é abordada na perspectiva da formação e evolução profissional nos artigos dessa revisão. Para Ries (2019) os arquivistas precisam se tornar especialistas em registro nato digitais. Yang *et al*. (2021) afirmam a necessidade de formar novos talentos, que devem se concentrar em quatro práticas, a saber, "prática de dados, prática de modelagem e simulação, prática de resolução de problemas baseada em computação e prática de pensamento sistêmico" (Weintrop, 2016 *apud* Yang *et al*., 2021, p. 9, tradução nossa).

Mas para isso, segundo Hutchinson (2020), será necessário que arquivistas atuem de forma interdisciplinar com cientistas da computação, compreendendo as metodologias de mineração de dados e análise computacional.

A inclusão da arquivologia computacional no currículo dos cursos de pós-graduação em gestão de documentos e arquivos é defendida por Ngoepe, Jacobs e Mojapelo (2024). De acordo com os autores, o conhecimento dessa ciência permitirá que os estudantes e profissionais "Adquiram habilidades na aplicação de ferramentas, métodos e recursos tecnológicos relativos ao processamento, análise, armazenamento, preservação de longo prazo e acesso de registros digitais" (Ngoepe; Jacobs; Mojapelo, 2024, p. 1, tradução nossa).

Tal proposição encontra amparo no trabalho de Talboom e Bell (2022, p. 410, tradução nossa), que estudou "blocos de notas em *phyton* hospedados em nuvem para fornecer uma maneira diferente de acessar um arquivo digital que vai além da interface de pesquisa de palavras-chave". Para os autores:

> [...] o formato permite que uma explicação narrativa seja incluída junto com o código, permitindo que os usuários acompanhem em seu próprio ritmo e em seu próprio tempo. Eles também são uma ótima maneira de defender o acesso computacional no ambiente de arquivamento (Talboom; Bell, 2022, p. 410, tradução nossa).

Carter *et al*. (2022) e Talboom e Bell (2022) tratam sobre essa nova

possibilidade tecnológica para a promover acesso aos grandes volumes de registros digitais existentes nos arquivos, especialmente em instituições de memória. Os blocos de notas computacionais, em Phyton ou Jupyter, são definidos como "ferramenta da web interativa, gratuita e de código aberto [...], que os pesquisadores podem usar para combinar código de software, saída computacional, texto explicativo e recursos multimídia em um único documento" (Perkel, 2018 *apud* Carter *et al,* 2022, p. 850).

Isto é, a arquivologia computacional pode ser vista como um alicerce para a construção de "ferramentas que auxiliem cognitivamente os arquivistas na tarefa de análise ou que auxilie os pesquisadores nas tarefas analíticas associadas a suas pesquisas" (Bell; Bunn, 2022, p. 961, tradução nossa).

García-Morales (2021), assevera que no contexto da gestão de registros e da governança de informações, a eficiência das organizações passa pela construção de estratégias que viabilizem o não armazenamento de informações "ROT" (redundantes, obsoletas e transitórias) e, para isso, há o conceito de "retenção de informações". Segundo a autora, a arquivologia computacional é um caminho para a comunicação interprofissional requerida para essa finalidade, por meio da integração das técnicas arquivísticas às novas soluções advindas da tecnologia da informação. Esse mesmo tema é tratado por Stančić (2018) que define a retenção e disposição de informações como uma preocupação a ser tratada com base na arquivologia computacional.

Além da "governança de informações" e da "retenção de informações", descritas anteriormente, a curadoria digital também foi identificada nos artigos. Bell e Bunn (2022) relatam que a curadoria digital obteve, até o momento, um maior êxito e aceitação dentro da comunidade arquivística profissional.

Para resumir e dar visibilidade aos principais elementos abordados nos artigos analisados, foi preparada a figura a seguir:

Figura 5 – As relações estabelecidas pela arquivologia computacional

Fonte: dados da pesquisa (2024)

Os conteúdos presentes nos trabalhos analisados versam sobre os fenômenos e conceitos que deram origem aos estudos sobre arquivologia computacional, as possibilidades tecnológicas vinculadas a essa área do conhecimento, as competências necessárias ao profissional que deseja atuar com essas possibilidades, e os recursos já existentes para aplicação desse conhecimento para o tratamento dos arquivos digitais de grande escala.

