

# PRESERVAÇÃO DIGITAL: PANORAMA DA PRODUÇÃO CIENTÍFICA NA BASE DE DADOS WEB OF SCIENCE (1992-2022)

DIGITAL PRESERVATION: AN OVERVIEW OF SCIENTIFIC PRODUCTION IN THE WEB OF SCIENCE DATABASE (1992-2022)



**Laura Vilela Rodrigues Rezende**

Universidade Federal de Goiás  
E-mail: [lauravil.rr@gmail.com](mailto:lauravil.rr@gmail.com)  
Goiânia – GO – Brasil



**Pablo Fernandes Castanheira**

Universidade Federal de Goiás  
E-mail: [pablofernandescastanheira@gmail.com](mailto:pablofernandescastanheira@gmail.com)  
Goiânia – GO – Brasil



**Rômulo Arantes Alves**

Universidade Federal de Goiás  
E-mail: [romuloarantes20@gmail.com](mailto:romuloarantes20@gmail.com)  
Goiânia – GO – Brasil



**Sonia Aguiar Cruz-Riascos**

Universidade Federal de Pernambuco  
E-mail: [lauravil.rr@gmail.com](mailto:lauravil.rr@gmail.com)  
Recife – PE – Brasil



## Resumo

**Introdução:** Trata-se de um panorama da produção científica sobre o tema Preservação Digital dada sua relevância na preservação do patrimônio científico, histórico e cultural da humanidade, registrado e armazenado em formato digital. **Objetivo:** O objetivo deste trabalho é identificar detalhadamente a produção científica sobre Preservação Digital a partir de artigos científicos publicados em periódicos indexados na base de dados Web of Science (WOS) no período de 1992 a 2022. **Metodologia:** Abordagem quantitativa, de caráter descritivo. O levantamento bibliográfico foi escolhido como técnica de pesquisa para realizar análise quantitativa baseada na elaboração de indicadores que proporcionam um melhor delineamento da produção científica sobre o tema. **Resultados:** A partir da coleta dos dados foram analisados 341 artigos que tratam diretamente do tema. Por meio da análise e interpretação dos dados foi possível verificar o número de artigos publicados ao longo dos anos, quantidade de artigos publicados segundo o idioma, país de origem dos pesquisadores, os periódicos com maior número de publicações, artigos por área do conhecimento, os autores mais produtivos na temática, o número de autores por artigo, os artigos mais citados e a frequência de palavras-chave. **Conclusão:** Os indicadores elaborados a partir deste estudo contribuem para aprimorar o conhecimento da comunidade científica sobre o tema: Preservação Digital. Acrescenta-se a isso o potencial de abrir caminhos para novas pesquisas, uma vez que um levantamento bem delimitado sobre a produção bibliográfica de uma temática específica facilita demasiadamente o trabalho dos pesquisadores no desenvolvimento de novos estudos com abordagens inovadoras.

**Palavras-chave:** Preservação digital. Objeto digital. Base de dados bibliográfica. Web of Science. Publicação científica.

## Abstract

**Introduction:** This is an overview of scientific production on the topic of Digital Preservation given its relevance in the preservation of the scientific, historical and cultural heritage of humanity, recorded and stored in digital format. **Objective:** The objective of this work is to analyze in detail the scientific production on Digital Preservation from scientific articles published in journals indexed in the Web of Science (WOS) database between the period from 1992 to 2022. **Methodology:** The bibliographical research was chosen as a data collection technique for performing quantitative analysis that is based on the development of indicators that provide a better outline of the scientific production on the subject. **Results:** From the collection of data, 341 articles were analyzed that deal directly with the topic. Through the analysis and interpretation of the data, it was possible to verify the number of articles published over the years, the number of articles published according to the language, the country of the researchers origin, the journals with the highest number of publications, articles by area of knowledge, the most productive authors on the subject, the number of authors per article, the most cited articles and the frequency of keywords. **Conclusion:** The indicators developed from this study contribute to improving the knowledge of the scientific community on the topic of Digital Preservation. Added to this is the potential to open paths for new researches, since a well-defined bibliographical research of a specific theme greatly facilitates the work of researchers in the development of new studies with innovative approaches.

**Keywords:** Digital preservation. Digital object. Bibliographical database. Web of Science. Scientific publication.

**LICENÇA DE USO**

Os autores cedem à [Revista Brasileira de Preservação Digital](#) os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença Creative Commons Attribution (CC BY) 4.0 International. Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

**PUBLISHERS**

Universidade Estadual de Campinas – Sistema de Bibliotecas / Instituto Brasileiro de Informação em Ciência e Tecnologia – Rede Brasileira de Serviços de Preservação Digital – Cariniana. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

**EDITORES**

Gildenir Carolino Santos, Miguel Angel Márdero Arellano.

**CREDIT**

**RECONHECIMENTOS:** Não aplicável.

**FINANCIAMENTO:** Não aplicável.

**CONFLITOS DE INTERESSE:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.

**APROVAÇÃO ÉTICA:** Não aplicável.

**DISPONIBILIDADE DE DADOS E MATERIAL:** Não aplicável.

**CONTRIBUIÇÕES DOS AUTORES:** Conceituação: CASTANHEIRA, P.; REZENDE, L.; Curadoria de Dados: CASTANHEIRA, P.; REZENDE, L.; Análise Formal: CASTANHEIRA, P.; REZENDE, L.; Investigação: CASTANHEIRA, P.; REZENDE, L.; Metodologia: CASTANHEIRA, P.; REZENDE, L.; Administração de Projetos: Laura Rezende; Software: CASTANHEIRA, P.; REZENDE, L., ALVES, R. A.; Redação – rascunho original: CASTANHEIRA, P.; REZENDE, L., R.; Redação – revisão & edição: CRUZ-RIASCOS, S. A.

Submetido em: 10/05/2023 – Aceito em: 19/05/2023 – Publicado em: 10/06/2023

## 1 INTRODUÇÃO

Com o desenvolvimento de novas tecnologias da informação e comunicação (TIC) na segunda metade do século XX, em especial com o computador e a Internet, profundas mudanças ocorreram na forma como o ser humano passa a lidar com a informação. As informações que anteriormente eram registradas em suportes analógicos passaram a ser produzidas e armazenadas em formato digital, dando origem ao artefato tecnológico chamado de objeto digital.

As TIC e os objetos digitais possibilitaram que as pessoas e organizações acessassem, registrassem, armazenassem e compartilhassem informações de qualquer lugar do planeta, com maior agilidade, ajudando a reduzir ou até a eliminar barreiras geográficas, através das redes de computadores (YAMAOKA; GAUTHIER, 2016).

Contudo, a informação registrada em formato digital possui algumas fragilidades, visto que o objeto digital depende de *hardwares* e *softwares*, além das mídias e dos formatos que estão sujeitos à degradação física e à obsolescência tecnológica dos suportes informacionais que podem impossibilitar gerações futuras a acessarem a informação armazenada em um determinado documento (MÁRDERO ARELLANO, 2004; BÁGGIO; FLORES, 2013, SANTOS; FLORES, 2015).

