

LYVIA ROCHA DE JESUS ARAUJO

Orientações para a publicação de artigos de dados:
primeiros passos para a construção de um padrão adaptável

Dissertação de mestrado
Março de 2024



LYVIA ROCHA DE JESUS ARAUJO

**ORIENTAÇÕES PARA A PUBLICAÇÃO DE ARTIGOS DE DADOS: PRIMEIROS
PASSOS PARA A CONSTRUÇÃO DE UM PADRÃO ADAPTÁVEL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Orientadora: Profa. Dra. Eloisa Príncipe (*in memoriam*)
Orientadora: Profa. Dra. Rosali Fernandez de Souza
Coorientadora: Profa. Dra. Maria Simone de Menezes Alencar

Rio de Janeiro

2024

A663o Araujo, Lyvia Rocha de Jesus
Orientações para a publicação de artigos de dados:
primeiros passos para a construção de um padrão
adaptável / Lyvia Rocha de Jesus Araujo. -- Rio de
Janeiro, 2024.
108 f.

Orientadora: Rosali Fernandez de Souza.
Coorientadora: Maria Simone de Menezes Alencar.
Dissertação (mestrado) - Universidade Federal do
Rio de Janeiro, Escola da Comunicação, Instituto
Brasileiro de Informação em Ciência e Tecnologia,
Programa de Pós-Graduação em Ciência da Informação,
2024.

1. Artigos de dados. 2. Revistas de dados. 3.
Acesso aberto. 4. Comunicação científica. 5. Ciência
da informação. I. de Souza, Rosali Fernandez,
orient. II. Alencar, Maria Simone de Menezes,
coorient. III. Título.

LYVIA ROCHA DE JESUS ARAUJO

**ORIENTAÇÕES PARA A PUBLICAÇÃO DE ARTIGOS DE DADOS: PRIMEIROS
PASSOS PARA A CONSTRUÇÃO DE UM PADRÃO ADAPTÁVEL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Aprovada em: 22 de março de 2024.

Profa. Dra. Rosali Fernandez de Souza (Orientadora)
Programa de Pós-Graduação em Ciência da Informação, Convênio Ibict/UFRJ.

Profa. Dra. Maria Simone de Menezes Alencar (Co-orientadora)
Programa de Pós-Graduação em Enfermagem e Biociências, UNIRIO.

Profa. Dra. Naira Christofolletti Silveira (Examinadora interna)
Programa de Pós-Graduação em Ciência da Informação, Convênio Ibict/UFRJ.

Profa. Dra. Maria Cristina Gonçalves Guardado (Examinadora externa)
Escola Superior da Tecnologia de Águeda, da Universidade de Aveiro, ESTGA-UA.

Em memória da estimada Profa. Dra. Eloisa da
Conceição Príncipe de Oliveira.

AGRADECIMENTOS

Em primeiro lugar agradeço à minha rede de apoio pelo suporte emocional e o carinho despendido a mim nesses três anos atípicos de curso, em particular aos meus amigos Lucas dos Santos, Lumie Zanazi, Ana Beatriz Vieira, Alexia Couto, Junior Lopes e Junior Rodrigues, à minha família e a todos que estiveram torcendo por mim.

Agradeço também ao Programa de Pós-graduação em Ciência da Informação, desenvolvido a partir do convênio estabelecido entre a Escola de Comunicação da Universidade Federal do Rio de Janeiro – UFRJ/ECO e o Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, aos meus caros docentes e a todos os colegas que trilharam toda ou parte dessa jornada comigo.

À minha orientadora e amiga, Profa. Dra. Eloisa Príncipe, por todo suporte, atenção e cuidado com que me orientou no curso desta pesquisa e por toda dedicação e devoção com que se doou no seu dia a dia em sala de aula e em toda sua carreira como pesquisadora, revisora e docente nas Ciências da Informação. Sua caminhada me inspirou e seguirá me inspirando e impulsionando a ofertar o meu melhor, em todas as áreas da minha vida.

À minha Coorientadora e grande amiga, Profa. Dra. Maria Simone de Menezes Alencar, pelos anos de parceria e aprendizado, trilhados desde a Iniciação Científica na Universidade Federal do Estado do Rio de Janeiro - UNIRIO. Admiro sua visão afiada para a inovação e o entusiasmo pelos caminhos da Ciência Aberta. Particularmente, sinto orgulho de ter sido sua aluna e seguir colocando em práticas todos os seus ensinamentos, da teoria às lições de vida.

À Profa. Dra. Dayane da Silva Prudencio, pela amizade, os ensinamentos, a torcida, os conselhos e o apoio tão genuína e carinhosamente a mim oferecidos desde a orientação do meu Trabalho de Conclusão de Curso na graduação.

À toda equipe do LabInova, com a qual colaborei durante os últimos anos e que me inspirou a seguir por este caminho na pesquisa, em especial aos meus amigos de investigação Diogo Vieira e Solange Campos.

À banca examinadora, especialmente à Profa. Dra. Rosali Fernandez de Souza, pela disponibilidade e as contribuições feitas para a completude desta pesquisa.

E a todos os autores direta e indiretamente citados nesta, por terem pavimentado o caminho.

*“From sprinkler splashes to fireplace ashes
I gave my blood, sweat, and tears for this.”*

- Taylor Alison Swift.

RESUMO

ARAÚJO, Lyvia Rocha de Jesus. **Orientações para a publicação de artigos de dados:** primeiros passos para a construção de um padrão adaptável. Orientadora: Rosali Fernandez de Souza. Coorientadora: Maria Simone de Menezes Alencar. 2024. 108f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2024.

Introdução: A evolução da comunicação científica e a emergência do movimento pelo acesso aberto impulsionou a criação de publicações que fomentam não apenas os resultados das pesquisas realizadas, mas também, o compartilhamento dos dados de suas pesquisas. Uma das práticas influenciadas por este movimento foi a submissão de artigos de dados à publicação em revistas científicas convencionais ou específicas para a publicação desse tipo de informação. **Objetivo:** Elaborar uma proposta para a implementação de um modelo de artigo de dados que sirva de subsídio às revistas científicas brasileiras. **Metodologia:** Foi realizada uma análise das orientações aos autores de 117 revistas indexadas na *Scopus* e na *Web of Science* que publicam artigos de dados. A descrição comparativa dos elementos para a elaboração de artigos de dados foi realizada de acordo com 3 categorias: propriedades gerais do conjunto de dados, descrição da produção/coleta dos dados e descrição sobre compartilhamento e reúso dos dados. **Resultados:** Dentre os elementos mais solicitados pelas revistas na composição de *data papers*, destacam-se: o título dos conjuntos de dados, a descrição da metodologia de coleta/produção e análise/tratamento dos dados, sua cobertura e o objetivo de sua produção/coleta, o contexto a que serviram primariamente, e os termos e condições de uso e reprodução dos dados de pesquisa compartilhados. **Conclusão:** as 117 revistas incluídas e exploradas nesta pesquisa, 20 estiveram presentes direta ou indiretamente em outras pesquisas sobre o tema. A análise atual destas revistas demonstrou diferenças sutis nos moldes disponibilizados para a elaboração de artigos de dados. Ao se ponderar a efetividade dos artigos de dados em seu objetivo principal, é importante levar em conta até que ponto as orientações e formatos disponíveis para construção e submissão de artigos de dados contemplam as necessidades informacionais para tomada de decisão dos pesquisadores sobre o reúso de dados em suas investigações. Apesar disso, entende-se que os primeiros passos para o surgimento de um padrão de campos e recomendações para a publicação de artigos de dados já vêm sendo fundamentados. Deste modo, a presente pesquisa sugere investigações mais apuradas e aprofundadas em contexto acadêmico-científico para melhor compreender as motivações que norteiam a construção das diretrizes para publicação de artigos de dados.

