

REPOSITÓRIOS DIGITAIS CONFIÁVEIS: uma revisão da literatura nacional e internacional publicada em periódicos científicos

TRUSTED DIGITAL REPOSITORIES: a review of national and international literature published in scientific journals

Luciana Gonçalves Silva Souza¹
Elisângela Cristina Aganette²

RESUMO

Repositórios digitais confiáveis são bibliotecas digitais que fornecem acesso às informações e, ao mesmo tempo, garantem autenticidade, confidencialidade e preservação por longo prazo dos documentos abrigados. Atualmente, existem critérios definidos por normas internacionais quanto à certificação desses repositórios. Em alguns países da América do Norte e Europa instituições se dedicam a auxiliar na aquisição do selo de certificação, construindo padrões e requisitos para a análise dos repositórios. Assim, o presente artigo discute o conceito de repositórios digitais confiáveis e sua exploração pela literatura nacional e internacional. Foi realizada uma pesquisa qualitativa, utilizando como método revisão sistemática em cinco bases de dados da área de Ciência da Informação, com o objetivo de identificar a aplicação dos estudos no cenário nacional e internacional e mapear os autores, periódicos e países que têm explorado o conceito e de que forma o tema tem sido trabalhado. Os resultados apontam que o tema é recente, com produções realizadas a partir de 2001. Há a predominância de citações em documentos que tratam de aspectos conceituais e da implementação da preservação digital. Pode-se afirmar que ainda é uma temática pouco explorada, pois, apesar de apontada por um grande quantitativo de autores, não há predominância de especialistas na publicação de documentos sobre o assunto. Conforme a filiação dos autores, os países que mais publicaram foram os Estados Unidos da América, Canadá, Brasil, Portugal e Inglaterra, mas houve a representação de países de todos os continentes do globo.

Palavras-chave: Repositórios digitais confiáveis. Preservação digital. Certificação de repositórios digitais. Confiabilidade da informação.

ABSTRACT

Trusted digital repositories are digital libraries that provide access to information and, at the same time, guarantee authenticity, confidentiality, and long-term preservation of sheltered documents. Currently, there are criteria defined by international standards regarding the certification of these repositories. In some countries in North America and Europe institutions are dedicated to assisting in the acquisition of the certification seal, building standards and requirements for the analysis of repositories. Thus, the present article discusses the concept of trusted digital repositories and their exploration by the national and international literature. A qualitative research was carried out, using as a method the systematic review in five databases of the Information Science area, with the objective of identifying the application of the studies in the national and international scenario and to map the authors, periodicals and countries that have explored the concept and how the theme has been worked on. The results point out that the theme is recent, with productions made from 2001. There is a predominance of citations in documents that deal with conceptual aspects and the implementation of digital preservation. It can be affirmed that it is still a little explored subject, because, although it is pointed out by a large number of authors, there is no predominance of specialists in the publication of documents on the subject. According to the authors' affiliation, the countries that most published were the United States of America, Canada, Brazil, Portugal and England, but there was representation of countries from all continents of the globe.

Keywords: Trusted digital repositories. Digital preservation. Certification of digital repositories. Reliability of information.

Artigo submetido em 15/04/2019 e aceito para publicação em 30/01/2020

1 Doutoranda em Gestão & Organização do Conhecimento. Universidade Federal de Minas Gerais, Brasil. ORCID <https://orcid.org/0000-0002-7532-6289>. E-mail: lucianags9@gmail.com

2 Professora Adjunta na Universidade Federal de Minas Gerais, Brasil. ORCID <http://orcid.org/0000-0003-4357-8016>. E-mail: elisange-laaganette@gmail.com

1 INTRODUÇÃO

A capacidade de registro, guarda e preservação das informações³ produzidas pela sociedade contemporânea foi ampliada em grande maneira por meio do suporte das Tecnologias de Informação e Comunicação (TICs). Contudo, manter e preservar registros confiáveis é uma reflexão ainda incipiente e pouco aplicada em bibliotecas e repositórios, principalmente quando se trata de recursos digitais, nos quais a informação é armazenada, mas é facilmente adulterada ou eliminada, sem possibilidade de recuperação (OTTO, 2014; MIQUEL TÉRMENS; DAVID LEIJA, 2017).

Estudos apontados por Otto (2014) sobre a preservação de conteúdos digitais em bibliotecas apresentam as dificuldades enfrentadas, concluindo que grande parte das organizações não realiza a prática ou possui limitações relativas aos métodos, processos, estratégias, sistemas ou tecnologias empregadas. Poucas instituições desenvolvem e aplicam políticas de preservação. Profissionais bibliotecários demonstraram insegurança quanto à adoção dos procedimentos e pouca conscientização sobre as consequências da ausência da preservação dos registros. O custo também foi apontado como um fator dificultador da realização do processo, já que a preservação por longo prazo implica em manutenção de procedimentos durante o período de guarda.

Nesse sentido, estratégias de preservação de documentos digitais vêm sendo estudadas com o propósito de desenvolver tecnologias que garantam a autenticidade e a confiabilidade das informações. Os repositórios digitais confiáveis são estruturados como plataformas que oferecem tais características, proporcionando a preservação por longo prazo dos recursos abrigados e segurança na disponibilização dos conteúdos (MAEMURA; MOLES; BECKER, 2017).