O risco da perda, parcial ou definitiva, da informação em formato digital instigou pesquisadores e organizações a desenvolverem estudos com o objetivo de propor estratégias de preservação digital a fim de garantir o acesso à informação em longo prazo (MÁRDERO ARELLANO, 2004). Verifica-se que esses estudos possuem um caráter multidisciplinar e contam com a contribuição de conhecimentos provenientes de diversas áreas do conhecimento, como: a Ciência da Informação, a Tecnologia da Informação e a Comunicação. Por esse motivo, depreende-se que esses estudos possuem abordagens e aplicações distintas que variam de acordo com o foco da investigação e a área de atuação do pesquisador.

A justificativa do presente estudo reside na necessidade de apresentar a produção científica sobre Preservação Digital com vistas a compreender a evolução do tema por meio dos artigos indexados na base de dados *Web of Science (WoS)* ao longo deste período. Acredita-se que esta análise fornecerá importantes subsídios para o desenvolvimento de novos estudos sobre Preservação Digital que podem beneficiar instituições públicas ou privadas que possuam a custódia de informações armazenadas em formato digital, sejam elas unidades de informação como: bibliotecas, arquivos e museus, ou como bases de dados, periódicos e demais recursos de informação.

Diante do exposto, segue-se a questão de pesquisa que orienta este estudo: Como está caracterizada a produção científica sobre Preservação Digital considerando periódicos indexados na base WoS nos últimos trinta anos?

Optou-se, portanto, por mapear o perfil e as características das pesquisas que abordam o tema Preservação Digital, apresentadas em artigos de periódicos indexados na *WoS*, no recorte temporal de 1992 a 2022.

Especificamente, buscou-se apontar a distribuição da publicação dos artigos ao longo dos anos, o número de artigos por idioma, o número de publicações por país, os periódicos mais produtivos na temática, o número de artigos por área de conhecimento, os autores mais produtivos em relação à temática, o número de autores por artigo, os artigos mais citados e a frequência de palavras-chave.

A pesquisa bibliográfica foi selecionada como técnica de coleta de dados por viabilizar a elaboração de indicadores que permitem avaliar a dinâmica e a evolução das publicações científicas. Dessa forma, esta pesquisa pode ser classificada como um estudo quantitativo com uma abordagem exploratória e descritiva da produção científica sobre Preservação Digital.

Para uma melhor compreensão, o presente artigo foi organizado em cinco seções: a primeira, composta por esta introdução; a segunda, apresenta a fundamentação teórica; a terceira, descreve os procedimentos metodológicos utilizados na coleta e organização dos dados; a quarta, apresenta a análise e interpretação dos dados, e por último, a quinta seção encerra este estudo com as considerações finais.

## 2 REFERENCIAL TEÓRICO

Discorre-se neste marco teórico sobre os temas preservação digital, objeto digital e estratégias de preservação digital representando os conceitos que compõem a base deste estudo.

### 2.2 O objeto digital e a preservação digital

Os estudos sobre preservação digital surgiram no século XX com o objetivo de garantir o acesso e a preservação das informações armazenadas em formato digital que estavam se deteriorando ou tornando-se obsoletas com o passar do tempo (BÁGGIO; FLORES, 2013).

Nesse sentido, Grácio, Fadel e Valentim (2013) compreendem a preservação digital como um conjunto de atividades e processos que visam garantir a acessibilidade e a autenticidade do objeto digital pelo maior período de tempo possível, a partir de plataformas tecnológicas diferentes daquelas disponíveis no contexto da sua criação.

A definição proposta por Grácio, Fadel e Valentim (2013) utiliza o termo objeto digital para fazer referência a toda e qualquer informação armazenada em formato

digital. De tal modo, para situarmos este estudo se faz necessário que se busque a definição do termo objeto digital, à luz do que já foi estudado sobre o tema.

Nesse contexto, Yamaoka e Gauthier (2016) informam que diversos termos têm sido utilizados pela literatura especializada para fazer referência ao objeto digital, e os mais conhecidos são: arquivo de computador, documento digital, artefato digital, recurso digital e material digital. Para este estudo, o termo objeto digital foi selecionado para atender à questão de organização e padronização dos conceitos.

Isto posto, Ferreira (2006) define o objeto digital como toda e qualquer informação representada por meio de dígitos binários. Na visão do autor, arquivos de imagem, áudio, vídeo, jogos e programas de computador são apenas exemplos do que pode ser considerado um objeto digital.

O objeto digital pode ser compreendido a partir de três níveis de abstração, o físico, constituído por *bits* em um suporte físico, nível lógico que segue regras e são interpretados por *softwares* e podem ser interpretados por seres humanos e por último o nível conceitual quando é interpretado pelo ser humano (THIBODEAU,2002).

Conforme Thibodeau (2002), a deterioração ou a obsolescência de qualquer um dos níveis do objeto digital pode impossibilitar, parcialmente ou definitivamente, o acesso à informação. Sayão (2005) reforça ao discorrer que os principais obstáculos para a preservação digital decorrem da complexidade dos objetos digitais que dependem de *hardwares*, *softwares*, das mídias e dos formatos que estão sujeitos à degradação física e à obsolescência tecnológica.

Por esse motivo, Márdero Arellano (2004) informa que a preservação digital deve incorporar a preservação física, lógica e intelectual dos objetos digitais. Para ele, a preservação física deve priorizar a preservação dos *hardwares*, das mídias e dos suportes, a preservação lógica deve focar a preservação dos *softwares* e formatos, e a preservação intelectual deve atentar para a preservação do conteúdo.

Arellano (2004) relaciona a degradação física com o desgaste do objeto digital no nível físico que pode ser causado por diversos fatores internos ou externos, como: a temperatura, umidade, mau uso, defeitos de fabricação, desastres naturais, entre outros. Contudo, Santos e Flores (2015) mencionam que a maior preocupação sobre a degradação física gira em torno do curto ciclo de vida dos objetos digitais. Segundo os autores, enquanto os objetos digitais duram apenas algumas décadas, suportes analógicos, como o pergaminho e o papel podem durar até centenas de anos. Assim, o curto ciclo de vida dos objetos digitais demanda que soluções sejam implementadas em um menor espaço de tempo, aumentando o custo e a complexidade dos processos de preservação.

Já Santos e Flores (2015) relacionam a obsolescência tecnológica com a substituição de uma tecnologia antiga por outra mais moderna. De acordo com os autores, a obsolescência tecnológica ocorre de forma cíclica, nos níveis de *hardware* e *software*, dos suportes e formatos, quando há sobreposição de uma tecnologia por outra, de forma que a tecnologia anterior se torna obsoleta e ultrapassada. Assim, os objetos digitais esquecidos ou deixados em suportes ou formatos obsoletos se tornam

incompatíveis com as novas tecnologias impossibilitando que as informações sejam acessadas.

Conforme Santos e Flores (2015), a degradação física e a obsolescência tecnológica demandam que as informações digitais sejam constantemente transferidas para novos suportes e formatos para serem preservadas, o que acaba envolvendo questões, como: incompatibilidade tecnológica, direitos autorais, segurança e confiabilidade dos dados, capacitação de técnicos, investimento em infraestrutura, disponibilização de recursos, planejamento, entre outros. Nesse sentido, Arellano alerta que:

O desafio é muito mais um problema social e institucional do que um problema técnico, porque, principalmente para a preservação digital, depende-se de instituições que passam por mudanças de direção, missão, administração e fontes de financiamento. (MÁRDERO ARELLANO, 2004, p. 21).