Palavras-chave: Artigos de dados. Revistas de dados. Acesso aberto. Comunicação científica. Ciência da informação.

ABSTRACT

ARAÚJO, Lyvia Rocha de Jesus. **Orientações para a publicação de artigos de dados:** primeiros passos para a construção de um padrão adaptável. Orientadora: Rosali Fernandez de Souza. Coorientadora: Maria Simone de Menezes Alencar. 2024. 108f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2024.

Introduction: The evolution of scientific communication and the emergence of the open access movement has led to the creation of publications that promote not only the results of research, but also the sharing of research data. One of the practices influenced by this movement has been the submission of data articles for publication in traditional or dedicated scientific journals. **Objective:** To develop a proposal for the implementation of a data article model to serve as a subsidy for Brazilian scientific journals. **Methodology:** An analysis was made of the guidelines for authors of 117 journals indexed in Scopus and Web of Science that publish data articles. The comparative description of the elements for preparing data articles was carried out according to 3 categories: general characteristics of the dataset, description of data production/collection, and description of data sharing and reuse. **Results:** The elements most requested by journals when writing data papers include: the title of the datasets, a description of the methodology used to collect/produce and analyze/process the data, the scope and purpose of their production/collection, the context in which they were primarily used, and the terms and conditions for the use and reproduction of the shared research data. **Conclusion:** Of the 117 journals included and examined in this study, 20 were directly or indirectly included in another research on the topic. The current analysis of these journals revealed subtle differences in the way data articles are produced. When considering the effectiveness of data articles in terms of their main objective, it is important to consider the extent to which the guidelines and formats available for preparing and submitting data articles meet the information needs of researchers when making decisions about reusing data in their research. Nevertheless, it is anticipated that the first steps towards the emergence of a standard set of fields and recommendations for the publication of data articles have already been taken. In this way, this research suggests more in-depth investigations in an academic-scientific context to better understand the motivations behind the construction of guidelines for publishing data articles.

Keywords: Data papers. Data journals. Open access. Scientific communication. Information science.

LISTA DE FIGURAS

Figura 1 - Taxonomia FOSTER tradução PT-BR.....	34
Figura 2 - Taxonomia dos dados de pesquisa	36
Figura 3 - Ciclo de vida dos dados de pesquisa	38
Figura 4 - Organograma do artigo de dados.....	40
Figura 5 - Fluxograma do processo de seleção das revistas incluídas na pesquisa.....	51

LISTA DE QUADROS

Quadro 1 - Licenças Creative Commons	37
Quadro 2 - Barreiras encontradas na publicação de resultados negativos	44
Quadro 3 - Características das revistas e elementos essenciais dos artigos de dados.....	48

LISTA DE GRÁFICOS

Gráfico 1 - Frequência dos termos adotados pelas revistas para se referir aos artigos de dados	54
Gráfico 2 – Revista solicita descrição da data de criação dos conjuntos de dados?.....	55
Gráfico 3 - Revista solicita descrição do formato dos conjuntos de dados?	56
Gráfico 4 - Revista solicita descrição do tipo dos conjuntos de dados?	57
Gráfico 5 - Revista solicita descrição da versão dos conjuntos de dados?.....	58
Gráfico 6 - Revista solicita descrição do título/nome dos conjuntos de dados?.....	58
Gráfico 7 - Revista solicita descrição do tamanho dos conjuntos de dados?	59
Gráfico 8 - Revista solicita descrição da linguagem dos conjuntos de dados?	59
Gráfico 9 - Revista solicita descrição do método de produção/coleta dos dados?	61
Gráfico 10 - Revista solicita descrição da cobertura dos dados?	62
Gráfico 11 - Revista solicita descrição da contribuição dos autores?	63
Gráfico 12 - Revista solicita descrição do método do processamento dos dados?.....	63
Gráfico 13 - Revista solicita descrição dos dados faltantes nos conjuntos de dados?.....	64
Gráfico 14 - Revista solicita descrição do objetivo da produção/coleta dos dados?	65
Gráfico 15 - Revista solicita descrição do contexto dos dados?.....	66
Gráfico 16 - Revista solicita descrição da proveniência dos dados?	67
Gráfico 17 - Revista solicita descrição da reputação e histórico dos repositórios?.....	68
Gráfico 18 - Revista solicita descrição da preservação dos dados?	69
Gráfico 19 - Revista solicita descrição do reúso de dados anteriores?.....	70
Gráfico 20 - Revista solicita descrição do reúso potencial dos dados?	70
Gráfico 21 - Revista solicita descrição dos termos e condições de uso dos dados?	71

LISTA DE TABELAS

Tabela 1 - Dispersão do número de revistas por editora científica	52
Tabela 2 - Dispersão do número de revistas em relação ao acesso	53

LISTA DE SIGLAS

ABEC Brasil	Associação Brasileira de Editores Científicos
APC	<i>Article Processing Charge</i>
BIREME	Centro Latino-americano de Informação em Ciências da Saúde
BOAI	<i>Budapest Open Access Initiative</i>
CC	<i>Creative Commons</i>
CEP	Comitê de Ética e Pesquisa do Centro de Filosofia e Ciências
CFCH/UFRJ	Humanas/Universidade Federal do Rio de Janeiro
CI	Ciência da Informação
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DOI	<i>Digital Object Identifier</i>
FAPESP	Fundação de Apoio à Pesquisa do Estado de São Paulo
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ISTA	<i>Information Science & Technology Abstracts</i>
LISTA	<i>Library, Information Science and Technology Abstracts</i>
OASIS	<i>Open Access Scholarly Information System</i>
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OECD	Organization for Economic Co-operation and Development
OSI	<i>Open Society Institute</i>
PGD	Plano de Gestão de Dados
PLoS	<i>Public Library of Science</i>
SciELO	<i>Scientific Electronic Library Online</i>
SPARC	<i>Scholarly Publishing and Academic Resources Coalition</i>
TI	Tecnologias da Informação

