

ESTADO DA ARTE DA PRESERVAÇÃO DIGITAL E AS TEMÁTICAS
PESQUISADAS POR PESQUISADORES BRASILEIROS

*STATE OF THE ART OF DIGITAL PRESERVATION AND THE THEMES
RESEARCHED BY BRAZILIAN RESEARCHERS*

 João Paulo Pastana Neves¹

 José Carlos Abbud Grácio²

¹ Doutorando em Ciência da Informação pelo PPGCI/Unesp e Mestre em Ciência da Informação pelo PPGCI/UFPA. Bibliotecário-documentalista efetivo do Centro de Memória da Faculdade de Medicina/ICM/UFPA.


E-mail: jp.neves@unesp.br

² Doutor em Ciência da Informação pelo PPGCI/Unesp. Prof. Dr. em Ciência da Informação pelo PPGCI/Unesp.

E-mail: jose.gracio@unesp.br



ACESSO ABERTO

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Não há.

Declaração de Disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 08 fev. 2024.

Aceito em: 17 set. 2024.

Publicado em: 15 nov. 2024.

Como citar este artigo:

NEVES, J. P. P.; GRÁCIO, J. C. A. Estado da arte da preservação digital e as temáticas pesquisadas por pesquisadores brasileiros. **Informação em Pauta**, Fortaleza, v. 9, p. 1-16, 2024.

RESUMO

Apresenta-se o estado da arte da preservação digital no Brasil. Nessa perspectiva, realizou-se uma revisão sistemática da literatura nas fontes de informação Web of Science, Scopus e Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação. Aplicou-se a análise bibliométrica e categorial nas publicações científicas (artigos, capítulos de livro e comunicações publicadas em anais de eventos), identificadas e selecionadas. Os termos aplicados na estratégia de busca na Web of Science e na Scopus foram 'digital preservation', 'archival documents scanning' e 'Brazil', e na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação utilizou-se apenas 'preservação digital'. Delimitou-se o período de 2000 até 2022. Recuperou-se 1.428 documentos no total, dos quais 324 foram selecionados devido à pertinência e alinhamento à pesquisa. Identificaram-se os autores mais produtivos e as principais temáticas pesquisadas. Evidenciou-se que a temática preservação digital tem sido objeto de pesquisa nos últimos anos devido a distintos fatores, entretanto, observou-se a preocupação em manter viva à memória institucional, bem como o incentivo da sustentabilidade digital.

Palavras-chave: preservação digital; digitalização; revisões de literatura; bibliometria. categorias; Brasil.

ABSTRACT

The state of the art of digital preservation in Brazil is presented.

From this perspective, a systematic review of the literature was carried out in the information sources Web of Science, Scopus, and Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação. Bibliometric and categorical analysis was applied to scientific publications (articles, book chapters and communications published in proceedings of events), identified, and selected. The terms applied in the search strategy on the Web of Science and Scopus were 'digital preservation', 'archival documents scanning' and 'Brazil', and in the Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação it used only 'digital preservation'.

The period from 2000 to 2022 was delimited. A total of 1,428 documents were retrieved, of which 324 were selected due to their pertinence and alignment with the research. The most productive authors and the main themes researched were identified. It was evident that the subject of digital preservation has been the subject of research in recent years due to different factors, however, there was a concern to keep the institutional memory alive, as well as the encouragement of digital Sustainability

Keywords: digital preservation; digitization; literature reviews; bibliometrics; categories; Brazil.

1 INTRODUÇÃO

Com a produção de documentos digitais e os fluxos informacionais em constante crescimento, devido ao acelerado avanço das Tecnologias de Informação e Comunicação (TIC), a Preservação Digital (PD) é essencial para preservar objetos digitais, visando a constituição e manutenção de memórias institucionais, evitando assim a perda desses objetos em função da obsolescência de hardware, software, suportes e formatos obsoletos como os Compact Disc Read Only Memory (CD-ROM), pen drives, hardware e software em geral. Nesse contexto, instituições públicas e privadas de diferentes segmentos vivenciaram situações de perda de conteúdos importantes sobre sua história.

Uma das iniciativas para evitar essas perdas é o uso de repositórios digitais, que estão cada vez mais presentes nas comunidades acadêmicas científicas brasileiras, tendo em vista a conscientização do valor agregado que possuem os objetos digitais, como documentos e dados científicos, e os riscos da perda desses objetos ao longo do tempo. Nesse contexto, é fundamental implementar uma infraestrutura eficiente voltada aos ambientes digitais nas universidades, visando propiciar apoio aos investigadores na gestão e disponibilização de, por exemplo, dados científicos, possibilitando a reutilização por outros cientistas. Nessa perspectiva, deve-se não somente formular suas políticas, como também os objetivos e as características da comunidade atendida (Monteiro; Sant'ana, 2018).

Nesse cenário, há a necessidade e a importância de se preservar a memória institucional e científica institucional em repositórios digitais para a salvaguarda, tratamento, armazenamento, preservação e recuperação desses objetos digitais, para o uso corrente e futuro de pesquisadores, cientistas, historiadores e acadêmicos em geral.

Nesse intuito, surge a seguinte pergunta de pesquisa: de que maneira os pesquisadores brasileiros vêm desenvolvendo as temáticas de pesquisa sobre PD? Resaltando que devemos desenvolver as atividades com o propósito de preservar os objetos digitais, não somente por questões, mas também para preservar a memória institucional. Nesse sentido, o objetivo desta pesquisa é analisar, por meio da literatura, o estado da arte da PD no País, a ocorrência dos temas pesquisados por pesquisadores brasileiros, desde as primeiras manifestações até os dias atuais.

2 PRESERVAÇÃO DIGITAL

Há muito tempo o homem criou o papiro e o papel e, por meio desses artefatos, conseguiu armazenar sua memória e história.

Entretanto, à medida que o volume documental/informacional cresceu, principalmente em função do surgimento da informação digital, surgiu a necessidade de armazenar grandes volumes em espaços reduzidos. Para tanto, as TIC têm papel fundamental nesse processo, pois além de possibilitar a redução de espaço no que tange ao armazenamento, também proporciona o rápido acesso tanto in loco quanto remotamente.

No entanto, o uso das TIC não garante a integridade dos conteúdos e/ou suportes, pois os objetos digitais são frágeis e podem ser danificados de várias maneiras. Para atenuar esse problema, a PD ganha destaque, pois atua visando manter os objetos digitais seguros sem o risco de perder sua autenticidade, a confiabilidade e a integridade.