Nesse contexto, Innarelli (2012) destaca o papel fundamental das instituições detentoras de acervo permanente como bibliotecas, arquivos, museus, universidades e outros centros de memória no desenvolvimento de estratégias de preservação digital, uma vez que essas instituições possuem experiência na gestão da informação por serem tradicionalmente responsáveis por reunir, organizar, conservar e disseminar a informação. As principais estratégias de preservação digital abordadas pela literatura estão apresentadas em próximo tópico.

6

### *2.3 As estratégias de preservação digital*

As estratégias de preservação digital podem ser classificadas como estruturais e operacionais (MÁRDERO ARELLANO, 2004; SANTOS, FLORES, 2015; 2018; FORMENTON, GRACIOSO, CASTRO, 2015; FORMENTON, GRACIOSO; 2020). As estratégias estruturais envolvem as questões culturais, legais, econômicas e administrativas e as operacionais envolvem as questões técnicas. Em seu estudo, Formenton e Gracioso (2020) reúnem as principais estratégias de preservação digital utilizadas atualmente de acordo com a literatura especializada. Sendo assim, os dois quadros, a seguir, contém o nome e a descrição das principais estratégias citadas pelos autores.

O Quadro 1 apresenta as principais estratégias estruturais utilizadas atualmente. Elas são mais abrangentes e envolvem todos os aspectos organizacionais relacionados à preservação digital.

Quadro 1. Estratégias estruturais

Estratégias estruturais	
Estratégias	Descrição
Adoção de padrões abertos	Utilização de softwares com padrões abertos e não proprietários.
Políticas e estratégias institucionais	Elaboração de documentos que vão nortear as práticas de preservação digital dentro de uma instituição, abrangendo as normas institucionais e a legislação vigente.
Orçamentos e custos de preservação	Análise e orçamentos das despesas com os recursos humanos e materiais necessários para a manutenção dos objetos digitais.
Seleção e conformidade legal	Seleção do que deve ser preservado levando em conta os diversos aspectos legais, dentre eles, os direitos autorais.
Treinamento e desenvolvimento de pessoas	Manutenção de um programa de formação continuada com o objetivo de especializar os profissionais envolvidos nas atividades de preservação digital.
Metadados para preservação digital	Elaboração de metadados que irão facilitar o processo de gestão da informação digital.
Investimento e Infraestrutura tecnológica	Previsão de investimentos na infraestrutura tecnológica.
Formação de redes de preservação	Colaboração entre organizações com o objetivo de compartilhar recursos, experiências, padronizações e estratégias de preservação digital.

Fonte: Elaborada pelos autores com base em Formenton e Gracioso (2020).

O Quadro 2 apresenta as estratégias operacionais mais utilizadas atualmente. Elas são mais técnicas e específicas, sendo utilizadas na preservação dos três níveis do objeto digital: o nível físico, o nível lógico e o nível conceitual.

Quadro 2. Estratégias operacionais

Estratégias Operacionais	
Estratégias	Descrição

Definição do meio de armazenamento	Avaliação da viabilidade técnica dos sistemas de armazenamento.
Migração	Transferência da informação digital de um hardware ou software obsoleto para outro mais moderno, com o objetivo de prolongar o ciclo de vida da informação digital.
Transferência para suportes analógicos	Transferência da informação digital para suportes analógicos de longa duração como o papel ou o microfilme.
Emulação	Replicação do funcionamento de hardwares e softwares obsoletos com o objetivo de recuperar a informação digital no seu contexto original.
Conservação de tecnologia	Conservação e manutenção dos hardwares e softwares em seus formatos originais.
Arqueologia Digital	Recuperação da informação digital de hardwares que se degradaram ou de hardware e softwares que se tornaram obsoletos.
Arquivamento Web	Preservação de conteúdos digitais disponíveis na WEB.

**Fonte:** Elaborada pelos autores com base em Formenton e Gracioso (2020).

Apesar dos estudos de Formenton e Gracioso (2020) apresentarem as principais estratégias utilizadas atualmente, Santos e Flores (2018) alertam que o constante desenvolvimento de novas tecnologias faz com que as próprias estratégias e métodos de preservação digital se tornem obsoletos. Nesse sentido, na visão dos autores as técnicas e os processos de preservação digital devem ser frequentemente atualizados e aperfeiçoados.

### 3 PROCEDIMENTOS METODOLÓGICOS

A WoS é considerada uma das maiores bases de dados científicos do mundo, possui 1.9 bilhão de referências citadas em mais de 171 milhões de registros, foi criada pelo *Institute for Scientific Information (ISI)*, e atualmente é mantida pela *Clarivate Analytics*, empresa que possui sede na Filadélfia, nos Estados Unidos (CLARIVATE, 2021).

A WOS foi selecionada como ferramenta de acesso ao conteúdo necessário para a realização deste estudo. Os dados foram coletados na base de dados da WoS no dia 31 de dezembro de 2021. Em seguida, em 18 de abril de 2023, foi feita uma atualização dos resultados anteriores visando incluir o ano de 2022 no recorte temporal, totalizando os últimos 30 anos. Identificou-se, em uma pesquisa inicial, que

o primeiro artigo sobre o tema “Preservação digital” adotado na presente pesquisa, foi indexado na WoS somente a partir de 1992. Assim, a coleta de dados foi realizada englobando o recorte temporal desde o dia 01 de janeiro de 1992 até o dia 31 de dezembro de 2022. A amostra possui como detalhamento ser composta apenas por artigos que tratam especificamente sobre o tema “Preservação Digital”. Por esse motivo, foram considerados apenas os artigos que fizeram alusão à preservação do objeto digital e às estratégias de preservação digital, abordadas no referencial teórico deste estudo, buscando compreender como o tema vem sendo estudado durante este período.

A estratégia de busca escolhida utilizou o termo: *Digital Preservation*. Assim, o termo foi inserido no campo de busca avançada da WoS, com letras maiúsculas e entre aspas retas, sendo grafado da seguinte forma: “DIGITAL PRESERVATION”. Destaca-se que o campo de busca avançada da WoS não faz distinção entre letras maiúsculas e minúsculas. Portanto, foi configurado para pesquisar o termo “DIGITAL PRESERVATION” no título, no resumo e nas palavras-chave dos artigos. Além disso, foram utilizados os seguintes filtros:

- a) Somente artigos de periódicos;
- b) Somente artigos publicados entre 01/01/1992 e 31/12/2022;
- c) Somente artigos escritos em inglês, português e espanhol.

Após a busca foram encontrados 508 artigos de periódicos. Todos os 508 artigos foram analisados por título, resumo e palavras-chave, de modo que foram excluídos os artigos duplicados ou que não tratavam diretamente sobre o tema Preservação Digital. Dessa maneira, após o refinamento restaram 341 artigos que efetivamente compõem a amostra desta pesquisa. Os registros bibliográficos dos artigos foram exportados para o *software* Microsoft Excel (editor de planilhas eletrônicas) para a organização e compilação dos dados em gráficos e tabelas de frequência absoluta e relativa.

## 4 ANÁLISE E INTERPRETAÇÃO DOS DADOS

Esta seção apresenta as análises das produções científicas sobre Preservação Digital, com base nos indicadores elaborados a partir dos dados bibliográficos de artigos científicos publicados em periódicos indexados na base de dados WoS no recorte temporal de 1992 a 2022.