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS.....	20
1.2 JUSTIFICATIVA	21
2 DA COMUNICAÇÃO CIENTÍFICA AOS ARTIGOS DE DADOS	23
2.1 COMUNICAÇÃO CIENTÍFICA.....	23
2.2 CIÊNCIA ABERTA E ACESSO ABERTO	29
2.3 DADOS ABERTOS DE PESQUISA.....	34
2.4 REVISTAS E ARTIGOS DE DADOS	40
3 PROCEDIMENTOS METODOLÓGICOS	46
3.1 PRÉ-ANÁLISE	46
3.2 EXPLORAÇÃO DO MATERIAL: Coleta e organização dos dados	47
3.2.1 Fase 1 - Levantamento das revistas que publicam artigos de dados	47
3.2.2 Fase 2 - Características das revistas e elementos dos artigos dados	48
3.3 TRATAMENTO E INTERPRETAÇÃO DOS RESULTADOS: análise dos dados	49
4 RESULTADOS	51
4.1 CARACTERÍSTICAS DAS REVISTAS QUE PUBLICAM ARTIGOS DE DADOS	52
4.2 DAS ORIENTAÇÕES PARA A SUBMISSÃO DE ARTIGOS DE DADOS	54
4.2.1 Propriedades gerais dos <i>datasets</i>	55
4.2.2 Contextualização da produção dos <i>datasets</i>/coleta dos dados	60
4.2.3 Informações sobre o compartilhamento e reúso dos dados	66
5 DISCUSSÃO	73
6 PRIMEIROS PASSOS PARA A CONSTRUÇÃO DE UM PADRÃO ADAPTÁVEL	79
7 CONCLUSÃO.....	83
REFERÊNCIAS	85
APÊNDICE A - Trajetória acadêmica da autora.....	90
APÊNDICE B - Plano de Gestão de Dados (PGD).....	93
APÊNDICE C - Revistas excluídas.....	96
APÊNDICE D - Síntese comparativa dos elementos relacionados às propriedades gerais do conjunto de dados	97
APÊNDICE E - Síntese comparativa dos elementos relacionados à contextualização da produção dos <i>datasets</i> /coleta dos dados	101
APÊNDICE F - Síntese comparativa dos elementos relacionados às informações sobre o compartilhamento e reúso dos dados.....	105

1 INTRODUÇÃO

Esta pesquisa nasce do interesse pessoal da autora pela discussão sobre a efetividade dos artigos de dados no ecossistema da ciência aberta e sua relevância no fomento as etapas de compartilhamento e reúso¹ dentro do ciclo de vida dos dados de pesquisa. No grupo de pesquisa do Labinova houve o contato mais próximo com a produção científica da área de enfermagem. Observou-se um grande volume de dados coletados que muitas das vezes encontram seu uso fim em seus estudos primários e são guardados com os autores/pesquisadores após os resultados delas serem publicados. A partir das discussões fomentadas no período de iniciação científica durante a graduação, as indagações anteriores tomaram corpo e se expandiram às questões que norteiam os passos desenvolvidos na presente pesquisa. A evolução das atividades desde a graduação encontra-se no Apêndice A.

Como ambiente de desenvolvimento e aprofundamento desta investigação, foi escolhido o Programa de Pós-Graduação em Ciência da Informação, estabelecido pelo convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Universidade Federal do Rio de Janeiro (UFRJ). Inserida no Programa, a temática foi trabalhada sob a luz dos conhecimentos explorados na Linha 1 - Comunicação, Organização e Gestão da Informação e do Conhecimento do programa, que dentre suas subáreas investiga assuntos relacionados à comunicação e divulgação em ciência e tecnologia; sistemas de organização e representação do conhecimento; e processos de busca, acesso, recuperação e uso da informação.

Para contextualizar os artigos de dados no panorama das publicações científicas, buscou-se embasamento na teoria da construção da comunicação científica, que historicamente se configurou a partir da dialética exercida na Grécia antiga, entre filósofos, matemáticos e outros estudiosos e entusiastas que compartilhavam suas descobertas em praça pública, discutindo possibilidades e hipóteses diante dos fenômenos por eles observados. Mais adiante, no século XVII, o surgimento das sociedades científicas reuniria estudiosos e pesquisadores profissionais em um núcleo de maior proximidade, facilitando a troca de informações científicas e a cooperação no desenvolvimento de novas ideias.

Até o advento dos periódicos científicos, publicações formais revisadas por pares, produzidos em período curto e atualizado com frequência determinada, a comunicação científica ocorria mediante troca de cartas entre os membros das sociedades científicas, de forma informal e desprendendo um tempo considerável até o retorno das considerações

¹ Conforme consulta com a ABL - Academia Brasileira de Letras, ficou esclarecido que o uso correto da palavra “reúso” inclui o acento agudo acima da letra “u”. Em revisão final foi percebida a necessidade de correção padronizada da gramática e foram realizadas as alterações necessárias.

almejadas. Na medida em que aumentava o volume de cartas científicas e crescia a demanda pela atualização mais rápida da evolução da ciência, sociedades científicas pensavam em uma alternativa acessível que melhor atendesse à comunidade acadêmico-científica da época.

Frente a este desafio, foram criados, em 1665, os dois primeiros periódicos científicos - o *Journal de Sçavans* (França) e as *Philosophical Transactions of the Royal Society* (Inglaterra), com a intenção de expandir a comunicação formal de novas descobertas e inovações científicas de forma mais aprofundada, econômica e em curto espaço de tempo, sendo atualizada com maior frequência e mantendo o rigor demandado pelos livros acadêmicos, por exemplo (Meadows, 1999).

Estas publicações abriram espaço para que outras sociedades científicas pudessem desenvolver e veicular seus próprios periódicos, até que, no século XX, editoras comerciais passaram a visualizar a possibilidade de lucro com estas publicações, aumentando exponencialmente o volume de revistas científicas publicadas ao redor do mundo. De lá para cá uma gama de mudanças alteraram o percurso de desenvolvimento destas publicações, as editoras ganharam cada vez mais força e impacto no mercado científico e o próprio escopo das publicações foi se tornando cada vez mais especializado.

A ascensão das Tecnologias da Informação (TI) no século XX, principalmente após o fim da Segunda Guerra Mundial, trouxe nova perspectiva à automatização de processos de coleta, armazenamento, busca e recuperação de informações. A essa altura o volume de informações circulando diariamente já era tão grande, que não podia ser processado sem o auxílio da tecnologia, tanto no meio acadêmico-científico quanto fora dele.

Na comunicação científica a adaptação dos formatos físicos de publicação para o eletrônico levou um pouco mais de tempo para se consolidar. Dentre os empecilhos encontrados pelos leitores até o fim da década de 1990, destacam-se: o tempo que demorava para se estabelecer uma boa conexão; a falta de mobilidade nos instrumentos utilizados pelos leitores (*desktops*, disquetes e CD-ROMs) para poder acessar textos digitais; e o déficit de entrosamento de pesquisadores mais velhos com a tecnologia. Nesse período, a prática mais utilizada pelos leitores era a impressão ou a cópia de arquivos eletrônicos (Meadows, 1999).