A PD visa salvaguardar, tratar, armazenar, preservar e recuperar os objetos digitais, para uso corrente e futuro de usuários, pesquisadores, cientistas, historiadores e acadêmicos em geral. Com esse entendimento, Ferreira (2006), Márdero Arellano (2008) e Grácio (2012) definem a PD como o conjunto de atividades ou processos responsáveis por garantir o acesso contínuo no longo prazo à informação e ao patrimônio cultural existente em formatos digitais, pois consiste na capacidade de garantir que a informação digital permaneça acessível, com qualidade e autenticidade suficiente para poder ser interpretada no futuro, independentemente da interface tecnológica utilizada.

Para Hedstron (1998), a PD é definida como o planejamento, alocação de recursos e aplicação de métodos e tecnologias de preservação necessária para garantir que a informação digital de valor contínuo permaneça acessível e utilizável.

Nessa perspectiva, a PD pode ser entendida como “[...] os processos de gestão envolvida na administração das atividades necessárias para garantir que um objeto digital possa ser acessado e utilizado no futuro, a partir das TIC existentes na época e com garantias de sua autenticidade e integridade” (Grácio et al., 2020a). Além disso, o autor menciona que há elementos significativos que podem definir PD, distribuídos em três eixos: o organizacional (institucional), o legal e o técnico (Figura 1).

Figura 1 - Elementos que definem a PD



Fonte: Grácio et al. (2020b, p. 570).

Para melhor compreensão sobre esses elementos, Grácio (2012) explica que o eixo organizacional (institucional) está relacionado a gestão estratégica de uma instituição e, portanto, está vinculado aos objetivos de PD à comunidade e às pessoas envolvidas:

- a. na definição de responsabilidades;
- b. na composição de equipe multidisciplinar que realize a gestão dos processos;
- c. na alocação de recursos financeiros;
- d. na garantia da autenticidade;
- e. na definição de uma política de preservação digital.

No que tange aos elementos legais, estes estão relacionados aos direitos autorais, às leis e atos administrativos que regem os objetos digitais, como as normativas internacionais e/ou nacionais, de modo a garantir a legalidade dos processos que envolvem a PD de uma instituição.

Em relação aos elementos técnicos, estes estão relacionados às questões técnicas envolvidas nas atividades informáticas, biblioteconômicas, arquivísticas e de outras áreas correlacionadas e envolvem: a) seleção do que preservar e do que descartar; b) definição de modelos e padrões; c) participação em iniciativas nacionais e internacionais que possibilitem a troca de experiência e conhecimento entre instituições; d) implementação da infraestrutura tecnológica para preservação e acesso aos objetos digitais preservados; e) aplicação de estratégias adequadas de PD; f) definição e aplicação de suporte adequado; g) representação por meio de metadados.

Desse modo, evidencia-se que esses são elementos fundamentais para a elaboração de uma política de PD para qualquer tipo de organização/instituição que almeje garantia de autenticidade e integridade de seus objetos digitais, que representam sua memória Institucional.

A PD eficiente sem dúvida se inicia com a formulação de uma política de PD compreendida como:

[...] documento que define os objetivos e diretrizes de uma instituição para a implantação de um programa de preservação de seus registros e objetos digitais. Essa política deve estar alinhada com os objetivos e com as outras políticas institucionais em vigor e ser revista ao longo do tempo, abrangendo todos os elementos relacionados à preservação digital, inclusive aquelas relacionadas à cultura organizacional e as mudanças nas TIC. Dessa forma, constitui um compromisso institucional para garantir o acesso e a integridade de um acervo, como no caso dos documentos de arquivo digitais (Grácio et al., 2020b, p. 569).

Além disso, a política de PD deve contribuir com os processos organizacionais, no intuito de garantir qualidade, integridade e autenticidade dos documentos inerentes, de modo que a informação seja concebida e interoperável no futuro por meio de uma plataforma distinta daquela na qual foi criada, contemplando todo o processo que envolve a preservação física, lógica e intelectual de um determinado objeto digital.

2.1 Objetos Digitais

Objetos digitais se referem a dados e informações armazenados em meio digital que podem ser representados por sequências de dígitos binários, podendo ser: textos científicos; fotos; banco de dados de conhecimento científico; vídeos; podcast; modelos de realidade virtual e aumentada; páginas web; software; e-mail; etc. (Ferreira, 2006; Grácio, 2012).

Esses exemplos de objetos digitais, por suas características distintas, evidenciam as diferentes maneiras de planejar e implementar a PD. Do mesmo modo, no que tange a busca e recuperação desses objetos digitais é essencial a descrição

dos metadados, como um dos elementos que garantem a autenticidade desses objetos.

2.2 Metadados

Os metadados podem ser definidos como “[...] conjuntos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação” (Grácio, 2012, p.122). Para Ferreira (2006), metadados são como metainformação utilizada para descrever um determinado objeto ou recurso. Grácio (2013) menciona os metadados são considerados fundamentais para garantir a autenticidade de um objeto digital.

Os metadados como estratégia de PD são classificados da seguinte maneira (Quadro 1):

Quadro 1 - Classificação dos tipos de metadados como estratégia de PD

METADADOS	AUTORES
Descritivos	<p>“[...] relativos ao conteúdo do objeto digital” (Grácio; Fadel; Valentim, 2013, p. 125).</p> <p>“[...] visam à pesquisa, recuperação e identificação. Podem conter elementos como, por exemplo, título, autor, assunto e palavras-chave” (Barbedo; Corujo; Sant’Ana, 2019).</p> <p>“[...] detalham um recurso digital para localização, identificação ou compreensão. Podem incluir propriedades ou elementos, tais como título, autor e assunto, em que os usos primários são descobertos, apresentação e interoperabilidade” (Formenton; Gracioso, 2022, p. 5).</p>
Estruturais	<p>“[...] vinculam de forma hierárquica distintos objetos digitais (textos, imagens, áudios etc.) integrantes de um mesmo documento ou recurso informacional” (Barbedo; Corujo; Sant’Ana, 2011).</p> <p>“[...] explicitam a estrutura interna do arquivo digital e as relações hierárquicas de partes integrantes de recursos entre si. Podem ter propriedades, como ordem e lugar na hierarquia, em que os usos primários são navegação e apresentação” (Formenton; Gracioso, 2022, p. 5).</p>
Administrativos	<p>“[...] documentam os atos de gestão ao longo do tempo para preservação do recurso” (Saramago, 2003, p.1 <i>apud</i> Grácio, 2012, p. 124).</p> <p>“[...] fornecem informações que apoiam a gestão do ciclo de vida (criação, seleção, descrição etc.) dos recursos informacionais. Podem incluir propriedades, tais como tipo e tamanho de arquivo, data/hora de criação, evento de preservação, status dos direitos autorais e termos de licença, em que os usos primários são interoperabilidade, gerência de objetos digitais e preservação” (Formenton; Gracioso, 2022, p. 5).</p>
Preservação	<p>“[...] é a informação que apoia e documenta a preservação de longo prazo de materiais digitais” (Sayão, 2010, p.10).</p> <p>“[...] incluem informações (por exemplo, as dependências de hardware e de software) exigidas para a gerência de um arquivo digital a longo prazo” (Formenton; Gracioso, 2022, p. 5).</p>

Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Apresenta-se no Quadro 1 as classificações de metadados definidas por autores que os compreendem como estratégias de PD. Além disso, no âmbito na PD os metadados

têm finalidades além da busca e recuperação da informação, visto que são essenciais na gestão e na preservação (Santos; Flores, 2020, p.81). Destaca-se, também, como os metadados podem atribuir mais qualidade aos objetos digitais no contexto da PD, sendo assim apresenta-se os principais padrões de metadados para a PD (Quadro 2):

Quadro 2 - Esquemas e padrões de metadados para preservação digital.