Assim, o conceito de repositórios digitais confiáveis inclui em sua definição a preservação digital, a confiabilidade e a autenticidade das informações, conforme documento referencial do assunto, o relatório produzido pelo Research Libraries Group (RLG) e pelo Online Computer Library Center (OCLC) intitulado *Trusted Digital Repositories: Attributes and Responsibilities* (2002). Por se tratar de uma temática ainda pouco explorada, observou-se a pouca literatura e a real necessidade de se aprofundar na referida temática. Diante disso, o presente artigo tem por objetivo identificar como tem sido tratada a nível nacional e internacional, visando sistematizar o assunto e possibilitar uma visão mais abrangente de sua aplicação.

3 Trabalharemos, neste artigo, com o conceito de informação abordado por Le Coadic (1996), onde o mesmo define informação como um conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica).

2 PRESERVAÇÃO DIGITAL

A preservação é um conceito que inspirou a afirmação de Conway (2001, p. 14) “Preservação é a aquisição, organização e distribuição de recursos a fim de que venham a impedir posterior deterioração ou renovar a possibilidade de utilização de um seletivo grupo de materiais”. Abrange a execução de técnicas que visam manter conteúdos acessíveis por determinados períodos de tempo. No contexto de unidades de informação, como bibliotecas, arquivos e museus, intensificou-se a ideia de preservação dos materiais abrigados tendo em vista garantias de evidências, manutenção de fatos históricos, registros de ações, pensamentos e conquistas.

A preservação aplicada ao ambiente digital trabalha o mesmo princípio da expectativa de manter conteúdos para acesso futuro. Contudo, a tecnologia inclui preocupações relativas à obsolescência de suportes, facilidade de alteração, dificuldade de identificação da originalidade. Conforme aponta Conway (2001, p. 12) “O universo digital transforma os conceitos da preservação tradicional: em vez de garantir a integridade física do objeto, passa a especificar a geração e a manutenção do objeto cuja integridade intelectual é sua característica principal”.

Assim, a preservação digital compreende processos de garantia da continuidade de materiais por períodos previamente estabelecidos, e a acessibilidade ao conteúdo é o propósito desta ação. Estratégias de preservação devem abranger acesso a informações autênticas (UNESCO, 2003).

Arellano (2004, p. 16) esclarece que, com o decorrer das ações de preservação digital houve alteração na prática desenvolvida: inicialmente buscava-se “garantir a longevidade dos arquivos, mas essa preocupação agora está centralizada na ausência de conhecimento sobre as estratégias de preservação digital e o que isso poderá significar na necessidade de garantir a longevidade dos arquivos digitais”.

Em vista disso, instituições detentoras de objetos digitais devem estar atentas ao trabalho de preservação de seus conteúdos. Conforme Grácio, Fadel e Valentim (2013, p. 113) o conceito está vinculado ao “processo de gestão organizacional que abrange várias atividades necessárias para garantir que um objeto digital possa ser acessado, recuperado e utilizado no futuro, a partir das TICs existentes na época e com garantias de autenticidade”.

Nesse sentido, a confiabilidade e a autenticidade podem ser conferidas aos documentos abrigados com a adoção de *softwares* de gestão documental como os repositórios digitais. No entanto, deve ser feito um estudo das funcionalidades oferecidas pelo *software* e a verificação da manutenção

da cadeia de custódia documental e preservação por longo prazo dos documentos abrigados. Aos repositórios que garantem essa autenticidade e acesso duradouro atribui-se o nome de repositórios digitais confiáveis (CCSDS, 2011).

3 REPOSITÓRIOS DIGITAIS CONFIÁVEIS

O conceito de confiabilidade atribuído aos repositórios digitais pode ser estabelecido pela definição da palavra confiança, que, conforme o Dicionário Houaiss (2009), é a “crença de que algo não falhará, de que é bem-feito ou forte o suficiente para cumprir sua função”. Assim, no contexto dos repositórios digitais, que possuem a função de “administrar, armazenar e preservar conteúdos informacionais em formato eletrônico”, garantindo, muitas vezes, acesso universal e gratuito a esses conteúdos (PINHEIRO; FERREZ, 2014, p. 195), a confiança está vinculada à garantia de que as informações armazenadas são íntegras e autênticas, refletem o conteúdo criado pelos autores que depositaram a confiança de manter seus dados em uma base de dados para a recuperação por vários usuários.

No contexto atual os repositórios digitais estão, em sua grande maioria, vinculados às instituições gestoras de informações como bibliotecas e arquivos. Reflexões a respeito da confiança depositada em instituições como bibliotecas, arquivos e museus são levantadas por Thomaz (2007), que indica que ao longo do tempo foi estabelecida a ideia de que a essas instituições podem ser creditadas a guarda e a disponibilização de objetos valiosos em termos culturais, emocionais e financeiros, pois essas organizações transmitem a segurança de que garantirão a preservação dos itens abrigados para acesso e conhecimento das futuras gerações.

No entanto, quando se trata de conteúdos digitais, a confiabilidade dos dados enfrenta ameaças que são mais difíceis de combater e identificar que as de documentos físicos. Posicionando-se quanto aos documentos arquivísticos digitais, o Conselho Nacional de Arquivos afirma:

Os documentos arquivísticos digitais apresentam dificuldades adicionais para presunção de autenticidade em razão de serem facilmente duplicados, distribuídos, renomeados, reformatados ou convertidos, além de poderem ser alterados e falsificados com facilidade, sem deixar rastros aparentes (CONSELHO NACIONAL DE ARQUIVOS, 2012, p. 1).

Diante disso, as instituições responsáveis por gerenciar documentos digitais enfrentam desafios ao tentarem manter seus conteúdos autênticos, como apontam Santos e Flores (2015, p. 198) “A complexidade e a especificidade do documento arquivístico digital implica na necessidade de um tratamento diferenciado se comparado aos documentos tradicionais em suportes analógicos”.