Antes que os periódicos e livros eletrônicos se consolidassem nos formatos atuais, outras questões passaram a ser consideradas. Apesar das otimizações no que diz respeito à organização de conteúdos interessantes em um mesmo volume e a atualização periódica destes mesmos conteúdos, a dificuldade que muitos leitores encontravam de arcar com o custo elevado das assinaturas de revistas científicas já era um problema existente desde a criação dos primeiros

periódicos e sua adaptação para o formato eletrônico não contribuiu muito com a redução destes custos.

As preocupações com o impacto causado no acesso às publicações científicas, devido às restrições e taxas excessivas de editoras comerciais, fizeram emergir, no início da década de 1990, o Movimento pelo Acesso Aberto à informação científica (*Open Access*). O movimento surgiu com a intenção de tornar os resultados publicados de pesquisas científicas acessíveis, gratuitos e livres de qualquer barreira que pudessem dificultar seu alcance para o público de interesse. Tal iniciativa esperava não apenas ampliar o volume de conhecimento cientificamente comprovado divulgado para um público maior de receptores em potencial, como também incentivar práticas colaborativas de produção de novos saberes.

O movimento pelo Acesso Aberto teve maior impulso e visibilidade após a publicação da Iniciativa de Acesso Aberto de Budapeste (*Budapest Open Access Initiative*, 2002), da Declaração de Bethesda sobre Publicação de Acesso Aberto (2003) e da Declaração de Berlim sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades (2003).

A Iniciativa de Acesso Aberto de Budapeste (BOAI), lançada em reunião promovida pelo *Open Society Institute* (OSI), teve como principal objetivo analisar como iniciativas isoladas de abertura do acesso a pesquisas científicas poderiam operar de forma colaborativa, e como a OSI e outras fundações poderiam contribuir com esta empreitada. Neste evento foram recomendadas duas soluções para as discussões ali tratadas: 1) o auto-arquivamento de documentos e informações em repositórios institucionais; 2) a publicação de artigos científicos em revistas de acesso livre, a partir das quais quaisquer leitores interessados poderiam ter acesso a conteúdos científicos.

No ano seguinte, foi lançada a Declaração de Bethesda, em reunião realizada no *Howard Hughes Medical Institute*, em Bethesda, objetivava delinear princípios para a obtenção de apoio financeiro das agências de financiamento governamentais e institucionais em prol do compartilhamento livre de resultados científicos. Em suma, esta Declaração propõe mudanças nas políticas institucionais e públicas relacionadas à publicação científica e reforça as condições apontadas na BOAI.

Ainda em 2003 surge a Declaração de Berlin, definida pela *Max Planck Society* (Alemanha), o OSI, a *Public Library of Science* (PLOS), e *Scholarly Publishing and Academic Resources Coalition* (SPARC), com o objetivo de endossar as declarações anteriores através da recomendação do uso consistente da Internet para divulgação e publicação científicas, encorajando pesquisadores a publicar seus estudos em acesso aberto. A Declaração de Berlin foi atualizada em 2005, reforçando a necessidade de se avaliar a produção disponível em acesso

aberto, definindo padrões de qualidade e reconhecimento para gerar métricas de progressão acadêmica no contexto das publicações abertas.

O delineamento das vertentes relacionadas à abertura dos processos científicos tornou o acesso aberto uma das iniciativas vinculadas ao “guarda-chuva” da Ciência Aberta, que por sua vez compreende toda a extensão das práticas abertas no contexto científico. Em essência, a Ciência Aberta nada mais é do que um conjunto de iniciativas e práticas que garantem a disseminação livre do conhecimento, que torna todos os componentes do processo científico acessíveis e reutilizáveis, permitindo seu uso e compartilhamento irrestrito por quaisquer pessoas interessadas (Albagli, 2019).

Segundo Albagli, Clinio e Raychtock (2014), a definição do que é Ciência Aberta tem como inspiração: a) o acesso aberto e livre a materiais científicos; b) a liberdade para redistribuir estes materiais; c) a liberdade para reutilizá-los e; d) a inexistência de restrições baseadas em características específicas (nacionalidade ou campo de atuação).

A Ciência Aberta fez com que os atores envolvidos no processo de comunicação científica, como pesquisadores, editores, agências de fomento, começassem a repensar o modo de se conduzir a ciência, em que não apenas as publicações, mas também as ferramentas, recursos e dados de pesquisa pudessem ser disponibilizados sem restrições de acesso, além de serem produzidos, avaliados, publicados e preservados de maneira aberta, possibilitando seu uso e citação dentro dos direitos e deveres de autoria e divulgação.

Segundo a taxonomia proposta por Pontika *et al.* (2015), dentre os principais eixos da Ciência Aberta, encontram-se os Dados Abertos, que compreendem o *Big Data* Aberto, a definição do que são Dados Abertos, as Revistas de Dados Abertos, os Padrões relacionados aos Dados Abertos o Uso e Reúso de Dados Abertos e os Dados Abertos Governamentais. No cenário concebido pelas revistas, padrões e o uso e reúso de dados, se inserem os Dados Abertos de Pesquisa, Repositórios de Dados, *Data Journals* e *Data Papers*, pontos de interesse desta pesquisa.

Compreende-se como dado de pesquisa “[...] todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica.” (Sales; Sayão, p. 36, 2019). Dentre os objetivos que norteiam a publicação de dados, o principal é o de torná-los reutilizáveis e reproduzíveis. Para Kim (2020) estes objetivos são cumpridos pelo compartilhamento de conjuntos de dados em repositórios e a elaboração de *data papers* (artigos de dados), que fornecem documentação extensa abordando informações contextuais suficientes para a interpretação dos conjuntos de dados compartilhados por pesquisadores.

Callaghan e colaboradores (2012), explicam que um *data paper* fornece informações sobre o quê, onde, por quê, como e de quem são os dados de pesquisa compartilhados, além de garantir o acesso a estes materiais através de um identificador permanente (*Digital Object Identifier* - DOI) que direcione o pesquisador ao repositório em que os dados de pesquisa que procuram estão armazenados. Os artigos de dados ou *data papers*², são orientados para apresentação dos metadados que descrevem a estrutura, o formato, os métodos utilizados na coleta, organização e processamento dos dados. Estes também incluem a autoridade responsável pelos dados, o licenciamento sob o qual estão disponibilizados, um DOI e seu potencial de reutilização (Candela *et al.*, 2015).

Os artigos de dados são publicados em revistas convencionais híbridas que aceitem sua submissão, publicações ampliadas³ ou revistas de dados (*data journals*). As revistas de dados surgem a partir da discussão sobre os benefícios do compartilhamento e o reúso de dados de pesquisa e as barreiras metodológicas, legais e técnicas que fazem com que pesquisadores não se sintam muito inclinados a publicar seus dados (Candela *et al.*, 2015).