PADRÕES	CARACTERÍSTICAS
Dublin Core	Existem dois níveis no padrão DC: Simple e Qualificado. Sendo assim, o DC Simple inclui quinze elementos; e o DC Qualificado compreende três elementos adicionais (Audiência, Proveniência e Detentor de Direitos), assim como um grupo de refinamentos ou qualificadores de elementos, que aperfeiçoam a semântica dos elementos de maneira que possam ser úteis na descoberta de recursos. O DC apresenta algumas características, que refletem também os seus objetivos, dentre os quais são as seguintes: simplicidade, interoperabilidade semântica, consenso internacional, extensibilidade e flexibilidade (Hillmann, 2005; Grácio, 2012).
MODS	O esquema MODS fornece informações bibliográficas importantes que apoiam outros padrões expressos em XML, como o <i>Metadata Encoding and Transmission Standard</i> (METS) e o <i>PREservation Metadata: Implementation Strategies</i> (PREMIS). Sob o enfoque da preservação digital, três elementos MODS merecem atenção: Informação de Origem, Item Relacionado e Condição de Acesso. Estes elementos registram informações que são úteis para auxiliar os metadados de preservação na corroboração da proveniência, autenticidade e integridade dos objetos digitais, bem como no reconhecimento dos direitos associados ao recurso que podem intervir, sobretudo, na realização das ações de preservação, de acesso e de utilização dos seus respectivos conteúdos.
EAD	Na preservação digital o uso do EAD é cabível pois, para Barbedo, Corujo e Sant'Ana (2011), trata-se de um esquema completo e que atende as normas de descrição arquivística. Os padrões de metadados descritivos, tal como o DC, o MODS e o próprio EAD, apesar de serem mais propícios à descoberta e recuperação de recursos ao invés da preservação de longo prazo, são esquemas importantes para o registro de algumas informações de apoio aos metadados de preservação, no tocante à proveniência, autenticidade, integridade, contexto técnico, direitos, e à custódia/propriedade dos recursos armazenados num repositório de preservação digital.
ANSI/NISO Z39.87	Na preservação digital o uso do padrão pode ser contextualizado através de dois projetos notáveis: em Portugal, o Repositório de Objectos Digitais Autênticos (RODA) da Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB) e da Universidade do Minho; e nos Estados Unidos, <i>University of Houston Digital Library</i> (UHDL).
METS	O METS é relevante na preservação digital, pois está escrito em XML, entretanto, a flexibilidade na implementação do METS pode inibir o intercâmbio de registros causando assim problemas de interoperabilidade entre sistemas; e na utilização do METS com o PREMIS pode ocorrer duplicações ou, ainda, uma parcial harmonia entre seus elementos (Rodrigues, 2008; Sayão, 2010; Lavoie; Gartner, 2013).
PREMIS	Promovido pela LC, o <i>PREMIS Maintenance Activity</i> oferece um esquema XML que corresponde diretamente ao Dicionário de Dados, permitindo a descrição de objetos, eventos, agentes e direitos, assim como a utilização do PREMIS representado em XML para a gerência e o intercâmbio de metadados entre sistemas de preservação.

Fonte: Adaptado de Formenton et al. (2017).

Os padrões de metadados apresentados no Quadro 2 são aplicados tanto na fase de produção de documentos digitais em relação ao formato, na fase de preservação, quanto

na fase de acesso, sem eles não há a ocorrência de interoperabilidade eficaz entre sistemas como, por exemplo, o Sistema de Gestão Arquivística de Documentos (SIGAD), o Repositório Arquivístico Digital Confiável (RDC-Arq) e o Access to Memory (AtOM). Em um destes exemplos, a norma ISO 14721:2012 'Space data and information transfer systems - Open archival information system (OAIS) - Reference model' informa que um RDC deve seguir como referência o modelo Open Archival Information System (OAIS) para PD de uso permanente.

Ao preservar objetos digitais no longo prazo, é necessário identificar primeiramente um modelo de PD que possa garantir não só a qualidade, mas também a autenticidade, a confiabilidade, a integridade e a segurança. Sendo assim, o modelo OAIS é referência como pré-requisito para o desenvolvimento de RDC (Arellano, 2004; Grácio, 2012; Flores; Pradebon; Cé, 2017; Santos; Flores, 2021; CONARq, 2022).

O modelo OAIS se destaca em ambiente tecnológico complexo por meio do RDC, cujo objetivo é preservar e manter a autenticidade e integridade dos objetos digitais no longo prazo, e garantir que os objetos digitais no futuro sejam acessados e interpretados pela comunidade interessada (Santos; Flores, 2021).

Nesse sentido, para uma instituição, a preservação digital tem sua importância social, pois, além de contemplar as demandas legais inerentes à legislação vigente, busca atender a missão de preservar a história e a memória institucional para o acesso a longo prazo.

2.3 Memória institucional

Antes de abordar a temática memória institucional, vale destacar o conceito atribuído por Felipe e Pinho (2018): "Memória é a faculdade que condiciona informações de caráter ligadas ao individual e ao coletivo". Costa (1997, p. 7) relata que:

[...] a memória é um elemento primordial no funcionamento das instituições. É através da memória que as instituições se reproduzem no seio da sociedade, retendo apenas as informações que interessam ao seu funcionamento. Há um processo seletivo que se desenvolve segundo regras instituídas e que variam de instituição para instituição. Tendo em vista que as instituições funcionam em rede no campo social, o limite de uma instituição é outra instituição.

A memória das instituições vista não necessariamente como instituição de memória, é relevante para a sociedade, pois ela está conectada à memória social e seu campo de pesquisa relacionado às áreas de conhecimento científico como a Ciência da Informação, Arquivologia, Comunicação, História, Educação e Administração (Felipe; Pinho, 2018). Parte-se do entendimento que, "[...] a Memória Institucional é uma memória que não é construída só internamente, ou seja, as informações provêm tanto de dentro quanto de fora da instituição" (Prado; Gracioso; Costa, 2019, p. 412). Essas definições enfatizam que as instituições necessitam entender que a memória é fundamental para a própria instituição, pois gera conhecimento sobre os registros de fatos e pessoas, evidenciando sua trajetória perante a sociedade.