Assim, a necessidade de reconhecer e manter organizações que garantam confiança de documentos digitais, explorada desde final dos anos de 1990, refletiu na constituição do *Open Archival Information System* (OAIS), modelo de referência do Sistema de Informações de Arquivamento Aberto. Tal modelo traz reflexões sobre o que seria necessário para manter a preservação por longo prazo das informações digitais e se constituiu na norma ISO 14721, que apresenta um modelo de referência para os sistemas de preservação de materiais digitais.

No contexto nacional, a referida norma foi publicada pela Associação Brasileira de Normas Técnicas como a NBR 15472, baseada também na Recomendação do Comitê Consultivo para Sistemas de Dados Espaciais (CCSDS) 650.0-B-1:2002 (2011). O modelo de referência para um Sistema Aberto de Arquivamento de Informação (SAAI) orienta quanto à preservação de informações para acesso e uso ressaltando a necessidade de compor políticas e procedimentos nas instituições que abrigam conteúdos digitais. Esclarece quais as necessidades de constituição de um sistema de preservação digital e sua manutenção, descrevendo o ambiente e os requisitos funcionais.

Desse modo conforme cita a recomendação 650.0-B-1:2002, mesmo antes da aprovação do padrão OAIS, que ocorreu em 2002, as organizações que se enquadravam no modelo o utilizavam como argumento para a confiabilidade de seus repositórios digitais. Assim, percebeu-se a necessidade de definir critérios detalhados para a constituição de Repositórios Digitais Confiáveis. O *Research Libraries Group* (RLG) e o *Online Computer Library Center* (OCLC) desenvolveram o documento *Trusted Digital Repositories: Attributes and Responsibilities* (2002), que teve como objetivo consolidar atributos e responsabilidades para os repositórios digitais que necessitavam trabalhar a preservação por longo prazo, desenvolvendo a confiabilidade de seu sistema. Discute-se ainda, estruturas potenciais para a certificação dos repositórios, recomendando programas de certificação e definição de critérios auditáveis para os repositórios.

Dando prosseguimento ao recomendado, o RLG, juntamente à entidade americana *National Archives and Records Administration* (NARA) buscou desenvolver os aspectos da certificação digital dos repositórios. Elaboraram um conjunto de critérios de certificação considerando várias tipologias de repositórios, que denominaram *Trustworthy Repositories Audit and Certification Checklist* (TRAC). Essa lista de verificação possibilita a avaliação objetiva das responsabilidades e dos riscos potenciais relacionados aos conteúdos digitais abrigados em repositórios, arquivos e provedores de conteúdos.

A partir do TRAC desenvolveu-se o padrão ISO, instituído pela norma ISO 16363 “Space data and information transfer systems – Audit and certification of trustworthy digital repositories” publicada em 2012 e revista em 2017. É utilizada como base para a certificação de repositórios ao definir os critérios de avaliação da confiabilidade aplicada aos repositórios digitais. Assim, a norma ISO 13636 é destinada à

aplicação em instituições que desejam mensurar a confiabilidade de seus repositórios ou que trabalham com auditorias nesse sentido. Também é instrumento para a reformulação de repositórios digitais.

Conforme o documento *Trusted Digital Repositories: Attributes and Responsibilities* um repositório digital confiável pode ser definido como “aquele cuja missão é fornecer acesso confiável e de longo prazo a recursos digitais para sua comunidade usuária, no presente e no futuro”. O texto define aspectos que considera essenciais aos repositórios digitais confiáveis: i) garantir o acesso a longo prazo aos recursos digitais abrigados; ii) manter um sistema que garanta a qualidade da informação da qual é responsável; iii) ter sustentabilidade financeira e ser viável para a organização; iv) construir o sistema com base em padrões estabelecidos garantindo o gerenciamento, acesso e segurança dos documentos; v) desenvolver metodologias de avaliação da confiabilidade pela comunidade usuária; vi) ser responsável pelas informações contidas e explicitar abertamente essa responsabilidade para os depositantes e usuários; vii) manter políticas e critérios de desempenho passíveis de auditoria e medição.

Térmens e Leija (2017) pontuam que o estudo de repositórios confiáveis é uma das linhas de pesquisa em preservação digital mais trabalhadas em nível internacional. Busca definir metodologias e ferramentas de avaliação de conformidade com padrões e boas práticas definidas. Os sistemas de auditoria podem ser considerados um dos principais métodos de definição de repositórios confiáveis, pois contam com a avaliação de especialistas que definem a segurança e confiabilidade dos sistemas informacionais.

Nessa perspectiva, buscou-se compreender como a literatura nacional e internacional tem trabalhado a aplicação da preservação em repositórios digitais avaliando, dessa forma, o uso do conceito de repositórios digitais confiáveis. Foram realizadas pesquisas nas seguintes bases de dados: Base de Dados em Ciência da Informação (BRAPCI), Education Resources Information Center (ERIC), Information Science & Technology Abstracts (ISTA), Library and Information Science Abstracts (LISA) e Library, Information Science & Technology Abstracts with Full Text (LISTA), como descrito na seguinte seção.

4 PROCEDIMENTOS METODOLÓGICOS

A presente investigação utilizou como técnica elementos da revisão sistemática, pois trata-se de uma estratégia que ressalta a representação da literatura com relação ao assunto, de maneira abrangente e imparcial. O procedimento metodológico aplicado compreende um processo de seleção de referências utilizando critérios explícitos e replicáveis, com o objetivo de minimizar o risco de viés, atribuindo validade científica à discussão produzida (EVANS; PEARSON, 2001).