Chavan e Penev (2011) discorrem sobre a função dos *data journals* como publicações de documentos de metadados pesquisáveis, que descrevem conjuntos de dados disponibilizados em formato digital dentro dos padrões de ações de divulgação científica já existentes. Essa função se soma às características comuns aos periódicos científicos convencionais, como a revisão por pares e a possibilidade de abranger conteúdos de uma ou mais áreas do conhecimento, mas os difere destes mesmos periódicos, na medida em que não compartilham da intenção de propagar os resultados de pesquisas acadêmicas e estudos científicos, limitando-se à descrição detalhada dos conjuntos de dados e de sua localização.

Uma revista de dados permite que os conjuntos de dados de pesquisa produzidos sejam compartilhados como uma “produção de pesquisa de primeira classe” publicada, revisada por pares, adequadamente indexada nas principais bases bibliográficas, que os tornem citáveis, facilmente detectáveis e reutilizáveis. Tanto as revistas quanto os próprios artigos de dados podem ser indexados e recuperados nas bases de dados tradicionais utilizadas para a disponibilização de periódicos convencionais de assuntos gerais ou especializados. O que os diferencia é a intenção que as revistas e artigos de dados têm de tornar um conjunto de dados reutilizável e reproduzível, mediante descrição aprofundada que permita sua exploração livre, garantindo os direitos de quem o tenha produzido (Candela *et al.*, 2015).

² Os termos “artigos de dados” e “*data papers*” são empregados indistintamente nesta pesquisa.

³ Publicação reforçada, melhorada, sobreposta, que possibilita a interatividade, a conectividade e a convergência de todas as partes da pesquisa (Oliveira, 2019, p.104).

Um número crescente de artigos de dados já é encontrado em revistas híbridas ou revistas de dados do exterior, como o *Data in Brief*, o *Scientific Data*, o *Earth System Science Data*, o *Data* e o *Biodiversity Data Journal*, que até janeiro de 2023 possuíam, respectivamente, 6.300, 2.070, 613, 312 e 300 artigos de dados indexados na *Web of Science*.

No Brasil o número de revistas que informam aceitar a publicação de artigos de dados ainda é muito reduzido. Até o momento poucos estudos foram realizados em âmbito nacional discutindo este assunto e apenas nos últimos dois anos observou-se a inclusão de artigos de dados no escopo de submissão de revistas científicas brasileiras⁴.

Diante do exposto, a presente pesquisa busca analisar as características dos artigos de dados de acordo com as orientações de submissão fornecidas pelas revistas de dados indexadas nas bases de dados *Scopus* e *Web of Science*, com vista à elaborar uma proposta de diretrizes para a implementação de um modelo de artigo de dados adaptável que sirva de apoio aos periódicos científicos brasileiros.

Considerando o contexto da Ciência Aberta e a publicação de artigos de dados visando uma ampla visibilidade, disseminação e o compartilhamento de dados de pesquisa (publicação) em revistas científicas, esta pesquisa é norteadas pelas seguintes questões: Quais elementos básicos/essenciais devem ser incluídos em um artigo de dados? Qual o impacto que a presença destes elementos pode ter na promoção do reuso e compartilhamento de dados de pesquisa, considerando a natureza dos artigos de dados?

1.1 OBJETIVOS

Elaborar uma proposta de diretrizes para a implementação de um modelo de artigo de dados adaptável que sirva de apoio aos periódicos científicos brasileiros.

Para tanto, buscar-se-á:

- a) Identificar títulos e características de revistas científicas que publicam artigos de dados;
- b) Analisar as orientações das revistas para submissão de artigos de dados, visando definir categorias de análise e os elementos de cada categoria.

⁴ A primeira foi a revista *Encontros Bibli*, uma publicação periódica do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (Brasil), que publicou o primeiro artigo de dados no volume 25, em 8 de maio de 2020. Em seguida foi criada a *Latin American Data in Science*, uma revista de dados editada pela Universidade Federal de Lavras (UFLA) em parceria com representantes da Universidad Mayor de San Andrés (UMSA - Bolívia), da Universidad de Antioquia (UdeA - Colômbia) e da Universidade de Münster (Westfälische Wilhelms-Universität - Alemanha), tendo publicado seu primeiro artigo de dados no volume 1, número 1, em 4 de julho de 2021.

1.2 JUSTIFICATIVA

A comunicação científica se transforma na medida em que novas ideias e tecnologias da informação são desenvolvidas, aceitas e aplicadas pela sociedade para viabilizar a disseminação do conhecimento. A adaptação das revistas científicas para o formato digital e sua difusão em plataformas *online* facilitou a globalização dos saberes, permitindo que pesquisadores do mundo inteiro tivessem acesso a publicações internacionais, rompendo com as limitações geográficas antes encontradas por autores que quisessem trabalhar em colaboração com membros de outras comunidades científicas, ter acesso aos documentos e dados dos seus pares etc.

Levando em conta o impacto da Ciência Aberta no atual contexto da comunicação científica, principalmente em relação ao compartilhamento de dados de pesquisa em periódicos científicos, acredita-se que a publicação de revistas e artigos orientados a dados pode facilitar a recuperação e o reúso dos conjuntos de dados compartilhados na internet. Mas, apesar da evolução da abertura dos processos científicos, alguns conceitos relacionados à Ciência Aberta e à abertura dos dados ainda causam estranhamento e dúvidas em pesquisadores brasileiros, que podem temer o compartilhamento dos dados integrais de suas pesquisas sem que estejam vinculados a uma publicação formal que os traga credibilidade e garanta a citação apropriada de seus autores.

As políticas nacionais de acesso aberto são recentes e a infraestrutura para garantir seu cumprimento ainda está em construção. Além disso, são raras as bases de dados que classificam artigos de dados como tipologia documental específica das revistas científicas que indexam, o que dificulta sua recuperação e diminui a visualização daqueles que buscam conhecê-los. Esses fatores podem contribuir para o baixo índice de compartilhamento de dados e um número reduzido de artigos de dados produzidos por autores de afiliação brasileira.

Nesse contexto, as informações fornecidas pelos *data papers* se tornam registros transparentes e creditáveis do conteúdo dos conjuntos de dados a que se referem, além de servir como guia para expansão ou exploração das amostras de dados previamente coletadas (Candela *et al.*, 2015; Chavan; Penev, 2011; García-García *et al.*, 2015).

A publicação de artigos de dados se faz relevante uma vez que oportuniza o reconhecimento não apenas de pesquisadores cujos dados compartilhados sejam provenientes de estudos primários, mas também de gestores de grandes conjuntos de dados (*big data*) e/ou dados negativos.

Outro ponto considerável a respeito dos artigos de dados é a perspectiva da garantia da confiabilidade, integridade e disponibilidade dos dados de pesquisa, prevista pela combinação do processo de revisão por pares, tanto dos conjuntos de dados quanto da estrutura e do conteúdo dos artigos; das políticas de preservação dos repositórios em que os dados são armazenados; e da atribuição de identificadores persistentes, que garante sua recuperação imediata (Candela *et al.*, 2015; Chavan; Penev, 2011; García-García, *et al.*, 2015).