3 CONSIDERAÇÕES E REFLEXÕES

Este trabalho teve como foco identificar e descrever como a arquivologia computacional é retratada nos artigos revisados por pares na área de conhecimento da ciência da informação. Obteve-se como resultado uma série de conceitos e aplicações sobre o tema, que iniciam, mas não esgotam as possibilidades de estudo sobre o assunto.

Os artigos analisados demonstram que a base da arquivologia computacional está na integração de conhecimentos da arquivologia com a ciência da computação, com foco no processamento, análise, armazenamento, preservação de longo prazo e principalmente no acesso aos arquivos digitais de

grande escala, ou ainda, nos conjuntos de dados presentes nos acervos de arquivos e instituições de memória.

Os estudos analisados apontam que para se beneficiar dos recursos computacionais aplicados aos acervos arquivísticos, os materiais de arquivo precisam ser digitalizados ou criados digitalmente seguindo critérios específicos. Posteriormente precisam ser descritos, com uso de metadados que os enriqueçam na perspectiva semântica, e estarem disponíveis para acesso aberto. Na sequência, há a necessidade de criação de uma infraestrutura de tecnologia e informações, que exige a participação cooperativa dos profissionais que entendem do acervo com aqueles que entendem das ferramentas de análise computacional, e é essa a contribuição da arquivologia computacional, integrar esses saberes e práticas em prol do acesso e do desenvolvimento da pesquisa científica em diversas áreas.

A arquivologia computacional favorece a dataficação dos materiais de arquivo, isto é, a transformação dos documentos digitalizados em dados e a conversão dos registros nascidos digitalmente em grandes conjuntos de dados arquivísticos. Desse modo, é possível criar ambientes digitais onde grandes volumes de dados e informações presentes nos arquivos poderão ser explorados e analisados com intermédio dos recursos computacionais, tecnologias colaborativas e novas formas de interação homem-máquina.

Tecnologias de processamento de linguagem natural, aprendizagem de máquina, modelagem, inteligências artificiais, visão computacional, mineração, blocos de notas computacionais são alguns dos recursos que se destacam.

Com foco na natureza dos registros digitais e não apenas no processamento dos dados, a arquivologia computacional pode contribuir no enriquecimento semântico desses registros, identificando informações importantes para diferentes perfis de usuários, e possibilitando o cruzamento de acervos com outros conjuntos de dados abertos. É uma revolução no acesso aos documentos de arquivo. Esse cenário irá permitir que os arquivos possam ser reutilizados para objetivos diferentes daqueles para os quais foram criados, ou seja, ampliar o escopo de possibilidades de uso dos arquivos.

Este estudo verificou que a arquivologia computacional é analisada em

diferentes perspectivas: como disciplina, área de pesquisa, aplicação de métodos da computação na arquivologia, transformação da arquivística e de forma preponderante como um novo campo de conhecimento interdisciplinar.

Com base na revisão podemos inferir que a **Arquivologia Computacional** é um campo interdisciplinar emergente que combina princípios da arquivologia e da ciência da computação para atender às demandas da gestão de acervos arquivísticos digitais em larga escala. Esse campo envolve o desenvolvimento e a aplicação de ferramentas tecnológicas que viabilizam a organização, preservação, descoberta e acesso a grandes volumes de dados arquivísticos, sendo relevante tanto para usuários humanos quanto para sistemas automatizados, os ditos usuários não-humanos.

1. Aplicações Práticas: Na prática, a arquivologia computacional implementa tecnologias como inteligência artificial, visão computacional, aprendizado de máquina e mineração de dados para otimizar a classificação, recuperação e preservação de registros digitais. Ela permite a criação de algoritmos de descrição automática, a análise de redes de documentos e a exploração de grandes volumes de informações arquivísticas de maneira rápida e eficiente.

2. Implicações nos Arquivos: A adoção da arquivologia computacional altera profundamente as operações dos arquivos, promovendo a transformação de documentos físicos e digitais em "coleções de dados" que podem ser analisadas e recontextualizadas. Isso amplia o papel dos arquivos como fontes de informação dinâmicas e acessíveis, favorecendo o desenvolvimento de novos métodos de preservação e tornando os acervos mais responsivos às necessidades da sociedade e da pesquisa científica.