A abordagem utilizada foi a qualitativa, de caráter exploratório e descritivo, buscando avaliar como a temática está sendo tratada em publicações científicas indexadas em bases de dados da área de ciência da informação nacionais e internacionais.

Assim, a construção da análise proposta se deu a partir de etapas que consistiram na condução das expressões de buscas, definição das bases de dados a serem exploradas e análises quantitativas e qualitativas, conforme descrições a seguir.

4.1 Expressões de busca

Para identificação das expressões de busca utilizadas na revisão de literatura procurou-se conceitos relacionados pelas seguintes fontes:

- a) Artigos de periódicos que abordem o tema preservação em repositórios digitais;
- b) Classificação Decimal de Dewey (CDD);
- c) Classificação Decimal Universal (CDU);
- d) Lista de cabeçalho de assunto da Biblioteca Nacional brasileira;
- e) Lista de cabeçalho de assunto da Library of Congress;
- f) Tesouro da base de dados LISA;
- g) Dicionário de Terminologia Arquivística (BRASIL, 2005);
- h) Dicionário de Biblioteconomia e Arquivologia (SAYÃO; SALES, 2012);
- i) Tesouro Brasileiro de Ciência da Informação (PINHEIRO; FERREZ, 2014).

Identificou-se 10 expressões, nas línguas portuguesa, espanhola e inglesa, que integram, de maneira mais restrita, a aplicação da preservação por longo prazo em repositórios digitais. Assim, a expressão “repositórios digitais confiáveis”, por remeter aos conceitos de preservação e de confiabilidade da informação armazenada em repositórios, reconhecendo o repositório digital confiável como o software que integra esses princípios, preservando a informação armazenada, além de disponibilizá-la, foi a única utilizada na busca, considerando as suas variações nas línguas inglesa e espanhola. Dessa maneira, utilizou-se as seguintes expressões na busca:

- a) Repositórios digitais confiáveis;
- b) Repositório digital confiável;
- c) Repositorio digital de confianza;

- d) Repositorios digitales de confianza;
- e) Trusted digital repository;
- f) Trusted digital repositories;
- g) Trustworthy repository;
- h) Trustworthy repositories;
- i) Trustworthy digital repository;
- j) Trustworthy digital repositories.

4.2 Seleção dos documentos

A pesquisa foi realizada durante o primeiro semestre de 2019 utilizando-se o Portal de Periódicos Capes como base de referência para acesso às bases de dados escolhidas. O Portal reúne e oferece às Instituições de Ensino Superior públicas brasileiras acesso a 130 bases de dados referenciais com publicações científicas mundiais, incentivando o acesso e a divulgação científica para fortalecimento da pós-graduação no país (PORTAL DE PERIÓDICOS CAPES/MEC, [201-]).

4.2.1 Bases de dados escolhidas

Foram escolhidas cinco bases de dados da área de Ciência da Informação, disponíveis online, de acesso aberto ou disponibilizadas pelo Portal de Periódicos da Capes. A seleção foi feita com base na representatividade das bases para a área de Ciência da Informação, em que a presente pesquisa se situa. São elas: BRAPCI, ERIC, ISTA, LISA, LISTA.

4.2.2 Critérios de elegibilidade

A busca foi realizada nos campos título, palavras-chave, resumo e texto completo. Para o filtro relacionado ao texto completo, as bases de dados que apresentam a opção de busca avançada das expressões normalmente são as que se enquadram na tipologia de bases de dados de fonte ou textuais, que possuem acesso ao texto completo dos documentos.

Considerou-se a tipologia documental “artigo”, incluindo a opção de seleção de revistas acadêmicas analisadas por especialistas. Somente a base de dados BRAPCI não apresenta esta forma

de busca, no entanto a base indica incluir apenas periódicos científicos em sua coleção. Os idiomas selecionados para a análise serão o espanhol, inglês e português.

A recuperação inicial retornou o total de 277 documentos, conforme demonstra a Tabela 1, sendo 234 documentos distintos e 43 repetidos. Foram retirados da análise notas de editores, notícias, resenhas, papers e apresentação de documentos.

Tabela 1 – Quantitativo de documentos recuperados por base de dados selecionada

Base de dados	Documentos
LISTA	158
LISA	100
BRAPCI	11
ISTA	6
ERIC	2

Fonte: Dados da pesquisa (2019).

Após avaliação inicial da amostra foram identificados 18 documentos que ainda se enquadravam em tipologias diferentes de artigos científicos, que foi a escolhida para integrar a presente pesquisa, apesar das bases de dados definirem como tal. Também durante o processo de análise de conteúdo, que será descrito na seção seguinte, foram identificados 11 documentos que foram recuperados, mas não possuem relação com o tema definido nos critérios da análise. Esses documentos foram retirados da seleção e a amostra trabalhada nas análises foi de 205 artigos.

4.3 Análise quantitativa e qualitativa

Buscou-se, neste trabalho, explorar alguns aspectos que demonstram como o tema está sendo trabalhado pelos pesquisadores a nível mundial. Dados quantitativos relativos ao número de periódicos selecionados, ano de publicação, quantidade de autores, quantidade de obras publicadas por autor e filiação dos mesmos indicam a predominância da exploração do assunto ao se considerar os aspectos geográfico e temporal.