Por outro lado, as funções administrativas de uma instituição se modernizaram, tendo o uso das TIC como aliadas na busca por eficiência e eficácia. Assim sendo, surgem documentos administrativos em meio digital para posterior cumprimento de seus valores primários e secundários, bem como no que tange a preservação de caráter permanente e, assim, integram o patrimônio documental (Santos; Flores, 2017). Esse processo tornar-se pertinente na maneira de realizar o quê preservar de fato e o quê eliminar, visando garantir a consistência

do que será preservado digitalmente para a constituição da memória institucional.

Grácio, Fadel e Valentim (2013) entendem que para a formação da memória institucional, a Instituição de Ensino Superior (IES) tem que dispor financeiramente de recursos permanentes para serem aplicados em infraestrutura, TIC e capacitação de pessoal, visando o desenvolvimento de atividades que envolvam a preservação digital e, portanto, devem ser vistos como um investimento permanente da IES para preservar sua produção administrativa, acadêmica, científica, artística e cultural, visto que por meio da preservação desses registros/documentos é constituída a história da instituição.

3 PROCEDIMENTOS METODOLÓGICOS

Em relação aos procedimentos metodológicos trata-se de uma pesquisa de natureza qualiquantitativa, do tipo descritiva e exploratória. Inicialmente foi realizada uma Revisão Sistemática de Literatura (RSL) em fontes de informação da área de Ciência da Informação sobre preservação digital no Brasil e, posteriormente, foram realizadas a análise bibliométrica e a análise categorial.

No que se refere à RSL, realizou-se buscas nas fontes de informação internacionais WoS e Scopus, e em âmbito nacional buscou-se na BRAPCI, uma vez que é considerada a maior base do campo científico da Ciência da Informação no País. A busca foi realizada por meio da aplicação dos descritores 'Digital preservation' OR 'Archival documents scanning' AND 'Brazil' nas bases WoS e Scopus; e 'Preservação digital' na BRAPCI.

Os tipos de documentos recuperados foram: artigos originais e de revisão; trabalhos publicados em anais de eventos científicos e capítulos de livro. No que se refere aos idiomas, os textos selecionados foram na língua inglesa nas bases internacionais (Scopus e WoS), e na língua portuguesa na base nacional (BRAPCI).

O recorte temporal cobriu o período de 2000 a 2022. Os filtros utilizados foram as aspas (") em todos os termos nas bases de conhecimento WoS e Scopus, no primeiro e segundo descritor e entre elas o operador booleano OR (OU), e no terceiro descritor o operador booleano AND (E). Enquanto na BRAPCI usou-se apenas o termo entre aspas 'Preservação digital'.

A organização dos dados foi realizada por meio da compilação das palavras-chave, títulos, autores e tipo de documento. Também foram identificados os autores mais produtivos e os termos mais usados por esses autores no período analisado.

4 RESULTADOS E DISCUSSÕES

As buscas realizadas nas bases de dados internacionais obtiveram os seguintes resultados: na WoS foram 1.057 (hum mil e cinquenta e sete) textos recuperados e na Scopus foram 20 (vinte) textos recuperados. Em relação à busca realizada na BRAPCI, recuperou-se 351 (trezentos e cinquenta e um) textos sobre a temática pesquisada. Do total de textos recuperados, selecionaram-se aqueles pertinentes e aderentes à pesquisa (Tabela 1).

Tabela 1 - Produções científicas indexadas nas Bases de dados Scopus, WoS e Brapci

Fontes de Informação	Qtde.	Tipos de documento	Período
WoS	52	46 artigos 6 trabalhos publicados em anais de eventos	2003 - 2022
Scopus	20	15 artigos 2 trabalhos publicados em anais de eventos 3 capítulos de livro	2003 - 2022
BRAPCI	252	250 artigos 2 capítulos de livro	2000 - 2022

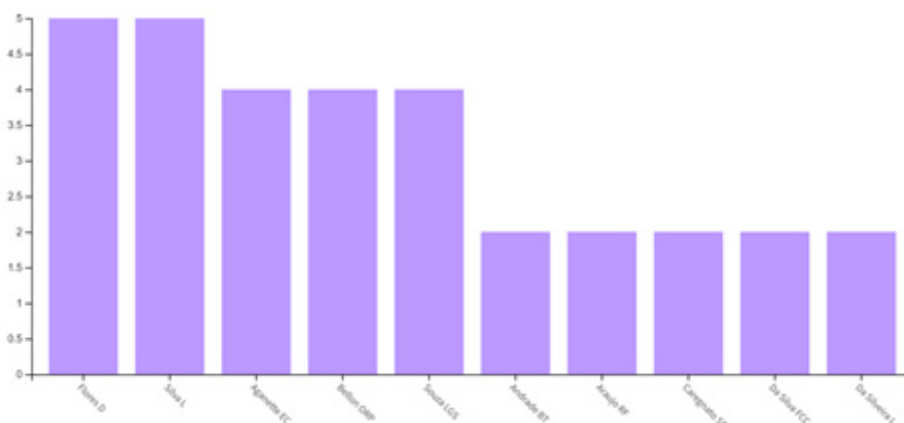
Total	324	311 artigos 8 trabalhos publicados em anais de eventos 5 capítulos de livro	2000 - 2022
--------------	------------	--	--------------------

Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Na Tabela 1, os critérios de exclusão e inclusão utilizados nas bases, ocorreram da seguinte maneira: na base de dados WoS foram descartados 1.005 (hum mil e cinco) textos científicos que não apresentavam alinhamento/aderência ao tema pesquisado, vale destacar que não foram identificados trabalhos produzidos por pesquisadores brasileiros como colaboradores ou produzidos no País, resultando na seleção de 52 (cinquenta e dois) textos científicos. No que tange a base de dados Scopus não se descartou nenhum texto.