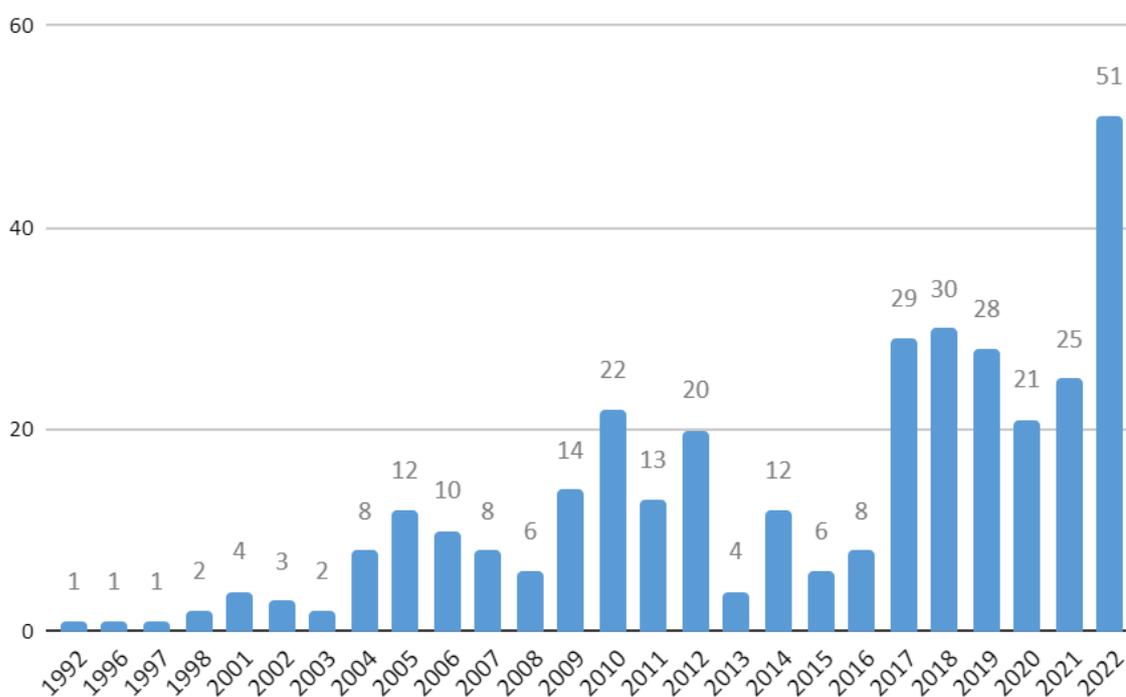
### 4.1 Distribuição dos artigos ao longo dos anos

O Gráfico 1 representa a produção anual de artigos indexados na WoS e foi elaborado com o objetivo de analisar o desenvolvimento da produção científica

indexada na WoS sobre o tema Preservação Digital no recorte temporal de 1992 a 2022.

O eixo horizontal do gráfico aponta o período em que os artigos foram publicados e o eixo vertical indica a quantidade de artigos publicados no período. Verifica-se que a primeira publicação relacionada com a temática de preservação digital definida anteriormente no escopo deste estudo surgiu em 1992. Percebe-se que não houve publicações nos anos de 1993, 1994 e 1995, com uma retomada em 1997 com uma publicação e em 1998 com duas publicações. Observa-se que não houve publicações nos anos de 1999 e 2000, contudo os estudos foram retomados em 2001 com quatro publicações. Entre 2001 e 2022 não houve mais períodos sem publicações. Com 51 estudos, 2022 foi o ano com o maior número de publicações na base de dados investigada.

**Gráfico 1.** Distribuição de artigos ao longo dos anos



Fonte: Dados da pesquisa (2023).

Apesar da variação do número de estudos publicados no período, percebe-se uma tendência de crescimento das produções científicas sobre Preservação Digital, pois 1.5% dos estudos foram publicados entre 1992 e 2000, 35.8% dos estudos foram publicados entre 2001 e 2011 e 62.8% dos estudos foram publicados entre 2012 e 2022.

Assim, pode-se inferir que os estudos sobre Preservação Digital surgiram no final do século XX, ganhando mais relevância no século XXI, em parte pelo constante desenvolvimento da tecnologia e em parte pelo grande volume de informação

armazenada em formato digital. Verifica-se, portanto, que o tema vem ganhando relevância ao longo do tempo e que cada vez mais pesquisadores estão desenvolvendo estudos sobre a temática visando encontrar melhores práticas para preservar o conteúdo digital por longo período de tempo.

#### 4.2 Número de artigos publicados segundo o idioma

A Tabela 1 foi elaborada com o objetivo de indicar o número de artigos publicados em um determinado idioma. A primeira coluna indica o idioma em que o artigo foi publicado, a segunda coluna aponta o número de artigos publicados em um determinado idioma e a terceira coluna apresenta o percentual de cada idioma em relação ao número total de artigos publicados.

**Tabela 1.** Número de artigos publicados segundo o idioma

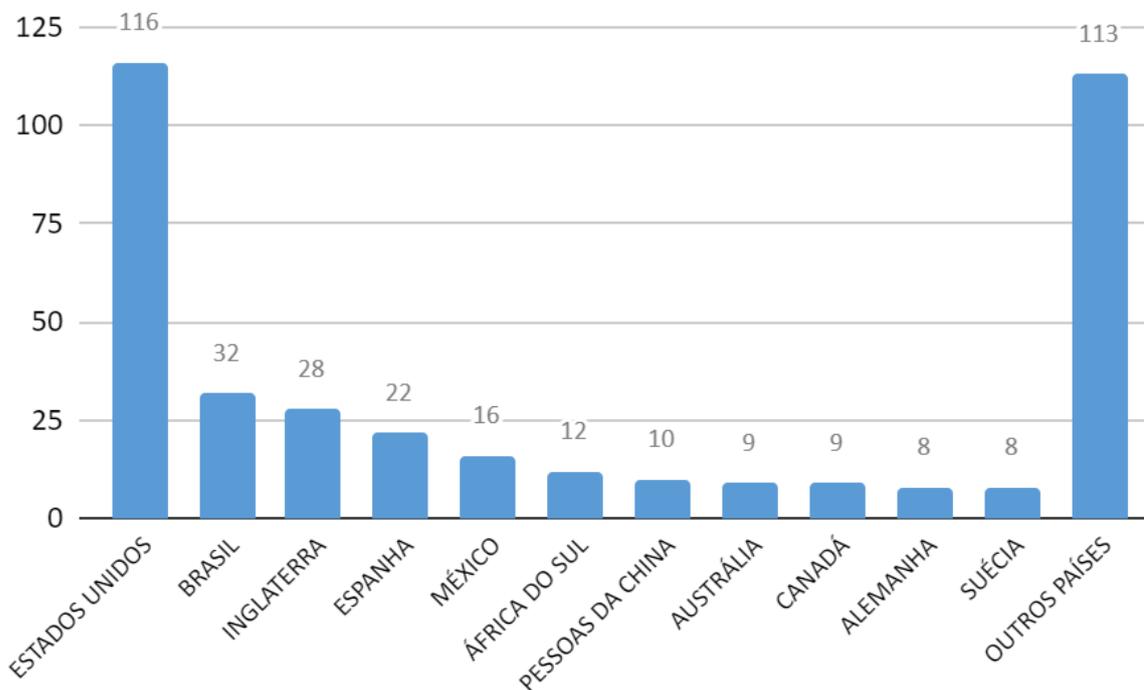
Idioma	Número de artigos	% de 341
Inglês	278	81,5
Espanhol	35	10,3
Português	28	8,2

Fonte: Dados da pesquisa (2023).