Além disso, com o objetivo de identificar os principais assuntos abordados pelos documentos recuperados foi realizada a análise do conteúdo baseada em Bardin (2011, p. 37). Segundo a autora, a análise de conteúdo é “um conjunto de técnicas de análises das comunicações”, sendo que todas as comunicações ditas ou registradas podem ser submetidas à análise. A partir da técnica de classificação

dos elementos de Bardin, foram definidos critérios de classificação que permitiram a definição de categorias que representam a totalidade dos textos explorados e permitem identificar os principais assuntos dos documentos. Considerando que a busca retornou trabalhos que continham a expressão “repositórios digitais confiáveis” e suas variações, como exposto no item 3.1, buscou-se identificar quais aspectos têm sido abordados pela literatura.

A análise se deu por meio da leitura sistematizada das seguintes seções dos documentos: resumo, introdução, conclusão / considerações finais. A esse respeito, Lancaster (2003) esclarece que, quando o indexador se depara com grande quantitativo de documentos para a extração de conteúdos, deve focar a atenção nas seções que representam o máximo de informações em um menor tópico: título, resumo, sinopse, introdução e conclusões, além de títulos de seções, palavras sublinhadas e legendas de ilustrações. O restante do texto deve ser considerado a partir de uma visão rápida para que o termo atribuído reflita todo o conteúdo.

5 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados obtidos a partir da análise dos documentos recuperados, considerando as características das produções, dos autores e o conteúdo abordado.

5.1 Caracterização dos documentos

Foram identificados 205 documentos distribuídos em 80 periódicos, sendo que as publicações que apresentaram maior número de artigos publicados com a temática foram o IASSIST Quarterly e o Library Trends, com 11 artigos. O IASSIST Quarterly é editado pela International Association for Social Science Information Services and Technology – Associação Internacional para Serviços e Tecnologia de Informação em Ciências Sociais (IASSIST). É uma associação que reúne profissionais que atuam com tecnologia da informação e serviços de dados para apoiar a pesquisa e o ensino em ciências sociais. Já o periódico Library Trends é editado pela Johns Hopkins University Press e está voltado aos profissionais bibliotecários e educadores, por apresentar conteúdos que abordam temáticas relativas às tecnologias, críticas e tendências de bibliotecas e unidades informacionais.

Com relação ao idioma, as expressões de busca e os critérios da pesquisa direcionaram quais línguas seriam trabalhadas. Entretanto, houve uma prevalência de obras em inglês, com 182 documentos. Dezoito artigos estão publicados em língua portuguesa e cinco são escritos em espanhol, conforme Tabela 2.

Tabela 2 - Idioma das publicações analisadas

Idioma	Quantitativo
Inglês	182
Português	18
Espanhol	5

Fonte: Dados da pesquisa (2019).

Quanto ao período de publicação do assunto, não foi estipulada uma data de corte, pois o objetivo era verificar quando o tema começou a ser explorado no cenário mundial. Assim, identificou-se artigos publicados a partir de 2001 até o ano de 2018, marco da finalização da pesquisa. Ressalta-se que o maior quantitativo publicado foi em 2015, com 31 artigos e os anos que tiveram menor número de publicações foram 2001 e 2002, com um artigo recuperado que menciona o tema. A Tabela 3 ilustra o quantitativo por ano de produção.

Tabela 3 – Total de publicações por ano de produção

Data	Quantitativo	Data	Quantitativo
2001	1	2010	10
2002	1	2011	17
2003	2	2012	18
2004	2	2013	9
2005	4	2014	22
2006	6	2015	31
2007	14	2016	17
2008	7	2017	23
2009	8	2018	13

Fonte: Dados da pesquisa (2019).

5.2 Caracterização dos autores

Os documentos analisados foram escritos por 368 autores. Infere-se, a partir dos dados levantados, que esse é um tema ainda difuso entre os autores, e, devido a sua especificidade, ainda pouco explorado.

Quanto aos referidos autores, tratam-se de profissionais que atuam em 38 países, sendo que o maior quantitativo, 135 autores, são dos Estados Unidos da América. O segundo país

que possui o maior quantitativo de autores interessados na presente temática é o Canadá, 39 autores, seguido pelo Brasil, com 20 autores, 16 em Portugal e 15 na Inglaterra. Os demais países apontaram um total de menos de 15 pesquisadores. No entanto, ainda assim, pode-se afirmar que o assunto vem sendo explorado em várias localidades geográficas, como demonstra a Tabela 4.

Tabela 4 – Nacionalidade dos autores dos documentos recuperados

País	Nº autores	País	Nº autores
EUA	135	Taiwan	5
Canadá	39	Bélgica	3
Brasil	20	Suíça	3
		Coreia do Sul	2
Portugal	16	França	2
Inglaterra	15	Hungria	2
Holanda	12	Lituânia	2
Irlanda	12	México	2
Reino Unido	12	Suécia	2
África do Sul	11	Botsuana	1
Alemanha	10	Cuba	1
Itália	10	Filipinas	1
Áustria	7	Gana	1
China	6	Grécia	1
Espanha	6	Irã	1
Austrália	5	Jordânia	1
Escócia	5	Kênia	1
Finlândia	5	Noruega	1
Índia	5	País de Gales	1
Nova Zelândia	5		

Fonte: Dados da pesquisa (2019).

Com relação ao quantitativo de documentos publicados por autor, aspecto que pode indicar especialistas que se dedicam ao assunto, identificou-se que um autor publicou seis artigos, três autores publicaram quatro artigos, sete autores publicaram três artigos, 26 autores publicaram dois documentos sobre o assunto e os demais, 331 autores, publicaram apenas um artigo sobre a temática.