Na BRAPCI foram selecionados somente artigos e capítulos de livros, sendo descartados 99 (noventa e nove) textos como, por exemplo: dossiês, pechas kucha e relatos de experiência, obtendo-se como resultado 252 (duzentos e cinquenta e dois) textos. A partir dos textos pertinentes selecionados, extraiu-se e analisaram-se informações conforme a pesquisa. O Gráfico 1 apresenta os autores mais produtivos no período pesquisado na base de dados WoS:

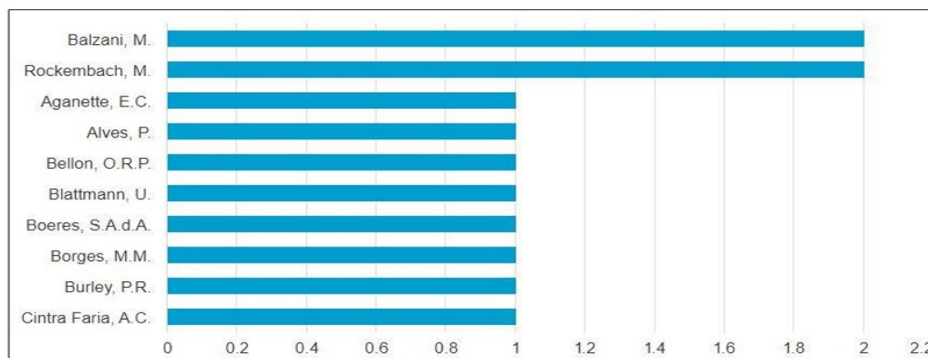
Gráfico 1 - Autores mais produtivos na base de dados WoS



Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Na base de dados WoS, os pesquisadores Daniel Flores e Luciano Silva foram identificados em 5 (cinco) textos cada, seguidos por Olga Bellon, Luciana Gonçalves Silva Souza e Elisângela Crisitina Aganette com 4 (quatro) textos respectivamente.

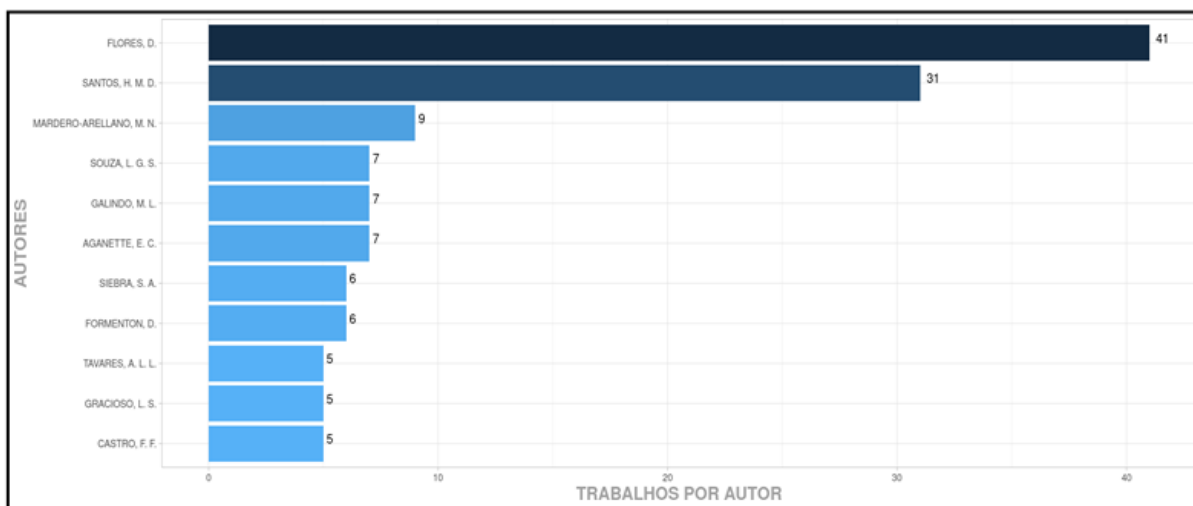
Gráfico 2 - Autores mais produtivos na base de dados Scopus



Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Na base de dados Scopus os autores mais produtivos identificados foram Balzani e Rockenbach, ambos com 2 (dois) textos, os demais com 1 (um) texto cada. duziram literatura sobre o assunto são Alemanha, Espanha e Estados Unidos. Destaca-se em quarto lugar os esforços de pesquisa oriundos do Reino Unido.

Gráfico 3 - Autores mais produtivos na base de dados BRAPCI.



Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

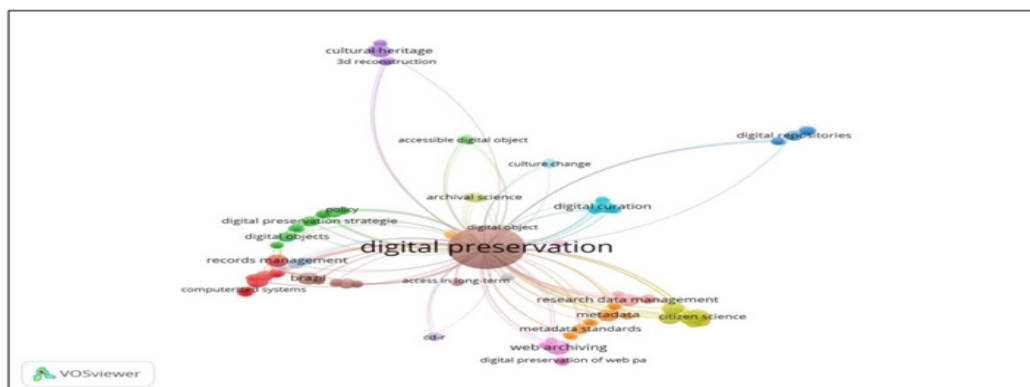
Na base de dados BRAPCI, o pesquisador Daniel Flores foi identificado com 41 (quarenta e um) textos, seguido de Henrique Machado dos Santos com 31 (trinta e um) textos, Miguel Ángel Mardero Arellano com 9 (nove) textos e Luciana Gonçalves Silva Souza com 7 (sete) textos respectivamente.

A partir da análise dos gráficos 1, 2 e 3, pode-se inferir que entre os autores mais produtivos identificados nas bases de dados WoS, Scopus e BRAPCI no período pesquisado foi o pesquisador Daniel Flores.

A partir da análise realizada por meio das palavras-chave atribuídas pelos autores, foi possível identificar as temáticas mais pesquisadas por eles sobre 'preservação digital', o que pode ser visualizado por meio da aplicação do software VOSviewer (Figuras 2, 3 e 4).

Na WoS foram identificadas 161 (cento e sessenta e uma) palavras-chave, evidenciando os temas que os autores mais abordaram (Figura 2).

Figura 2 - Temáticas abordadas pelos autores na WoS.