3. Necessidade de Formação: A atuação efetiva em arquivologia computacional demanda que os arquivistas possuam habilidades específicas em tecnologias de informação, incluindo conhecimentos em análise de dados, modelagem computacional e gestão de metadados. Isso implica uma formação interdisciplinar que permita a esses profissionais não só gerir os acervos, mas também colaborar com especialistas em ciência da computação para desenvolver soluções tecnológicas alinhadas às práticas e aos princípios

arquivísticos.

A razão de ser da arquivologia é dar acesso aos documentos arquivísticos. Numa sociedade marcada pela produção e consumo de documentos de forma preponderante no ambiente digital, em que a computação é essencial para criar, armazenar e acessar o acervo, talvez a arquivologia computacional seja apenas uma adaptação ou evolução necessária para que a arquivologia continue podendo desempenhar o seu papel. Se esse rótulo favorece a valorização do conhecimento da arquivologia e a encontrabilidade de registros arquivísticos, vemos de forma positiva.

REFERÊNCIAS

SEKEFF, Gisela. O emprego dos sonhos. **Domingo**, Rio de Janeiro, ano 26, n. 1344, p. 30-36, 3 fev. 2002.

BELL, Mark. From tree to network: reordering an archival catalogue. **Records Management Journal**, [S. l.], v. 30, n. 3, p. 379–394, 1 jan. 2020. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/RMJ-09-2019-0051/full/html. Acesso em: 29 out. 2024.

BELL, Mark.; BUNN, Jenny. Dark archives or a dark age for reasoning over archives? **AI and Society**, [S. l.], v. 37, n. 3, p. 959–966, 1 set. 2022. Disponível em: https://doi.org/10.1007/s00146-021-01365-z. Acesso em: 29 out. 2024.

CARTER, Kirsten Strigel; GONDEK, Abby; UNDERWOOD, William; RANDBY, Teddy; MARCIANO, Richard. Using AI and ML to optimize information discovery in under-utilized, Holocaust-related records. **AI and Society**, [S. l.], v. 37, n. 3, p. 837–858, 1 set. 2022. Disponível em: https://doi.org/10.1007/s00146-021-01368-w. Acesso em: 29 out. 2024.

CHEN, Haihua; KIM, Jeonghyun (Annie); CHEN, Jiangping; SAKATA, Aisa. Demystifying oral history with natural language processing and data analytics: a case study of the Densho digital collection. **The Electronic Library**, [S. l.], v. 42, n. 4, p. 643–663, 1 jan. 2024. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/EL-12-2023-0303/full/html. Acesso em: 29 out. 2024.

CUSHING, Amber. L.; OSTI, Giulia. So how do we balance all of these needs?: how the concept of AI technology impacts digital archival expertise. **Journal of Documentation**, [S. l.], v. 79, n. 7, p. 12–29, 1 jan. 2023. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/JD-08-2022-0170/full/html. Acesso em: 29 out. 2024.

FENG, Huiling; LIAN, Zhiying; PAN, Weimei; QU, Chunmei; ZHOU, Wenhong; WANG, Ning; LI, Mengqiu. Retrospect and prospect: the research landscape of archival studies. **Archival Science**, [S. l.], v. 21, n. 4, p. 391–411, 1 dez. 2021. Disponível em: https://doi.org/10.1007/s10502-021-09364-1. Acesso em: 29 out. 2024.

GARCÍA-MORALES, Elisa. Retención de la información en la empresa: cuestión de sostenibilidad. **Anuario Think EPI**, [S. l.], v. 15, p. 1–5, jan. 2021. Disponível em: https://doi.org/10.3145/thinkepi.2021.e15f03. Acesso em: 29 out. 2024.

HEGEDUS, István. How artificial intelligence and machine learning can help rethink archives? **Atlanti+**, [S. l.], v. 30, n. 2, p. 57–64, 2020. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85147496843&origin=inward&txGid=deb13850670fcc867791f0cfbb531591. Acesso em: 29 out. 2024.

HOU, Yumeng; SEYDOU, Fadel Mamar.; KENDERDINE, Sara. Unlocking a multimodal archive of Southern Chinese martial arts through embodied cues. **Journal of Documentation**, [S. l.] v. 80, n. 5, p. 1148–1166, 2024. Disponível em: https://doi.org/10.1108/JD-01-2022-0027. Acesso em: 29 out. 2024.