Ao todo, foram analisados 341 artigos, dos quais 278 foram publicados em inglês, 35 em espanhol e 28 em português. O grande número de artigos publicados em inglês pode ser justificado em parte porque a *WoS* é mantida pela *Clarivate Analytics*, empresa que possui sede nos Estados Unidos, país que possui o inglês como idioma dominante, e também, porque o inglês é utilizado por autores e pesquisadores de outros países como estratégia de internacionalização da comunicação científica.

#### 4.3 Número de publicações por país

O Gráfico 2 representa o número de artigos publicados em periódicos de um determinado país. O gráfico foi elaborado com o objetivo de verificar quais foram os países que mais desenvolveram estudos sobre Preservação Digital.

**Gráfico 2.** Número de publicações por país

Fonte: Dados da pesquisa (2023).

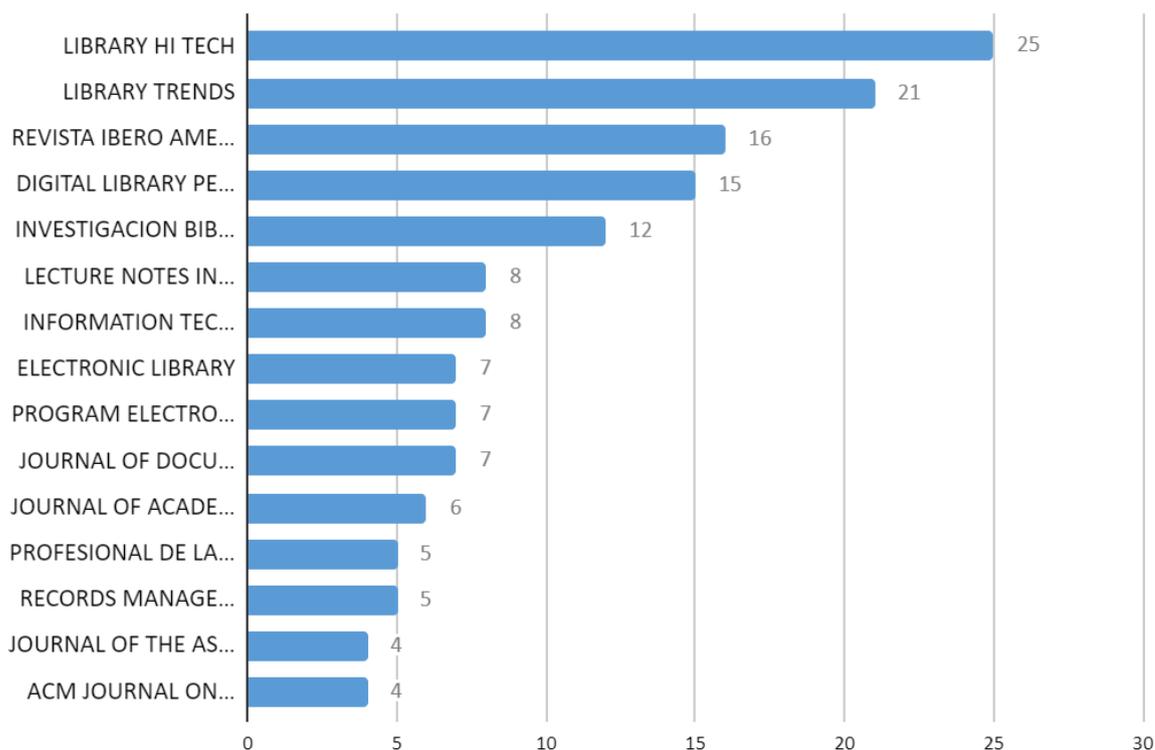
O eixo horizontal do gráfico indica o nome dos países e o eixo vertical informa a quantidade de artigos publicados por país. Ao todo, 55 países foram responsáveis por publicar 341 artigos sobre Preservação Digital. Os Estados Unidos com 116 estudos é o país que mais publicou estudos sobre o tema, seguido pelo Brasil com 32 estudos, a Inglaterra com 28 estudos, a Espanha com 22 estudos, o México com 16 estudos, a África do Sul com 12 estudos, a China com 10 estudos, a Austrália e o Canadá com 9 estudos, e a Alemanha e Suécia com 8. Os 113 artigos restantes foram publicados de maneira difusa por outros 44 países.

A análise do número de publicações por país possibilitou a identificação dos 11 países que mais se destacaram no desenvolvimento dos estudos sobre Preservação Digital. Nesse sentido, os Estados Unidos, Brasil, Inglaterra e Espanha constituem o grupo de países mais influentes na publicação de estudos sobre o tema, sendo responsável por 51.7% do total de publicações.

#### 4.4 Periódicos mais produtivos

O Gráfico 3 representa os periódicos que mais publicaram artigos sobre Preservação Digital.

Gráfico 3. Periódicos mais produtivos



Fonte: Dados da pesquisa (2023).

13

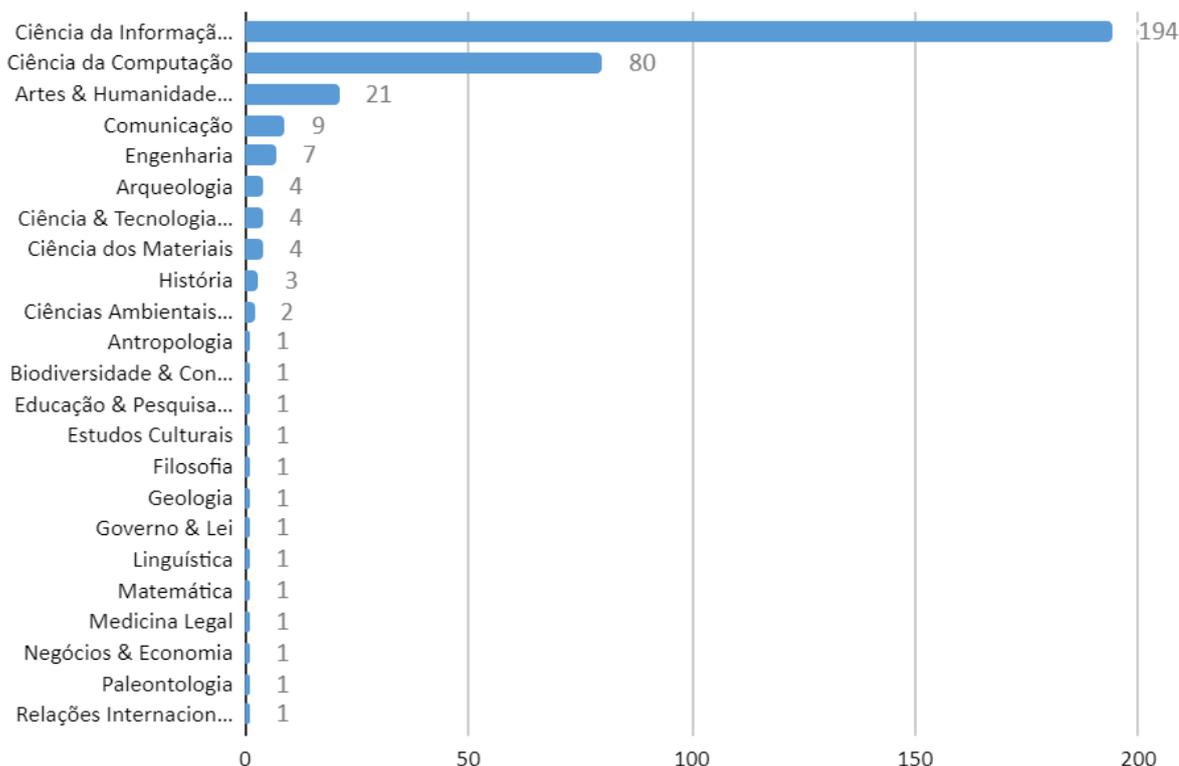
O eixo horizontal do gráfico informa a quantidade de artigos publicados ao longo de todo o período e o eixo vertical informa o nome dos periódicos. Verifica-se que os 341 artigos analisados por este estudo foram publicados por 156 periódicos. O periódico mais produtivo foi o *Library Hi Tech* com 25 estudos, seguido pelo *Library Trends* com 21 estudos, a *Revista Ibero Americana de Ciência da Informação* com 16 estudos, o *Digital Library Perspectives* com 15 estudos, o *Investigacion Bibliotecologica* com 12 estudos, o *Lecture Notes In Computer Science* e o *information Technology and Libraries* com 8 estudos, o *Electronic Library*, *Program Electronic Library And Information Systems* e *Journal of Documentation* com 7 estudos, o *Journal of Academic Librarianship* com 6 estudos, o *Profesional de La Información* e *Records Management Journal* com 5 estudos, e o *Journal of the Association for Information Science and Technology* e *ACM Journal on Computing and Cultural Heritage* com 4 estudos.