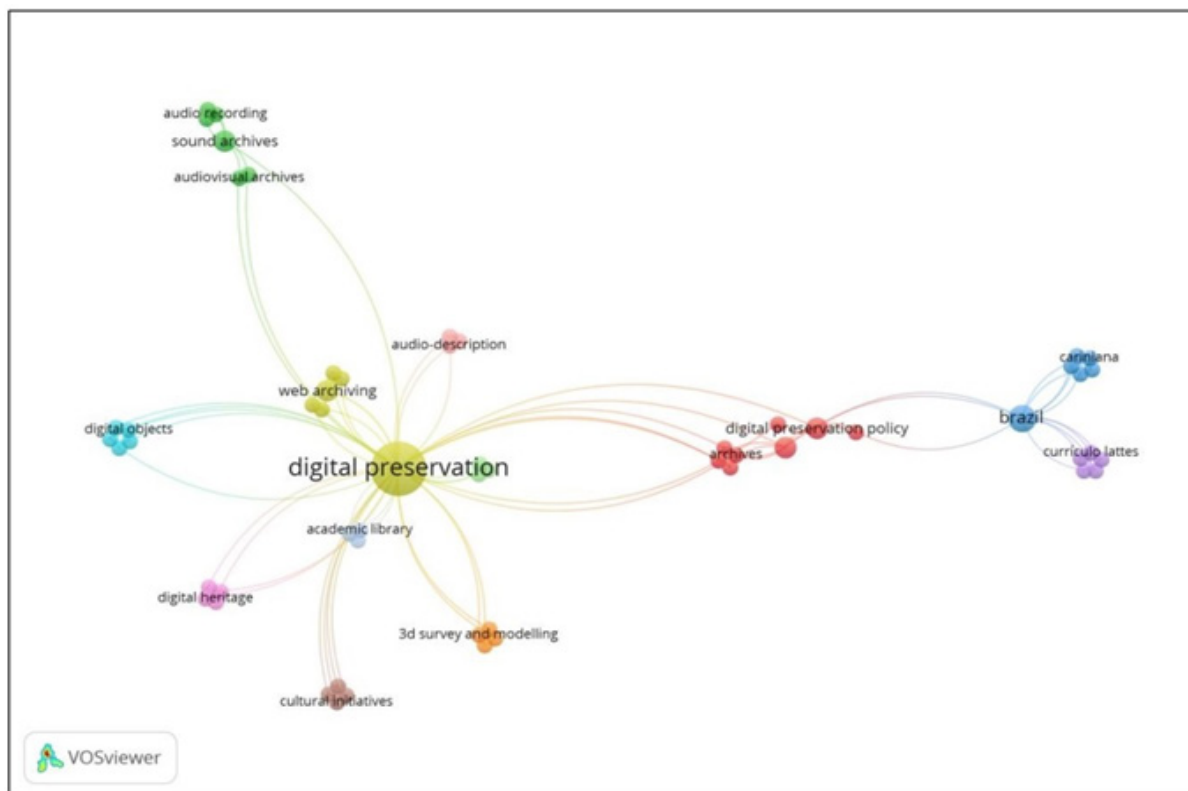


Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Na WoS os temas que mais predominaram e são associados aos autores mais produtivos são: (Metadata) Metadados (5); Archiving Web (Arquivamento na Web) (4); Digital curation (Curadoria digital) (3); Cultural heritage (Patrimônio cultural) (3); Research data management (Gestão de dados de pesquisa) (3); Digital repositories (Repositórios digitais) (2); Digital preservation policy (Política de preservação digital) (2); Information management (Gestão da informação) (2); e Memory (Memória) (2).

Na Scopus foram encontradas 72 (setenta e duas) palavras-chave, que possibilitou visualizar os temas mais estudados pelos pesquisadores (Figura 3).

Figura 3 - Temáticas abordadas pelos autores na Scopus

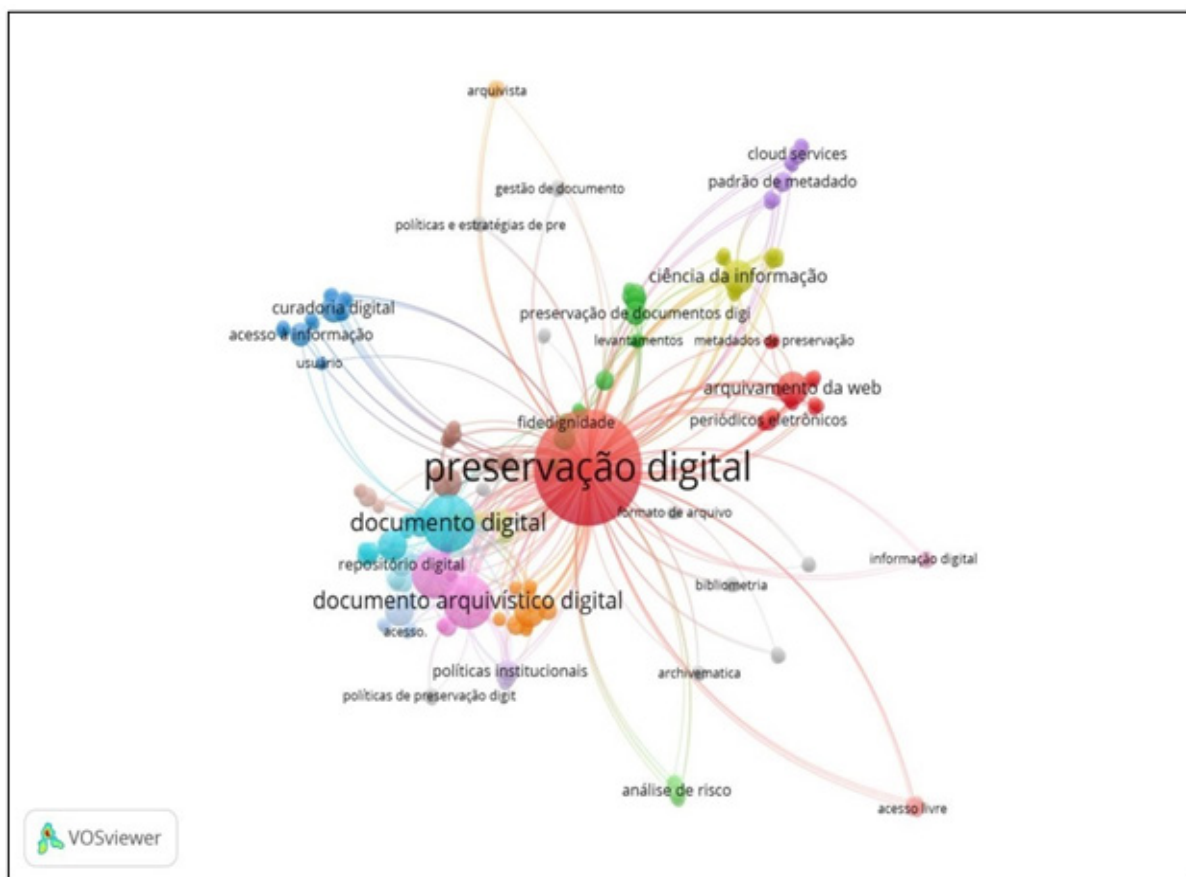


Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Na Scopus os temas que os autores mais pesquisaram foram: Archiving Web (Arquivamento da Web) (2); Digital preservation policy (Política de preservação digital) (2); e Audiovisual archives (Arquivos audiovisuais) (2).