HUTCHINSON, Tim. Natural language processing and machine learning as practical toolsets for archival processing. **Records Management Journal**, [S. l.], v. 30, n. 2, p. 155–174, 2020. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85085025120&doi=10.1108%2fRMJ-09-2019-0055&origin=inward&txGid=2edea74e6edeb6e24a73ba66c99cd138. Acesso em: 29 out. 2024.

MOHER, David; LIBERATI, Alessandro; TETZLAFF, Jennifer; ALTMAN, Douglas G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **Plos Medicine**, [S. l.], v. 6, n. 7, p. 1-6, 2009. DOI: https://doi.org/10.1371/journal.pmed.1000097. Disponível em: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097#s5. Acesso em: 15 out. 2024.

MCLEOD, Julie; LOMAS, Elizabeth. Record DNA: reconceptualising digital records as the future evidence base. **Archival Science**, [S. l.], v. 23, n. 3, p. 411–446, 1 set. 2023. Disponível em: https://doi.org/10.1007/s10502-023-09414-w. Acesso em: 29 out. 2024.

NGOEPE, Mpho.; JACOBS, Lorette.; MOJAPELO, Makutla. Inclusion of digital records in the archives and records management curricula in a comprehensive open distance e-learning environment. **Information Development**, [S. l.], v. 40, n. 2, p. 190–201, 24 jun. 2024. Disponível em: https://journals.sagepub.com/doi/10.1177/02666669221081812. Acesso em 29 out. 2024.

PADHY, Smruti; JANSEN, Greg; ALAMEDA, Jay; BLACK, Edgar; DIESENDRUCK, Liana; DIETZE, Mike. Brow Dog: leveraging everything towards autocuration. *In:* 2015 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2015, Santa Clara, CA, USA. **Anais** [...]. Santa Clara, CA, USA: IEEE, 2015 (pp. 493-500). DOI: 10.1109/BigData.2015.7363791. Disponível em: https://ieeexplore.ieee.org/document/7363791. Acesso em: 29 out. 2024.

RIES, Thorsten.; PALKÓ, Gábor. Born-digital archives. **International Journal of Digital Humanities**, [S. l.] v. 1, n. 1, p. 1–11, abr. 2019. Disponível em: https://doi.org/10.1007/s42803-019-00011-x. Acesso em: 29 out. 2024.

STANČIĆ, Hrvoje. Computational archival science. **Moderna Arhivistika**, [S. l.], v. 2018, n. 2, p. 323–330, 2018. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85107473833&origin=inward&txGid=dbb450a07bffc404c46e5c9493c5a1e2. Acesso em: 29 out. 2024.

TALBOOM, Leontien.; BELL, Mark. Keeping it under lock and keywords: exploring new ways to open up the web archives with notebooks. **Archival Science**, [S. l.], v. 22, n. 3, p. 393–415, 2022. Disponível em: https://doi.org/10.1007/s10502-022-09391-6. Acesso em 29 out. 2024.

THOMAS, Will R.; GALEWSKY, Benjamin; SATHEESAN, Sandeep Puthanveetil; JANSEN, Gregory; MARCIANO, Richard; BRADLEY, Shannon; LEE, Jhong; MARINI, Luigi; MCHENRY, Kenton. Petabytes in Practice: Working with Collections as Data at Scale. **Data and Information Management**, [S. l.], v. 3, n. 1, p. 18–25, 2019. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85144411248&doi=10.2478%2fdim-2019-0004&origin=inward&txGid=8fc203ce25375a1edfd54005df5362fd. Acesso em: 29 out. 2024.

YANG, Jianliang; YUENAN, Liu; SIHUAN, He; QI, Tianjiao. Computational Archival Science: The New Development of Archival Science. **Documentation, Information and Knowledge**, [S. l.], v. 38, n. 3, p. 4–13, 2021. Disponível em: http://dik.whu.edu.cn/jwk3/tsqbzs/EN/10.13366/j.dik.2021.03.004. Acesso em: 29 out. 2024.