Constata-se que os 15 periódicos mais produtivos foram responsáveis por publicar 150 artigos, ou seja, 44% do total de publicações. A Análise dos periódicos mais produtivos permite verificar que a maioria deles estão associados às áreas de Ciência da Informação e Tecnologia da Informação.

4.5 Número de artigos por área do conhecimento

O Gráfico 4 representa o número de artigos por área de conhecimento e foi elaborado com o objetivo de verificar quais áreas do conhecimento desenvolveram mais estudos sobre o tema Preservação Digital.

Gráfico 4. Número de artigos por área do conhecimento



Fonte: Dados da pesquisa (2023).

O eixo horizontal do gráfico informa o número de artigos publicados por área do conhecimento e o eixo vertical indica as áreas do conhecimento. As áreas do conhecimento que mais produziram estudos sobre Preservação Digital foram: A Ciência da Informação com 194 estudos, seguido pela Ciência da Computação com 80 estudos, Artes & Humanidades - Outros Tópicos com 21 estudos, Comunicação com 9 estudos, Engenharia com 7 estudos, Ciência & Tecnologia - Outros Tópicos, Arqueologia e Ciência dos Materiais com 4 estudos, História com 3 estudos e Ciências Ambientais com 2 estudos.

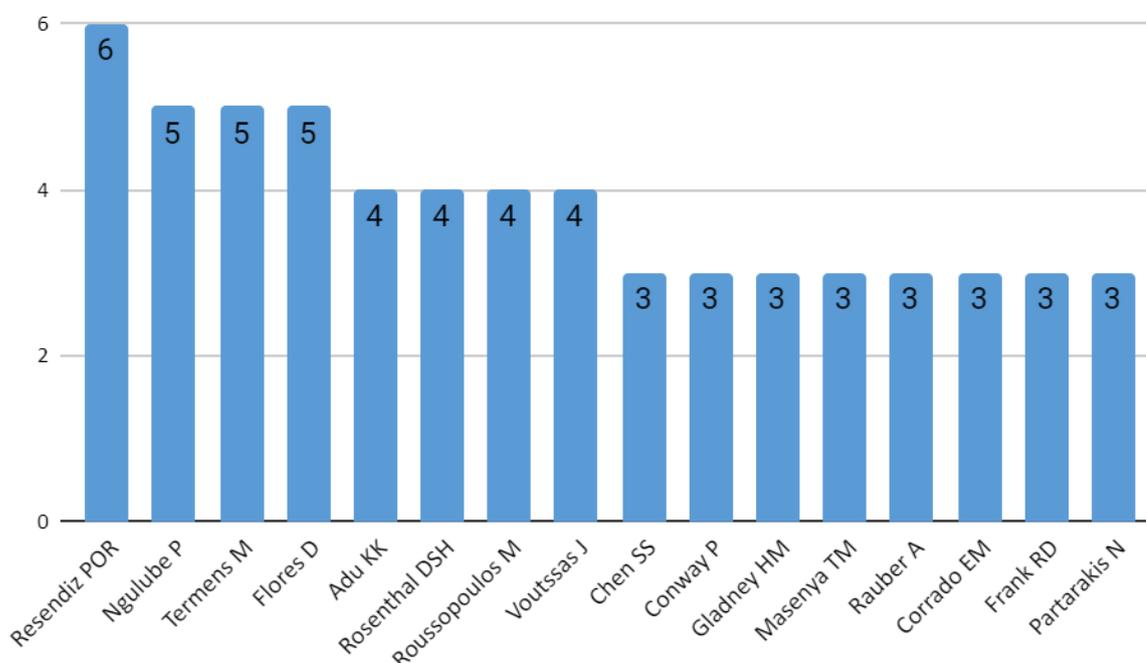
Verifica-se que a Ciência da Informação foi a área do conhecimento que mais se destacou no desenvolvimento de estudos sobre Preservação Digital, sendo responsável por 56.9% dos estudos publicados. Esses estudos desenvolvidos na área de Ciência da Informação podem ser segmentados em algumas subáreas como a

Arquivologia, a Biblioteconomia, a Museologia e a Gestão da Informação. Assim, pode-se inferir que as áreas e subáreas do conhecimento que de alguma forma se relacionam com instituições que possuem a informação digital em sua custódia, visando a perpetuação da memória imaterial, são as áreas e subáreas do conhecimento que mais desenvolveram estudos sobre o tema aqui estudado.

#### 4.6 Autores mais produtivos

O Gráfico 5 foi elaborado com o objetivo de apresentar os autores que mais publicaram artigos sobre o tema Preservação Digital.

Gráfico 5. Autores mais produtivos



Fonte: Dados da pesquisa (2023).

O eixo horizontal do gráfico apresenta o nome dos autores e o eixo vertical indica a quantidade de artigos publicados pelos autores no período estudado. Identificou-se que 718 autores foram responsáveis por publicar os 341 artigos analisados neste estudo. O autor mais produtivo é Resendiz, com seis estudos publicados; seguido por Ngulube, Termens e Flores, com cinco estudos; Adu, Rosenthal, Roussopoulos e Voutssas, com 4 estudos, e Chen, Conway, Gladney, Masenya, Rauber, Corrado, Frank e Partakis, com 3 estudos. Os 16 autores mais produtivos e responsáveis por publicar 61 estudos, representando 17.9% do total de publicações.

Pode-se observar que não há uma grande concentração de publicações em um grupo específico de autores. Essa difusão de autores pode ser considerada positiva, pois pode-se inferir que pesquisadores com perfis profissionais distintos estão publicando os resultados de suas pesquisas e fornecendo diferentes contribuições para o desenvolvimento dos estudos sobre preservação digital.