Na BRAPCI identificou-se 192 (cento e noventa e duas) palavras-chave atribuídas pelos autores, cujos temas de interesse foram: Documento digital (17); Documento arquivístico digital (15); Arquivologia (12); Obsolescência tecnológica (6); Arquivamento da Web (6); Ciência da Informação (6); Repositórios digitais (5); Metadados (4); Curadoria digital (4); Memória (1) (Figura 4).

Figura 4 - Temáticas abordadas pelos autores na BRAPCI



Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

Análise categorial que propiciou evidenciar as temáticas produzidas sobre o tema ‘preservação digital’ pelos autores mais produtivos (Tabela 2).

Tabela 2 - Temáticas estudadas pelos autores mais produtivos no período pesquisado

Fontes de Informação / Temáticas	Scopus	WoS	BRAPCI	Total	%
Documento digital	0	1	17	18	16,7
Documentos digitais arquivísticos	0	1	15	16	14,8
Arquivologia	0	0	12	12	11,1
Arquivamento da Web	2	4	6	12	11,1
Repositórios digitais	2	2	5	9	8,3
Metadados	0	5	4	9	8,3
Ciência da Informação	1	1	6	8	7,4
Curadoria digital	0	3	4	7	6,5
Política de preservação digital	2	2	2	6	5,5
Memória	1	2	1	4	3,7
Gestão da informação	1	2	0	3	2,8
Total	9	23	76	108	100,0%

Fonte: Elaboração própria a partir dos dados da pesquisa (2023).

As temáticas mais estudadas pelos autores mais produtivos indicam o que tem sido relevante para a evolução do estado da arte do tema ‘preservação digital’, destaca-se: ‘documentos digitais arquivísticos’, ‘documento digital’ e ‘Arquivologia’

são as temáticas mais estudadas pelos autores Flores e Santos, e a maioria das produções científicas aparece em colaboração entre estes autores, que representam na soma de 46 (quarenta e seis) textos (42,6%) do total. Temas como 'arquivamento na Web' com 12 (doze) textos (11,1%) do total; 'repositórios digitais' 9 (nove) (8,3%); 'metadados' 9 (nove) (8,3%); 'Ciência da Informação' 8 (oito) (7,4%); e 'curadoria digital' 7 (sete) (6,5%) do total, configuram-se na média de interesse. Por outro lado, os temas que aparecem com menor ocorrência são: 'política de preservação digital' 6 (seis) (5,5%); 'memória' 4 (quatro) (3,7%); e 'gestão da informação' 3 (três) (2,8%) do total.

5 CONSIDERAÇÕES FINAIS

Os estudos sobre preservação digital ainda são recentes no Brasil, no entanto, por meio de vários estudos publicados em periódicos indexados na WoS, Scopus e BRAPCI demonstra um avanço, principalmente, no que tange ao 'documento digital' e 'documentos arquivísticos digitais' que, por sua vez, se constituem em ativos digitais significativos para a memória patrimonial de uma instituição/organização.

As temáticas 'documento digital' e 'documentos arquivísticos digitais' aparecem de modo significativo, pois juntas somam 34 (trinta e quatro) textos, cuja frequência é de 31,5%. Observaram-se poucos trabalhos relacionados às temáticas 'política de preservação digital', 'memória' e 'gestão da informação', representando 6 (seis) (5,5%), 4 (quatro) (3,7%) e 3 (três) (2,8%) textos respectivamente, e isso se manteve em todas as três bases de dados pesquisadas.

Em relação à análise sobre os temas estudados pelos autores mais produtivos, conclui-se que há diversificadas temáticas sobre o tema 'preservação digital' no Brasil, porém a ocorrência de determinados temas estão concentrados nas bases de dados como 'repositórios digitais', 'políticas de preservação digital' e 'arquivamento na Web', representando a somatória de 27 (vinte e sete) trabalhos, cuja frequência é de 24,9%.

Dos autores analisados em relação aos temas estudados por eles, evidenciou-se uma discussão acerca da diferença entre 'documento digital' em relação a 'documento arquivístico digital'. Um dos autores mais produtivos, qual seja, Flores explica a diferença em um de seus estudos, o Conselho Nacional de Arquivos (CONARq) também explica que documento digital se refere a informação registrada, codificada em dígitos binários, acessível e interpretável por meio de sistema computacional, enquanto, o documento arquivístico digital se refere ao documento digital reconhecido e tratado como um documento arquivístico (Kroth; Flores, 2018).

Outro aspecto relevante é relacionado ao tema 'preservação digital' no âmbito da Ciência da Informação, pois obteve significativa ocorrência nas três bases de dados pesquisadas.

Nessa perspectiva, evidencia-se que a temática 'preservação digital' continua em construção no Brasil, mas há temáticas inter-relacionadas que vêm sendo estudadas por pesquisadores brasileiros ao longo do tempo. Esta pesquisa demonstra que ainda há a necessidade de desenvolver estudos relacionados à memória, à gestão da informação e à política de preservação digital, pois se demonstraram tímidas, porém promissoras, visto que sem elas não há como fazer preservação digital eficiente e eficaz em uma instituição que valorize a informação de fato.

Dessa forma, com intuito de garantir à memória a gerações futuras com uma excelente gestão da informação nas instituições embasada pela política de preservação digital, é preciso que se promovam iniciativas que viabilizam essas finalidades, uma que é ligada a essas atmosferas no Brasil, é a Dríade da Rede Brasileira de Serviços de Preservação Digital (Cariniana) do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Este grupo vem desenvolvendo soluções tecnológicas de

pesquisas voltadas para a rede de serviços de preservação digital de documentos eletrônicos no Brasil e mantém parcerias com outras instituições nacionais e internacionais.

REFERÊNCIAS

BARBEDO, F.; CORUJO, L.; SANT'ANA, M. **Recomendações para a produção de planos de preservação digital**. Lisboa: DGLAB, 2019. E-book. Disponível em: https://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2019/08/Recomendacoes_PPD_v2.pdf. Acesso em: 31 agos. 2023.

CONARq - Conselho Nacional de Arquivos. **Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos e-Arq Brasil**. Versão 2 2021, 2022. 224 p. Disponível em: <https://www.gov.br/conarq/pt-br>. Acesso em: 28 maio 2023.

COSTA, I. T. M. **Memória institucional: a construção conceitual numa abordagem teórico- metodológica**. 1997. 169 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1997. Disponível em: <https://ridi.ibict.br/handle/123456789/686>. Acesso em 26 jun. 2023.

FELIPE, C. B. M.; PINHO, F. A. Fotografia como dispositivo da memória institucional. **Logeion: filosofia da informação**, v. 5, n. 1, p. 89-101, 2018. DOI: 10.21728/logcion.2018v5n1.p89-101. Disponível em: <https://brapci.inf.br/index.php/res/v/32941>. Acesso em: 26 jun. 2023.