Nesse sentido, os editores internacionais apresentam modelos de diretrizes que podem inspirar a ampliação do escopo das revistas brasileiras para o aceite destas publicações. Considerando a baixa adesão à publicação de artigos de dados no Brasil, por desconhecimento por parte dos autores e editores, pela falta do estabelecimento de diretrizes que orientem a "elaboração" de artigos de dados, esta pesquisa é motivada a explorar a concepção dos editores associados à Associação Brasileira de Editores Científicos (ABEC Brasil) sobre *data journals* e *data papers*, e quais as perspectivas futuras de editoras brasileiras quanto à adequação dos escopos de suas revistas para a recepção de artigos de dados.

É evidente que existem lacunas a serem preenchidas no que tange à publicação de artigos de dados por pesquisadores, instituições e indivíduos brasileiros. Ao discutir os conceitos relacionados às revistas e aos artigos de dados e conhecer a visão das editoras científicas brasileiras sobre este cenário, espera-se diminuir parte dos obstáculos encontrados por brasileiros no que se refere à disponibilização e registro dos seus dados de pesquisa. Além disso, a exploração de periódicos internacionais que aceitam estes documentos, oferece alternativas para a atual falta de políticas específicas nas revistas científicas brasileiras e de políticas nacionais que dêem conta de suprir as necessidades desse tipo de publicação.

No âmbito da Ciência da Informação (CI), entende-se que a adaptação de revistas científicas para o aceite de artigos de dados e a criação de periódicos voltados exclusivamente para sua publicação evidencia o valor de um dos principais elementos do ciclo da pesquisa: os dados de pesquisa. Toda informação é gerada a partir de dados em um determinado contexto, e uma vez que esses dados estejam disponibilizados em sua forma primária, novas informações podem ser geradas em contextos distintos.

Partindo do que fora defendido nesta justificativa, a seção 2 apresenta o embasamento teórico e esclarecimento dos conceitos abordados nesta pesquisa. Em seguida, são apresentados os procedimentos metodológicos utilizados (seção 3) para o alcance dos resultados obtidos (seção 4) e discutidos na seção 5. Por fim, a seção 6 apresenta as principais conclusões percebidas ao fim da mesma.

2 DA COMUNICAÇÃO CIENTÍFICA AOS ARTIGOS DE DADOS

Esta pesquisa se insere no contexto da evolução da Ciência da Informação, como consequência do modo como esta influencia e é influenciada pelo desenvolvimento de meios de produção e comunicação científica da sociedade através do tempo. O termo Ciência da Informação (CI) foi reconhecido a partir de 1960, consolidando-se como campo de exploração dos fenômenos de produção, comunicação, recuperação e uso da informação (Le Coadic, 1996; Pinheiro; Loureiro, 1995; Saracevic, 1996).

A CI é observada como uma combinação dos conceitos de ciência pura e ciência aplicada. Isso porque envolve em seus estudos questões científicas, conceituais e epistemológicas relacionadas à informação, bem como técnicas e práticas profissionais de construção do conhecimento a partir de informações confiáveis (Saracevic, 1996).

Ao possuir a informação como ponto focal, a ciência da informação consegue elaborar insights que despontam em diversas áreas do conhecimento. A interdisciplinaridade da CI é percebida desde a emergência do campo. Em seu estudo, Saracevic (1996) explica a relação interdisciplinar entre a CI, a Biblioteconomia, a Computação, a Ciência Cognitiva e a Comunicação, mas esse cenário não se limita a estas áreas. Com o passar do tempo, a CI expandiu as fronteiras das relações informação x área do conhecimento, se desdobrando em estudos que alcançam as mais variadas áreas.

Inicialmente voltada para o entendimento das relações da informação com a ciência e o emprego de técnicas de organização e recuperação da informação, entre as décadas de 1980 e 1990, a CI passou a agregar disciplinas voltadas para a disseminação da informação e a mensuração de sua aplicação na construção do conhecimento científico, corroborando com a visão de Le Coadic (1996) de que as atividades científicas existiriam a partir do registro e comunicação das informações científicas, e vice-versa.

Dentre as disciplinas especializadas nesta discussão, destaca-se a comunicação científica, que elucida a construção da sociedade através dos meios que utiliza para veicular os saberes produzidos através do tempo. A Comunicação Científica é uma subárea de interesse da CI que explora as nuances das práticas de publicação científica e as diretrizes que garantam aos pesquisadores e autores os direitos previstos de acesso e propriedade intelectual.

2.1 COMUNICAÇÃO CIENTÍFICA

A maneira como as informações são compartilhadas pode variar de acordo com alguns fatores, como o meio utilizado, a confidencialidade e a criticidade da informação a ser veiculada

e o(s) receptor(es) desta informação. No que se refere à informação científica, Meadows (1999) pontua que o modo como pesquisadores compartilham suas informações depende do canal utilizado (formal ou informal), da natureza das informações e que público se pretende alcançar. Como consequência deste processo, muda-se também a estrutura de formulação dos registros de informações e o acondicionamento necessário para sua preservação.

Especificamente, a comunicação científica se desdobra na criação de soluções que facilitem o processo de distribuição e acesso da/à informação. Em retrospecto, a comunicação científica não possui uma data para o seu surgimento, mas o leque de assuntos sobre o qual se debruça esteve definido, indiretamente, desde a criação dos primeiros registros escritos da informação teórica (científica) ocidental, na Grécia Antiga.

Nesse mesmo período, a discussão verbal entre filósofos e eruditos em assembleias públicas também podia ser observada, prática esta que evoluiu até alcançar os moldes observados nas conferências científicas atuais. Nessa perspectiva começavam a se fundamentar as duas vias na qual se dividem os meios de comunicação da informação: a comunicação formal e a comunicação informal. A primeira representada pelos registros documentais oficiais que salvaguardavam a história e os saberes da época, e a segunda como um caminho simplificado, mais acessível e rápido para troca de ideias, construção e desconstrução de hipóteses.

Em um panorama mais recente, a comunicação formal é compreendida como aquela na qual a credibilidade da informação comunicada tenha sido comprovada pela avaliação de especialistas. Em paralelo, a comunicação informal é aquela pela qual as informações são veiculadas livremente, em apresentações verbais, cartas ou informes, sem que precise ser aprovada por pares ou distribuída por sistemas oficiais (Fjällbrant, 1997; Meadows, 1999). Com o decorrer do tempo e a evolução dos processos de comunicação científica, a relação entre as duas formas foi se tornando, gradualmente, cada vez mais tênue. Nos dias atuais, mesmo algumas conferências têm todo conteúdo apresentado avaliado por especialista, documentado e difundido em revistas científicas ou sistemas oficiais de instituições de ensino e pesquisa.

A mudança entre os níveis de formalidade, contudo, não foram as únicas ocorridas no percurso da difusão da informação teórico-científica. A primeira grande alteração no processo de disseminação da informação foi o surgimento da imprensa na Europa do século XV, que impulsionou a reprodução de obras consagradas historicamente e “barateou” os custos de produção da literatura popular e científica.