#### 4.7 Número de autores por artigo

A Tabela 2 representa o número de autores por artigo e foi elaborada com o objetivo de verificar o número de artigos publicados em co-autoria. Dessa forma, a primeira coluna indica o número de autores assinando o artigo, a segunda coluna apresenta o número de artigos publicados, e a terceira coluna informa o percentual em relação ao número total de artigos.

**Tabela 2.** Número de autores por artigo.

Número de autores assinando o artigo	Número de artigos	% de artigos n=341
1 autor	126	37,00
2 autores	107	31,40
3 autores	62	18,20
4 autores	18	5,30
5 autores	10	2,90
6 autores	3	0,90
7 autores	2	0,60
8 autores	3	0,90
9 autores	3	0,90
10 autores	3	0,90
12 autores	2	0,60
15 autores	1	0,30
16 autores	1	0,30

Fonte: Dados da Pesquisa (2023).

No que diz respeito à quantidade de autores por artigo, pode-se observar que 313 artigos (91.8%) possuem entre 1 e 4 autores, e 28 artigos (8.2%) possuem cinco ou mais autores. Observa-se que 215 artigos (63%) foram publicados em colaboração. Isto, por sua vez, aponta a importância da colaboração entre os autores para o desenvolvimento dos estudos sobre Preservação Digital, considerado uma temática recente e que carece de compartilhamento de experiências e estudos visando aprofundamento e amadurecimento das abordagens e melhores práticas.

#### 4.8 Artigos mais citados

Os 341 artigos analisados para este estudo foram citados 1512 vezes em todo o período. A análise dos artigos mais citados foi realizada com o objetivo de identificar os estudos mais relevantes sobre Preservação Digital, do ponto de vista da quantidade de citações, e estimar seu alcance junto à comunidade científica. Nesse sentido, verificou-se que os 10 artigos mais citados, ou seja, 3,44% do total de estudos analisados foram citados 443 vezes, sendo responsáveis por 30.05% de todas as citações realizadas no período.

A Tabela 3 apresenta uma classificação com os 10 artigos mais citados pela comunidade científica. A primeira coluna aponta a colocação do estudo, a segunda indica o ano de publicação, a terceira informa o número de vezes que o artigo foi citado, a quarta identifica os autores e a quinta nomeia os artigos.

**Tabela 3.** Artigos mais citados.

Classificação	Ano	Citações	Autores	Título do artigo
1º	2005	82	MANIATIS, P et al.	THE LOCKSS peer-to-peer digital preservation system.
2º	1997	73	HEADSTRO N, M.	Digital preservation: a time bomb for digital libraries.
3º	2007	52	IKEUCHI, K et al.	The great buddha project: digitally archiving, restoring, and analyzing cultural heritage objects.
4º	2010	47	CONWAY, P.	Preservation in the age of google: digitization, digital preservation, and dilemmas.
5º	2011	38	ABEL, RL et al.	Digital preservation and dissemination of ancient lithic technology with modern micro-ct.

6º	2014	34	MALONE, J et al.	The software ontology (swo): a resource for reproducibility in biomedical data analysis, curation and digital preservation.
7º	2002	33	LEE, KH et al.	The state of art and practice in digital preservation.
8º	2007	32	BRADLEY, K.	Defining digital sustainability.
9º	2006	30	HOCKX-YU, H.	Digital preservation in the context of institutional repositories.
10º	2009	22	HEDGE, M; BLANKE, T; HASSA, A.	Rule-based curation and preservation of data: a data grid approach using irods.

Fonte: Dados da pesquisa (2023).

O artigo mais citado foi o “*THE LOCKSS peer-to-peer digital preservation system*” com 82 citações, seguido pelo “*Digital preservation: a time bomb for digital libraries*” com 73 citações e o “*The great buddha project: digitally archiving, restoring, and analyzing cultural heritage objects*” com 52 citações.

Destaca-se que, dentre os autores com os artigos mais citados, apenas Conway, P. consta na lista de autores mais produtivos, representado no Gráfico 5. Assim, parece correto afirmar que não há uma relação direta entre os estudos mais citados e os autores mais produtivos.

#### 4.9 Frequência de palavras-chave

A Imagem 1 apresenta uma nuvem de palavras-chave que foi elaborada com o objetivo de identificar os termos mais recorrentes nos estudos sobre Preservação Digital. Por esse motivo, a Imagem 1 relaciona apenas os termos que apresentaram mais de cinco recorrências. Das palavras-chave identificadas através dos 341 artigos analisados, apenas 21 obtiveram cinco ou mais incidências. Os termos que mais se destacaram foram o *digital preservation* com 151 recorrências, o *digital libraries* e *Archives* com 21, *Preservation* com 19, e o *digital storage* com 17.

Imagem 1. Nuvem de palavras-chave



Fonte: Dados da pesquisa (2023).

As palavras-chave mais utilizadas para indexar os artigos sobre o tema Preservação Digital fazem referência a elementos específicos das áreas de Ciência da Informação e Tecnologia da Informação, quais sejam: *academic libraries*, *archives*, *digital libraries*, *institutional repository* e *digital repository*. Pode-se inferir que estes elementos identificados nas palavras-chave se justificam devido à importância da participação das instituições custodiais em projetos que estudam e implementam ações de preservação da informação armazenada em formato digital. Além disso, essas instituições em geral são responsáveis pela custódia definitiva de bens culturais cujas referências descritivas cada vez mais se encontram em formato digital.

19

## 6 Considerações Finais

Este estudo teve o objetivo de mapear e caracterizar a produção científica sobre Preservação Digital de artigos científicos publicados em periódicos indexados na base de dados WoS no período de 1992 a 2022.

A pesquisa bibliográfica possibilitou elaborar diversos indicadores que permitiram analisar aspectos como: (1) a distribuição dos artigos ao longo dos anos; (2) o número de artigos publicados segundo o idioma; (3) o número de publicações por país; (4) os periódicos mais produtivos na temática em questão; (5) o número de artigos por área do conhecimento; (6) os autores mais produtivos; (7) o número de autores por artigo; (8) os artigos mais citados; e, (9) a frequência de palavras-chave.

A análise e interpretação dos dados permitiu verificar que o interesse acadêmico sobre o tema Preservação Digital é crescente e multidisciplinar. Constatou-

se que apesar de haver uma certa concentração de estudos nas áreas de Ciência da Informação e Tecnologia da Informação, outras áreas do conhecimento estão desenvolvendo estudos sobre o tema reforçando sua relevância.

Os 341 artigos científicos publicados em inglês, português e espanhol analisados, foram publicados por 156 periódicos em 55 países. Tal fato revela a preocupação da comunidade científica em realizar estudos objetivando garantir que a informação digital permaneça acessível no futuro. A caracterização destes estudos é fundamental para aprimorar o conhecimento da comunidade científica sobre o tema Preservação Digital e viabilizar o compartilhamento de conhecimentos e experiências entre acadêmicos de diferentes áreas do conhecimento.

Os dados aqui quantificados abrem caminho para que novos estudos sobre Preservação Digital sejam desenvolvidos uma vez que este trabalho possui o intuito de auxiliar pesquisadores a terem maior familiaridade com o tema.