FERREIRA, M. Introdução à preservação digital: conceitos, estratégias e actuais. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. **E-book**. Disponível em: <chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 26 maio 2023.

FLORES, D.; PRADEBON, D. S.; CÉ, G. Análise do conhecimento teórico-metodológico da preservação digital sob a ótica da oais, saai, iso 14721 e nbr 15472. **Brazilian Journal of Information Science**, v. 11, 2017. DOI: 10.36311/19811640.2017.v11n4.11.p73. Disponível em: <https://brapci.inf.br/index.php/res/basket>. Acesso em: 26 jun. 2023.

FORMENTON, D.; CASTRO, F. F.; GRACIOSO, L. S.; FURNIVAL, A. C. M.; SIMÕES, M. G. M.. Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios (Peru)**, n. 68, p. 82-95, 2017. Disponível em: <https://brapci.inf.br/index.php/res/v/70537>. Acesso em: 13 jun. 2023.

FORMENTON, D.; GRACIOSO, L. S.. Padrões de metadados no arquivamento da Web: recursos tecnológicos para a garantia da preservação digital de websites arquivados. **Revista Digital de Biblioteconomia & Ciência da Informação**, n. 22001, v. 20, 2022. Disponível em: <https://brapci.inf.br/index.php/res/v/170063>. Acesso em: 13 jun. 2023.

GRÁCIO, J. C. A. **Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior**. São Paulo: Cultura Acadêmica, 2012. Disponível em: <http://hdl.handle.net/11449/113727>. Acesso em: 19 jun. 2023.

GRÁCIO, J. C. A.; FADEL, B.; VALENTIM, M. L. P. Preservação digital nas instituições de ensino superior: aspectos organizacionais, legais e técnicos. **Perspectivas em Ciência da Informação**, v. 18, n. 3, p. 111-129, 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/38897>. Acesso em: 23 jun. 2023.

GRÁCIO, J. C. A. Preservação Digital nas Instituições de Ensino Superior: aspectos organizacionais, legais e técnicos. In: SEMINÁRIO GESTÃO DE DADOS DE PESQUISA, 1., 2020a. (Slides). Disponível em: <https://bibliotecas.unifesp.br/images/eventos/SGDP/JoseGracio.pdf>. Acesso em: 23 jun. 2023.

GRÁCIO, J. C. A.; TROITIÑO, S.; MADIO, T. C. de C.; BREGA, J. R. F.; MORAES, M. B. Modelo para elaboração de políticas de preservação-digital de documentos de arquivo por instituições de ensino superior: o caso da Unesp. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, [S. l.], v. 14, n. 3, 2020b. DOI: 10.29397/reciis.v14i3.2111. Disponível em: <https://www.reciis.icict.fiocruz.br/index>.

php/reciis/article/view/2111.
Acesso em: 08 set. 2023.

HILLMANN, D. **Using Dublin Core**. 2005. Disponível em: <https://www.dublin-core.org/specifications/dublin-core/usageguide/>. Acesso em: 08 set. 2023.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 14721**: Space data and information transfer systems: open archival information system – Reference model. Geneva: ISO, 2012.

KROTH, M. L.; FLORES, D. Autenticidade de documentos arquivísticos digitais: análise de um processo de afastamento. **Biblios** (Peru), n. 72, p. 67-79, 2018. DOI: 10.5195/biblios.2018.361. Acesso em: 20 jun. 2023.

LAVOIE, B.; GARTNER, R. Preservation metadata. 2nd edition. **Digital Preservation Coalition Technology Watch Report**. [s.l.], v. 13, n. 3, p. 1-36, 2013. Disponível em: www.dpconline.org/docs/dpc-technology-watch-publications/technology-watch-reports-1/894-dpctw13-03/file. Acesso em: 08 set. 2023.

MÁRDERO ARELLANO, M. Á.. Critérios para a preservação digital da informação científica. 2008. 356 f. Tese (Doutorado em Ciência da Informação)-Universidade de Brasília, Brasília, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/1518>. Acesso em: 25 ago. 2022.

MÁRDERO-ARELLANO, M. N. Preservação de documentos digitais. **Ciência da Informação**, v. 33, n. 2, 2004. DOI: 10.18225/ci.inf.v33i2.1043. Disponível em: <https://brapci.inf.br/index.php/res/v/20219>. Acesso em: 26 jun. 2023.

MONTEIRO, E. C. de S. de A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, [S. l.], v. 23, n. 53, p. 160-173, 2018. DOI: 10.5007/1518-2924.2018v23n53p160. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160>. Acesso em: 24 maio. 2023.

PRADO, S.; GRACIOSO, L. S.; COSTA, L. S. F. O papel da memória institucional para a gestão universitária: contribuições para a consolidação da umma na ufscar. **Informação & Informação**, v. 24, n. 3, p. 409-432, 2019. DOI: 10.5433/1981-8920.2019v24n3p409. Disponível em: <https://brapci.inf.br/index.php/res/v/134179>. Acesso em: 26 jun. 2023.

POLLAK, M. Memória, esquecimento, silêncio. **Estudos Históricos**, Rio de Janeiro, v. 2, n. 3, 1989, p. 3-15. Disponível em: https://moodle.ufsc.br/pluginfile.php/819737/mod_resource/content/1/POLLAK%2C%20Michael%20-%20Memória%2C%20esquecimento%2C%20silêncio.pdf. Acesso em: 05 maio 2023.

SANTOS, H. M. D.; FLORES, D. Gestão de objetos no repositório arquivístico digital confiável: um diálogo com a iso 16.363. **Páginas A&B**, Arquivos e Bibliotecas (Portugal), n. 13, p. 67-99, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/145242>. Acesso em: 13 jun. 2023.

SANTOS, H. M. D.; FLORES, D. Open archival information system: análise do modelo funcional no contexto da arquivística. **Revista Brasileira de Preservação Digital**, v. 2, 2021. DOI: 10.20396/rebpred.v2i00.15814. Disponível em: <https://brapci.inf.br/index.php/res/v/165877>. Acesso em: 26 jun. 2023.

SANTOS, H. M.; FLORES, D. Preservação do patrimônio documental arquivístico em ambiente digital. **Palavra Clave (La Plata)**, La Plata, v. 7, n. 1, p. e029, 2017. DOI: 10.24215/18539912e029. Disponível em: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCe029>. Acesso em: 26 jun. 2023.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, [S. l.], v. 15, n. 30, p. 1-31, 2010. DOI: 10.5007/1518-2924.2010v15n30p1. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>. Acesso em: 19 jun. 2023.

VALENTIM, M. L. P. Análise de conteúdo. In: VALENTIM, M. L. P. (org.). **Métodos qualitativos de pesquisa em Ciência da Informação**. São Paulo: Polis, p. 119-134, 2005.