YANG, Yuchen. Write What You Want: Applying Text-to-Video Retrieval to Audiovisual Archives. **Journal on Computing and Cultural Heritage**,[S. l.], v. 16, n. 4, 2023. Disponível em: https://www.scopus.com/record/display.uri?eid=2-s2.0-85184770102&doi=10.1145%2f3627167&origin=inward&txGid=a0c31b0008b0a7208bdc2cccbc63bee3. Acesso em: 29 out. 2024.

YANG, Yuchen. Datafication of audiovisual archives: from practice mapping to a thinking model. **Journal of Documentation**, [S. l.], v. 80, n. 5, p. 1119–1132, 3 set. 2024. Disponível em:

https://www.emerald.com/insight/content/doi/10.1108/JD-04-2022-0093/full/html. Acesso em 29 out. 2024.

COMPUTATIONAL ARCHIVAL SCIENCE AND THE FUTURE OF DIGITAL RECORDS

ABSTRACT
Objective: This study aims to explore how computational archival science has been addressed in peer-reviewed publications, analyzing the integration of digital technologies in the management of digital archival collections and identifying the main trends, tools, and challenges associated. **Methodology:** A systematic literature review was conducted using the PRISMA protocol, with searches in the Web of Science, Scopus, Emerald, LISTA, Science Direct, and Springer Link databases. After a screening process, including title, abstract, and full-text analysis, 18 articles were selected. Bibliometric analysis were performed using *VOSviewer* software, mapping keyword co-occurrences and co-authorship relationships. **Results:** The review highlighted that computational archival science is a growing interdisciplinary field characterized by the use of artificial intelligence, machine learning, natural language processing, and data mining to manage and preserve digital records on a large scale. It identified the increasing demand for trained professionals and the need to integrate these technologies into educational curricula. Additionally, it was observed that the United States and the United Kingdom lead publications in this area. **Conclusions:** Computational archival science represents a necessary evolution to address the challenges of the digital environment, promoting accessibility and efficient processing of large volumes of data. The integration between archival science and computing enables new possibilities for exploring and using digital collections, expanding the impact of archival science in the digital age. Future research could focus on developing specific tools and further epistemological depth within the field.

Descriptors: Computational archival science. Digital record. Information and Knowledge Management. Digital Information Technologies.

LA ARCHIVÍSTICA COMPUTACIONAL Y EL FUTURO DE LOS DOCUMENTOS DE ARCHIVO DIGITALES

RESUMEN
Objetivo: Este estudio tiene como objetivo explorar cómo se ha abordado la archivología computacional en publicaciones revisadas por pares, analizando la integración de tecnologías digitales en la gestión de colecciones archivísticas digitales e identificando las principales tendencias, herramientas y desafíos asociados. **Metodología:** Se realizó una revisión sistemática de la literatura utilizando el protocolo PRISMA, con búsqueda en las bases de datos Web of Science, Scopus, Emerald, LISTA, Science Direct y Springer Link. Después del proceso de selección, que incluyó el análisis de títulos, resúmenes y textos completos, se seleccionaron 18 artículos. El análisis bibliométrico y de contenido se realizó con el apoyo del software VOSviewer, que mapeó las coocurrencias de palabras clave y las relaciones de coautoría. **Resultados:** La revisión destacó que la archivología computacional es un campo

interdisciplinario en crecimiento, caracterizado por el uso de inteligencia artificial, aprendizaje automático, procesamiento de lenguaje natural y minería de datos para gestionar y preservar registros digitales a gran escala. Se identificó una demanda creciente de profesionales capacitados y la necesidad de integrar estas tecnologías en los planes de estudio educativos. Además, se observó que Estados Unidos y Reino Unido lideran las publicaciones en el área. **Conclusiones:** La archivología computacional representa una evolución necesaria para enfrentar los desafíos del entorno digital, promoviendo la accesibilidad y el procesamiento eficiente de grandes volúmenes de datos. La integración entre archivología e informática permite nuevas posibilidades para la exploración y el uso de colecciones digitales, ampliando el impacto de la ciencia archivística en la era digital. Las investigaciones futuras podrían centrarse en el desarrollo de herramientas específicas y en el mayor profundización epistemológica del campo.

Descriptores: Archivología computacional. Registro digital. Gestión de la Información y el Conocimiento. Tecnologías Digitales de la Información.

Recebido em: 30.10.2024
Aceito em: 21.11.2024

.