Tabela 5 – Quantidade de publicações sobre a temática por autor

Quantitativo/Artigo	Quantitativo/Autor
6	1
4	3
3	7
2	26
1	331
Total	368

Fonte: Dados da pesquisa (2019).

Os dados indicam que a produção sobre o tema ainda está dispersa, contudo, por se tratar de um aspecto específico da preservação digital, acredita-se que a infreqüência de desenvolvimento do assunto pelos autores não indica a ausência de literatura sobre a grande área, preservação digital, mas sim sobre a definição dos repositórios confiáveis. Assim, os autores que publicaram sobre o assunto, especificamente aqueles que produziram até três artigos sobre a temática, produziram outros conteúdos sobre preservação digital, curadoria digital e preservação por longo prazo de documentos digitais, conforme pesquisa realizada pelos pesquisadores em momento posterior ao resultado acima encontrado.

Com o objetivo de buscar relação entre o interesse dos quatro autores que mais publicaram sobre o tema e o espaço temporal dessas publicações, foram identificadas as seguintes datas:

- O autor que produziu seis documentos realizou as publicações nos anos de 2004, 2006, 2007, 2011, 2013 e 2016;
- o autor A, que possui quatro artigos, realizou as publicações em 2008, 2012 e dois documentos em 2016;
- os autores B e C que também publicaram quatro documentos estiveram sempre em coautoria como orientador e orientando, apresentando dois artigos em 2015 e um documento em cada ano seguinte: 2016 e 2017.

Dessa forma, pode-se inferir pelos dados que, entre esses quatro autores, dois estudam o tema há cerca de dez anos, demonstrando interesse pela área e dois realizaram pesquisas durante o período de realização de pós-graduação realizada recentemente.

5.3 Análise do conteúdo das publicações

Tendo em vista que os documentos trabalhados no presente artigo possuem correspondência com a temática da preservação digital, pois citam a expressão “repositórios digitais confiáveis” e suas variações, o presente estudo buscou também verificar quais os principais assuntos abordados, uma vez que tal aspecto revela as preocupações dos pesquisadores e as áreas específicas possivelmente ainda pouco exploradas pela literatura.

A organização dos trabalhos por temas deu-se a partir da extração de categorias dos próprios conteúdos. Em alguns casos os documentos se enquadram em mais de uma categoria, no entanto, verificou-se e optou-se por aquela que melhor representa o assunto principal. A classificação não aconteceu de maneira rígida e procurou-se discriminar um número mais extenso de categorias que representam os conteúdos. Foram identificadas quatro categorias que se desdobram em 14 subcategorias:

Tabela 6 – Categorização dos documentos recuperados por assunto

Categorias	Subcategorias	Quantidade	Percentual
Preservação digital	Aspectos conceituais	38	47%
	Planejamento e implantação de ações	35	
	Metadados de preservação e integração de sistemas	12	
	Profissionais que atuam com preservação	9	
	Estudos de usuários e avaliação dos serviços	3	
Preservação e gestão de conteúdos específicos	Audiovisuais	9	26%
	Dados de pesquisa	27	
	Dados governamentais	10	
	Dados geoespaciais	5	
	Documentos XML	2	
Repositórios digitais confiáveis	Exemplos, avaliação	34	22%
	Métricas	3	
	Metodologias de auditoria e certificação	8	
Curadoria digital	Aspectos conceituais e desenvolvimento	10	5%

Fonte: Dados da pesquisa (2019).

Conforme apresentado na Tabela 6, o assunto principal mais trabalhado entre os documentos recuperados foi a preservação digital, que, devido às perspectivas abordadas, demandou um desmembramento em cinco subcategorias: as que exploravam os aspectos conceituais, as que descreviam aplicações da atividade de preservação digital, os metadados específicos de preservação e integração entre sistemas, os estudos de usuários de serviços de preservação e o estudo dos profissionais que lidam com a prática. Dos 205 documentos analisados, 97 trataram do assunto “Preservação digital”, sendo 38 referentes à subcategoria “aspectos conceituais”, 35 referentes à “planejamento e implantação de ações”, 12 aos metadados e integração entre sistemas, nove remetem aos profissionais que atuam na execução da gestão e técnicas de preservação e três são estudos de usuários de serviços de preservação digital. Esses temas abordados compreendem 47% do total de documentos analisados.

Nesta categoria estão revisões sobre a elaboração de políticas institucionais de preservação digital e reflexões sobre as orientações relacionadas ao assunto. O conteúdo dos trabalhos ressalta a definição de preservação digital, as normas internacionais que buscam padronizar as práticas a serem desenvolvidas, as estratégias utilizadas pelas instituições como o *backup*, emulação, encapsulamento, alguns programas nacionais de preservação já desenvolvidos, o papel da Ciência da Informação no contexto da preservação.

Os documentos que abordam a subcategoria “metadados” buscam reforçar a necessidade de conhecer e preencher corretamente os metadados de preservação. Os artigos selecionados conceituam metadados e apontam sua importância, são apresentados estudos de inclusão dos mesmos pelos profissionais da informação, demonstradas as diferenças entre os metadados administrativos e de preservação, os padrões de metadados utilizados atualmente e as correspondências utilizadas para a integração de registros entre repositórios.

As publicações enquadradas na subcategoria “profissionais” compreendem artigos que abordam a identidade dos profissionais que lidam com os repositórios e fazem a gestão dos documentos digitais. Estão direcionados aos profissionais da informação como arquivistas, administradores, elaboradores de políticas, bibliotecários, buscando apontar as ações para adquirir a confiança dos usuários nas informações abrigadas e a identidade profissional que devem assumir.