A imprensa não serviu somente aos interesses dos leitores cultos e estudiosos, mas tornou mais acessível a comunicação em geral. Nesse mesmo decurso foram produzidos exemplares religiosos, informativos e folhetins variados, fazendo com que em 1613 já se

questionasse o volume de informações “inúteis” impressas na época. Apesar da contradição apontada, na metade do século XV já podia-se notar a melhor e mais rápida difusão de obras científicas, cujo efeito pode ser notado pela origem da *Oxford University Press*, na Inglaterra, nesse mesmo período (Meadows, 1999).

A constituição da imprensa tornou mais curto o prazo de espera pela publicação de obras físicas, mas até o advento das revistas científicas não existia um modelo de publicação estruturado de rápida atualização cuja credibilidade pudesse ser comprovada ou contestada com facilidade. Nesse aspecto, o surgimento das sociedades científicas com foco na difusão do conhecimento científico global, abriu portas para novas perspectivas.

As sociedades científicas surgiram na segunda metade do século XVII, independentemente de concepções políticas, religiosas e de associações profissionais. Estas sociedades recebiam leigos, cientistas e especialistas, e abria espaço para a discussão de temas consolidados e emergentes na ciência e tecnologia. Como prática de coleta de dados científicos para disseminação, as sociedades científicas buscavam explorar o que vinha sendo descoberto em outros países através da troca de correspondência pessoais ou em missões de campo, nas quais cientistas eram enviados pessoalmente para observar a realidade de outras nações. Além disso, era comum a essas organizações a republicação e tradução de obras consagradas (Fjällbrant, 1997; Meadows, 1999).

As primeiras revistas científicas produzidas, inicialmente, possuíam ampla abrangência de assuntos em seu escopo. O *Journal des Savants* se dedicava à publicação de tabelas de classificação de livros para catalogação, resumos dos livros mais lidos no continente europeu, obituários de personalidades, metodologias de estudos em desenvolvimento, decretos jurídicos e tudo mais que pudesse interessar aos eruditos da época. Por sua abrangência e posterior declínio da cobertura de temas científicos, o *Journal des Savants* é tido como precursor das revistas de humanidades. Enquanto isso, o *Philosophical Transactions* dava mais holofotes à divulgação de estudos experimentais, sendo reconhecido como precursor dos periódicos científicos modernos (Meadows, 1999).

Com o novo modelo de publicação ganhando abrangência e alcance entre as sociedades científicas, novas revistas foram sendo criadas. Mas, ainda assim, até a segunda metade do século XVIII o crescimento da publicação de periódicos científicos ocorreu de forma lenta (Fjällbrant, 1997). É digno de nota que as revistas científicas, desde sua origem, não suprimiram outros meios de comunicação científica, como a troca de correspondências ou e-mails, e os livros, por exemplo. Porém, com o amadurecimento de sua estrutura, as revistas

científicas passaram a ser mais aceitas pelos atores principais da cadeia de comunicação científica.

Na categoria dos atores desta cadeia, Fjällbrant (1997) apontava nove pontos focais: 1) os cientistas/autores, produtores primários da informação científica; 2) os pesquisadores, consumidores das informações produzidas; 3) os estudantes, aspirantes à uma área científica; 4) o público em geral, amadores e eruditos interessados nos avanços científicos; 5) os editores científicos, produtores secundários das publicações produzidas; 6) as bibliotecas, facilitadoras do acesso e da recuperação da informação disponibilizada; 7) organizações legais, públicas e industriais que se beneficiam dos produtos da ciência; 8) instituições acadêmicas que utilizam o conhecimento científico diariamente em suas atividades; e 9) instituições religiosas filantrópicas, que entre os séculos XVII e XVIII tinham grande influência em práticas voltadas para o desenvolvimento filosófico e científico.

Atualmente percebe-se uma redução no número de atores realmente envolvidos no processo de comunicação científica ou no próprio ciclo de vida da construção do conhecimento. Ao que Freitas e Leite (2019) expõem na seguinte divisão: 1) autores/pesquisadores, 2) editoras, 3) bibliotecas universitárias e seus bibliotecários, 4) agências de fomento e universidades.

Nesse universo, os autores-pesquisadores iniciam o ciclo de comunicação ao pesquisar na literatura existente o embasamento necessário para a fundamentação de uma hipótese (Freitas; Leite, 2019). Para que alcance um resultado satisfatório nesta etapa, é preciso que o autor tenha a sua disposição toda a informação de que necessita, ou uma seleção de informações que atenda às suas necessidades de pesquisa. Uma vez concluída esta fase, é preciso que o autor utilize de meios formais e informais de comunicação para compartilhar seus achados com as partes interessadas e obter o feedback analítico de seus pares.

Concomitantemente, as editoras científicas dão continuidade ao ciclo de comunicação, coordenando todas as etapas de recepção, avaliação, diagramação e publicação dos manuscritos enviados pelos autores (Freitas; Leite, 2019). As editoras são responsáveis pela continuidade do processo formal de comunicação científica, favorecendo a comprovação da propriedade dos autores sobre os trabalhos publicados, o exame da credibilidade das informações neles contidas e sua utilização no embasamento de investigações posteriores.

Na sequência, as bibliotecas universitárias ou especializadas fazem a ponte entre as editoras e os leitores, adquirindo obras relevantes e disponibilizando-as de forma mais acessível e mais facilmente recuperável para a comunidade de usuários que atendem (Freitas; Leite, 2019). Para atender à constante demanda informacional, as bibliotecas precisam manter seus acervos atualizados, preservados e indexados, com o propósito de tornar o conhecimento

recuperável em tempo hábil, em boas condições e pertinentes às alterações científicas e tecnológicas das áreas cobertas.

Por último, as agências de fomento e as universidades vão influenciar quase todo o ciclo de comunicação, fornecendo a infraestrutura técnica e política necessárias para o desenvolvimento de novas pesquisas, orientando padrões de produção, sugerindo veículos de distribuição e métricas de uso das pesquisas financiadas (Freitas; Leite, 2019). As agências de fomento concedem aporte financeiro à pesquisas promissoras em diversas áreas do conhecimento. Para a concessão deste apoio geralmente são consideradas: a pertinência do tema de pesquisa, a experiência dos autores envolvidos e a estrutura do projeto por eles apresentado. De forma mais simples, às universidades cabe a orientação de aspirantes à pesquisa para que amadureçam suas habilidades de coleta e interpretação de dados com o intuito de transformá-los em conhecimento novo.