Destaca-se que diversos estudos de pesquisadores brasileiros foram utilizados como referência para compor a fundamentação teórica deste artigo. Uma análise minuciosa dos dados bibliográficos dos artigos indexados na *WoS*, no recorte temporal observado, revelou que, embora a primeira publicação indexada na base seja de 1992, as primeiras pesquisas de autores brasileiros sobre o tema Preservação Digital foram indexadas na *WoS* somente a partir de 2007, contudo apesar do curto período de tempo, o Brasil foi identificado como o segundo país que mais publicou estudos sobre o tema.

Ainda no contexto brasileiro, destaca-se a participação de pesquisadores em importantes grupos de pesquisa sobre Preservação Digital, como a Rede Cariniana e o grupo de pesquisa “Estudos e Práticas de Preservação Digital - Triade”. A constituição de grupos de pesquisa e o número crescente de publicações sobre o tema indicam que o Brasil terá um desempenho ainda mais relevante no futuro. Um fato concreto que caminha nesta direção é a Revista Brasileira de Preservação Digital (RBPD) que teve seu primeiro número publicado em 2020, e até o momento, já publicou 23 artigos, em sua maioria de pesquisadores brasileiros. Destaca-se que os artigos deste periódico não constam nos dados coletados deste estudo uma vez que ele ainda não está indexado na base pesquisada. Dito isto, temos realmente um futuro promissor no contexto brasileiro uma vez que há 4 anos existe um periódico exclusivo para a temática de preservação digital, assuntos correlatos e seus desdobramentos.

A utilização da *WoS* como única base de dados se constitui como um fator limitante para os resultados desta pesquisa, pois impede que os resultados sejam generalizados. Contudo, a escolha de uma única base de dados se justifica devido à utilização de metadados padronizados, algo fundamental para a elaboração de indicadores bibliométricos. A falta de padronização de metadados entre as bases científicas prejudica o intercâmbio entre informações e dificulta a análise da produção científica em maior escala.

Como proposta para estudo futuro, recomenda-se que a metodologia aplicada neste estudo seja replicada para analisar outras bases de dados científicas com o intuito de gerar um conjunto de dados mais abrangente acerca das produções científicas relacionadas ao tema da Preservação Digital, a fim de viabilizar uma análise com outras abordagens que permitam a geração de mais conhecimento sobre a área.

## Referências

BAGGIO, Claudia Carmen; FLORES, Daniel. Documentos digitais: preservação e estratégias. **BIBLOS** - Revista do Instituto de Ciências Humanas e da Informação, Rio Grande, RS, v. 27, n.1, p. 11-24, 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/23959>. Acesso em: 22 maio. 2023.

CLARIVATE ANALYTICS. Web of Science, 2021. Disponível em: <https://clarivate.com/webofsciencgroup/solutions/web-of-science/>. Acesso em: 22 maio. 2023.

FERREIRA, Aurélio Buarque de Holanda. **Mini Aurélio**: o dicionário da língua portuguesa. Curitiba: Positivo, 2010.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <https://bit.ly/42dUVrb>. Acesso em: 22 maio. 2023.

FORMENTON, Danilo. GRACIOSO, Luciana de Souza; CASTRO, Fabiano Ferreira de. Revisitando a preservação digital na perspectiva da ciência da informação: aproximações conceituais. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v.13, n.1, p. 170-191, 2015. Disponível em: <https://brapci.inf.br/index.php/res/v/40069>. Acesso em: 22 maio. 2023.

FORMENTON, Danilo; GRACIOSO, Luciana de Souza. Preservação digital: desafios, requisitos, estratégias e produção científica. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v. 18, p. 1-27, 2020. Disponível em: <https://bit.ly/3MLEJYv>. Acesso em: 22 maio. 2023.

GRÁCIO, José Carlos Abbud. **Preservação digital na gestão da informação**: um modelo processual para as instituições de ensino superior. São Paulo: Cultura Acadêmica, 2012. Disponível em: <https://repositorio.unesp.br/handle/11449/113727>. Acesso em: 22 maio. 2023.

GRÁCIO, José Carlos Abbud; FADEL, Bárbara; VALENTIM, Marta Lígia Pomim. Preservação digital nas instituições de ensino superior: aspectos organizacionais, legais e técnicos. **Perspectivas em Ciência da Informação**, v. 18, n.3, p. 111-129, jul./set. 2013. Disponível em: <https://bit.ly/41Z19Lj>. Acesso em: 22 maio. 2023.

GUEDES, V. L. S.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento em sistemas de informação, de comunicação e de avaliação científica e tecnológica. *In*: **ENCONTRO NACIONAL**

**DE CIÊNCIA DA INFORMAÇÃO**, 6., 2005, Salvador. Anais [...]. Salvador: UFBA, 2005. Disponível em: <https://bit.ly/3ovkq9j>. Acesso em: 22 maio. 2023.

INNARELLI, Humberto Celeste. Preservação Digital: a influência da gestão dos documentos digitais na preservação da informação e da cultura. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 8, n.2, p. 72-87, jan./jun. 2011. Disponível em: <https://bit.ly/3Mub5GX>. Acesso em: 22 maio. 2023.

MÁRDERO ARELLANO, Miguel Angel. Preservação de documentos digitais. Preservação de documentos digitais. **Ciência da Informação**, Brasília, DF, v. 33, n. 2, p. 15-27, 2004. Disponível em: <http://revista.ibict.br/ciinf/article/view/1043>. Acesso em: 22 maio. 2023.

MEDEIROS, José Mauro Gouveia; VITORIANO, Maria Albeti Vieira. A evolução da bibliometria e sua interdisciplinaridade na produção científica brasileira. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 13, n.3, p. 491-503, set./dez. 2015. Disponível em: <https://bit.ly/3OxLlqq>. Acesso em: 22 maio. 2023.

SANTOS, Henrique Machado dos; FLORES, Daniel. Políticas de preservação digital para documentos arquivísticos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 4, p. 197-217, out./dez. 2015. Disponível em: <https://bit.ly/420BXDY>. Acesso em: 22 maio. 2023.

SANTOS, Henrique Machado dos; FLORES, Daniel. A obsolescência do conhecimento em preservação digital. **Ciência da Informação em Revista**, Maceió, v. 5, n. 1, p. 41-58, mai. 2018. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/3337>. Acesso em: 22 maio. 2023.

SAYÃO, Luis Fernando. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: MARCONDES, Carlos H. *et al.* **Bibliotecas digitais: saberes e práticas**. Salvador, BA: EDUFBA; Brasília, DF: IBICT, 2005. p. 115-146. Disponível em: <https://bit.ly/42T5Oj4>. Acesso em: 22 maio 2023.

THIBODEAU, Kenneth. Overview of technological approaches to digital preservation and challenges in coming years. In: THE STATE OF DIGITAL PRESERVATION: AN INTERNATIONAL PERSPECTIVE. **Anais** [...] Washington: CLIR and Library of Congress, 2002. Disponível em: <https://bit.ly/3BMDpzh>. Acesso em: 22 maio. 2023.

YAMAOKA, Eloi Juniti; GAUTHIER, Fernando Ostuni. Objetos Digitais: em busca da precisão conceitual. **Informação & Informação**, João Pessoa, v. 18, n. 2, p. 77-97, 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/33936>. Acesso em: 22 maio. 2023.