Integrando e finalizando a categoria “preservação digital” a subcategoria “usuários” demonstra a avaliação de repositórios digitais pelos usuários e a perspectiva da confiança de conteúdos a partir da visão das comunidades usuárias das informações disponibilizadas. Essa correspondência entre o

usuário e a confiabilidade da informação é expressa no conceito de repositório digital confiável descrita por RLG / OCLC (2002): “um repositório digital confiável é aquele cuja missão é fornecer acesso confiável e de longo prazo a recursos digitais gerenciados para sua comunidade designada, agora e no futuro” (p. 5).

A segunda categoria identificada com maior representatividade foi a preservação e gestão de conteúdos específicos, e suas subcategorias “audiovisuais”, “dados de pesquisa”, “dados governamentais”, “dados geoespaciais” e “documentos XML”. A opção por desmembrá-la em subcategorias específicas se deu porque ressaltam a importância já identificada pelos teóricos em trabalhar especificamente com a preservação destes conteúdos. Foram incluídos 53 documentos nesta categoria, representando 26% do total de documentos recuperados nesta revisão de literatura. Os conteúdos audiovisuais são abordados por nove documentos que tratam da preservação no contexto de histórias orais, de arquivos de mídias, de documentações sonoras, de imagens em movimento para exibição em canais online e de televisões. A discussão sobre a preservação dos conteúdos gira em torno da gestão e do arquivamento das imagens e sons, considerando o papel de espaços além de bibliotecas e arquivos.

No contexto crescente de valorização e disponibilidade dos dados utilizados no desenvolvimento das pesquisas científicas, com a possibilidade de reuso e economia de recursos, a preservação dos conteúdos é um aspecto importante e foi trabalhado em 27 documentos. Os dados governamentais também passam a ser considerados no contexto de transparência e prestação de contas das ações dos governos, sendo essencial a disponibilização e a preservação das informações produzidas, conteúdo trabalhado em sete documentos. Acredita-se que os dados geoespaciais são conteúdos especificamente abordados em cinco artigos por um motivo pontual, devido ao histórico das agências espaciais nacionais, preocupadas com a preservação dos dados capturados por satélites e suas manipulações, em um esforço dos pesquisadores em compreender a dinâmica e os processos evolutivos do Universo (SWEETKIND-SINGER; LARSGAARD; ERWIN, 2006), conforme histórico das normativas de preservação digital, já citado no referencial teórico deste artigo. Por fim, há dois documentos que apresentam estruturas para descrição e gerenciamento dos documentos XML, descrevendo as vantagens de se manter registros organizacionais em formato aberto e passíveis de preservação e recuperação (SALMINEN; JAUHAINEN; NURMEKSELA, 2014).

O assunto principal “repositórios digitais confiáveis” esteve presente em 45 documentos recuperados, correspondendo a 22% do total. Considera-se interessante analisar esse percentual já

que este foi o termo utilizado para a busca dos conteúdos. Contudo, acredita-se que o reduzido índice de documentos que tratam especificamente deste tipo de repositório, indicando exemplos, avaliações, métricas e metodologias de auditorias, deve-se ao fato de que ainda são poucos repositórios que alcançam a certificação. Aspectos que justificam essa condição estão entre os abordados por Miquel Térmens e David Leija (2017) quanto à complexidade das auditorias e altos custos por exigirem profissionais especializados, o que faz com que as organizações fiquem relutantes quanto à aplicação, além de pressuporem ações efetivas de preservação digital, que ainda não são realidade na maioria dos repositórios.

Os documentos recuperados indicam experiências de repositórios confiáveis de países como Irlanda, Estados Unidos e Alemanha, incluindo iniciativas como as das bibliotecas digitais da Library and Archives Canada (LAC), JSTOR, GESIS Data Archive, Digital Repository of Ireland. São apresentadas também na subcategoria “métricas” as formas de avaliação da relevância e utilidade dos repositórios digitais, seguindo aspectos baseados em modelos consolidados como o OAIS e outros criados e descritos nos documentos. E, finalizando a categoria, na subcategoria “metodologias de auditoria e certificação” estão documentos que descrevem os métodos desenvolvidos especificamente para auditar sistemas de preservação, entre os quais os modelos Drambora, Nestor, TRAC, ISO 16363, Deakin, NDSA Levels.

Por fim, a categoria “curadoria digital” identificada em 10 artigos, o que corresponde a 5% do total de documentos, inclui textos que abordam a curadoria de conteúdos dispostos em *big datas*, arquivos e infraestruturas de materiais dispostos na *web*, além da aplicação da curadoria com documentos produzidos em universidades.

6 CONSIDERAÇÕES FINAIS

A produção de documentos digitais é prática frequente em instituições, no entanto, pouco tem sido trabalhado com relação à garantia de acesso por longo prazo aos conteúdos produzidos. Essa condição opõe-se ao propósito de organizações informacionais como bibliotecas, arquivos e museus, que prezam pelo registro, guarda e resgate de memórias e patrimônios que lhes são conferidos.

Assim, ressalta-se a importância de trabalhar conteúdos relativos à preservação digital e de ferramentas que possibilitem a gestão de documentos e a realização de processos e técnicas de desenvolvimento das práticas de manutenção de registros por longo prazo, garantindo a confiabilidade dos conteúdos.

Buscou-se, no presente artigo, contribuir com a temática utilizando como aspecto referencial o conceito de repositórios digitais confiáveis. A pesquisa por documentos científicos que abordavam a expressão retornou um número significativo de itens e, a partir deles, foi possível identificar como a literatura vem trabalhando o tema.

A partir do estudo realizado pode-se perceber que as publicações sobre o assunto ainda são incipientes, grande parte aborda aspectos conceituais e implementação da preservação digital. Citam a possibilidade de preservação realizada pelos repositórios, mas não exploram o conceito de repositórios digitais confiáveis. Acredita-se que isso acontece porque ainda há um número reduzido de repositórios que se enquadram nesta categoria, pois exige o cumprimento de normas e padrões e a realização de auditorias para a certificação da confiabilidade dos softwares, um processo dispendioso para muitas instituições gestoras de repositórios.

Dessa forma, espera-se que a pesquisa contribua para apresentar lacunas na literatura científica e possibilidades de exploração de conteúdos relativos ao tema e possibilite a síntese das dimensões trabalhadas em nível nacional e internacional.

REFERÊNCIAS

ARELLANO, Miguel Ángel Márdero. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, dez. 2004. Disponível em: <http://revista.ibict.br/ciinf/article/view/1043/1113>. Acesso em: 05 dez. 2018.

BARDIN, Laurence. **Análise de conteúdo**. Tradução Luís Antero Reto, Augusto Pinheiro. São Paulo: Edições 70, 2011.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. **Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais**. Rio de Janeiro: Arquivo Nacional, 2012. Disponível em: http://www.conarq.arquivonacional.gov.br/media/diretrizes_presuncao_autenticidade_publicada.pdf. Acesso em: 05 dez. 2018.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Audit and certification of trustworthy digital repositories**: recommended practice: CCSDS 652.0-M-1. Washington, DC: Magenta Book, Sep. 2011. Disponível em: <https://public.ccsds.org/pubs/652x0m1.pdf>. Acesso em: 07 dez. 2018.

CONWAY, Paul. **Preservação no universo digital**. 2. ed. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos, 2001.

DIGITAL LIBRARY FEDERATION. **METS**: Metadata encoding and transmission standard: primer and reference manual. Washington, DC, 2010. Disponível em: <http://www.loc.gov/standards/mets/METSPrimer.pdf>. Acesso em: 05 maio 2018.

EVANS, D.; PEARSON, A. Systematic reviews: gatekeepers of nursing knowledge. **Journal of Clinical Nursing**, v. 10, n. 5, p. 593–599, 2001.

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.

LE COADIC, Yves François. **A Ciência da Informação**. Tradução de Maria Yêda F. S. de Filgueiras Gomes. Brasília: Briquet de Lemos, 1996.

MAEMURA, Emily; MOLES, Nathan; BECKER, Christoph. Organizational assessment frameworks for digital preservation: a literature review and mapping. **Journal of the Association for Information Science and Technology**, v. 68, n. 7, p. 1619-1637, 2017.

ONLINE COMPUTER LIBRARY CENTER (OCLC). Trustworthy Repositories Audit and Certification Checklist is Published. **News from OCLC Research**, [201-]. Disponível em: <https://www.oclc.org/research/news/2007/03-12.html>. Acesso em: 03 dez. 2018.

OTTO, Jane Johnson. Administrative metadata for long-term preservation and management of resources: a survey of current practices in ARL Libraries. **Association for Library Collections & Technical Services**, v. 58, n. 1, 2014.

PORTAL DE PERIÓDICOS CAPES/MEC. Missão e objetivos. Brasília, [201-]. Disponível em: https://www.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pcontent&alias=missao-objetivos&Itemid=109. Acesso em: 13 abr. 2019.

RESEARCH LIBRARIES GROUP; ONLINE COMPUTER LIBRARY CENTER. **Trusted Digital Repositories**: attributes and responsibilities. Mountain View, CA, May 2002. Disponível em: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>. Acesso em: 06 dez. 2018.

SALMINEN, Airi; JAUHAINEN, Eliisa; NURMEKSELA, Reija. A Life Cycle Model of XML Documents. **Journal of the Association for Information Science and Technology**, v. 65, n. 12, p. 2564–2580, 2014.

SANTOS, Henrique Machado dos; FLORES, Daniel. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 2, p. 198-218, jun. 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362015000200198&lng=es&nrm=iso. Acesso em: 03 dez. 2018.

SWEETKIND-SINGER, Julie; LARSGAARD, Mary Lynette; ERWIN, Tracey. Digital Preservation of Geospatial Data. **Library Trends**, v. 55, n. 2, p. 304–314, 2006.

TÉRMENS, Miquel; LEIJA, David. Auditoría de preservación digital con NDSA Levels: methodology of digital preservation audits with NDSA Levels. **El profesional de la información**, v. 26, n. 3, Mayo/ Junio 2017. Disponível em: <https://fd.ub.edu/pub/termens/docs/EPI-v26n3.pdf>. Acesso em: 05 dez. 2018.

THOMAZ, Kátia de P. Repositórios digitais confiáveis e certificação. **Arquivística.net**, Rio de Janeiro, v. 3, n. 1, p. 80-89, jan./jun. 2007. Disponível em: http://www.brapci.inf.br/_repositorio/2010/05/pdf_fed0720dbb_0010726.pdf. Acesso em: 05 dez. 2018.

THE LIBRARY OF CONGRESS. **PREMIS**: Data Dictionary for Preservation Metadata, Version 3.0. Washington, DC, Nov. 2015. Disponível em: <<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>>. Acesso em: 05 maio 2018.

UNESCO. **Guidelines for the preservation of digital heritage**. 2003. Disponível em: <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>>. Acesso em: 06 nov. 2018.