Diante das ações que precisam ser desenvolvidas por cada um dos atores da cadeia de comunicação científica, fica evidente o motivo de sua boa receptividade ao formato das revistas científicas. As vantagens originais dos periódicos científicos envolviam o alcance ampliado de seu conteúdo a leitores de diferentes localidades, o detalhamento de informações sobre a metodologia utilizada para coleta e análise dos dados das pesquisas, avaliação crítica do conteúdo por pares, atribuição de autoridade aos autores sobre os assuntos por eles discutidos e viabilidade da citação de estudos utilizados como referencial teórico dos saberes produzidos. Todas estas vantagens já atendiam às dificuldades encontradas pelos atores no exercício de seus papéis.

Tudo isso em um espaço até então ocupado somente pelas revistas científicas impressas. Foi somente no século XX, com o crescimento exponencial da produção informacional e o surgimento da comunicação em rede, que se começou a pensar na possibilidade de propagação de versões digitais dos periódicos impressos. Vale salientar que os avanços das indústrias baseadas no progresso do conhecimento já vinham sendo notados desde a segunda metade do século XIX, na qual expandiram-se as sociedades científicas e profissionais, as editoras comerciais e, paralelamente, suas atividades de publicação (Meadows, 1999).

O crescimento desenfreado do volume de informações compartilhadas após o fim da segunda guerra mundial apenas ratificou alguns dos problemas encontrados nas revistas científicas impressas, como: a demora para a publicação, a que Fjällbrant (1997) registrava a média de 8 meses entre a submissão do manuscrito e a impressão do artigo na revista; o alto custo despendido no funcionamento das editoras, tanto as universitárias quanto as comerciais, convertido em altas taxas de assinatura para acesso aos volumes publicados e o espaço limitado

para o armazenamento dos exemplares nas bibliotecas universitárias/de instituições de pesquisa.

Como cenário alternativo para esta situação, surgiam as primeiras publicações eletrônicas em bases de dados administradas por editoras comerciais já consolidadas, como a *Elsevier* e a *Springer*, fundadas em 1880 e 1842, respectivamente. Isso só foi possível devido ao crescimento dos custos de impressão nas décadas de 1980 e 1990 e a redução de ferramentas de computação, que já ocupavam as residências da classe média. Assim tiveram início as primeiras ofertas de conteúdo dos periódicos científicos em formato digital, primeiro com a automatização dos catálogos das bibliotecas, seguida da disponibilização de resumos dos artigos impressos e mais tarde com a publicação dos primeiros textos completos em edições eletrônicas (Fjällbrant, 1997).

Apesar de promissora, a adaptação das revistas científicas do formato impresso para o formato eletrônico enfrentava a resistência do seu público. Isso porque nos anos iniciais de sua produção, ainda não existia uma estrutura padronizada para sua diagramação e disponibilização. Os sistemas operacionais ainda estavam sendo desenvolvidos e nem todos os softwares de leitura de texto disponíveis na época conseguiam reter as configurações dos documentos recuperados na rede. Outra problemática significativa era a falta de um padrão técnico para a citação e referência dos trabalhos publicados em formato digital (Fjällbrant, 1997; Meadows, 1999).

Em geral, a falta de políticas e de infraestrutura consolidadas para a disseminação de publicações científicas em formato eletrônico fazia com que os estudos publicados somente neste meio não tivessem validade probatória para o currículo acadêmico-científico de pesquisadores institucionalizados e docentes universitários. Fjällbrant (1997) evidencia que, somando-se a esta desconfiança, outras três situações colocavam a credibilidade das publicações digitais em risco: 1) a restrição do acesso ao conteúdo publicado em rede, que até a década de 1990 se limitava a quem possuísse assinatura das bases de dados ou revistas disponíveis, dispusesse do acervo de uma biblioteca próxima ou mantivesse vínculo institucional com universidades e/ou instituições de pesquisa; 2) As restrições do direito de copyright passado dos autores às editoras científicas; e 3) O risco de plágio, uma vez que os textos digitados acabavam sendo mais vulneráveis à cópias e impressões não autorizadas.

É claro que, com a maturação da tecnologia computacional e de seus sistemas, grande parte dos problemas estruturais foram sendo sanados. Assim como novas políticas e padrões de normalização, indexação e citação para documentos digitalizados e nativos digitais foram sendo fixadas na comunidade acadêmico-científica.

Seguindo o que Meadows (1999) considerou como opinião de grande parte dos leitores, os componentes estruturais dos artigos científicos impressos foram mantidos passados para o formato eletrônico com quase total semelhança, exceto pela contagem de páginas e volumes e outras características relacionadas às revistas em si. A comunicação em rede tornou muito mais fácil e rápida a pesquisa colaborativa, bem com a avaliação dos artigos submetidos para publicação e a disponibilização dos trabalhos aceitos.

Por muito tempo era possível encontrar periódicos científicos que são publicados tanto em formato impresso quanto em formato digital, e até hoje esta realidade se faz presente. Nesse decurso, uma problemática que cruzou as fronteiras dos dois modelos de distribuição das revistas científicas, foi o preço elevado das taxas de assinatura paga pelo acesso a estas publicações. Até o estopim do Movimento *Open Access*, as limitações de acesso ao conhecimento científico em circulação causavam frustração nos pesquisadores que precisavam se manter atualizados da literatura produzida em suas áreas e não conseguiam vislumbrar muito mais do que lhes era recuperado nos acervos ao qual tinham à disposição. Nessa condição, teve início o paradigma científico moderno: a Ciência Aberta ou *Open Science*.

2.2 CIÊNCIA ABERTA E ACESSO ABERTO

Historicamente, a Ciência Aberta existe desde o século XVII, em que se deu o advento das revistas científicas. Nesse período a abertura dos processos científicos não fazia alusão à disponibilização gratuita do conteúdo publicado, mas se preocupava principalmente com o compartilhamento do conhecimento entre partes interessadas e a cooperação dos pesquisadores em prol do desenvolvimento científico (Green, 2017).

Isso porque, por muito tempo, os cientistas foram desencorajados a compartilhar seus achados, na tentativa de preservar seu prestígio e o de seus mestres e financiadores. Sem um sistema de comunicação científica estruturado na época, os pesquisadores reivindicavam a propriedade de suas descobertas através do envio de anagramas aos seus pares. Este método atuava como uma prevenção caso outros viessem a alcançar os mesmos achados posteriormente. De modo que, quando a descoberta estivesse amadurecida e começasse a ser difundida na sociedade, os anagramas fariam sentido e o financiamento da pesquisa em andamento já estaria assegurado (Bartling; Friesike, 2014).

Com o reconhecimento de que novos saberes poderiam ser gerados a partir de descobertas anteriores, a expansão da demanda pelo conhecimento fez com que academias, sociedades e comunidades científicas comesçassem a se formar, e os saberes antes restritos a

grupos seletos de estudiosos passou a ser transmitido a círculos cada vez maiores. (Green, 2017). Isso fez com que a ciência passasse a galgar espaços e meios de comunicação antes impensáveis, como o periódico científico, tornando-a mais “aberta” e acessível, de certa forma.

A Ciência Aberta observada nos dias atuais, carrega os princípios do movimento *open access*, cuja discussão se consolidou na década de 1990, três séculos após o surgimento das revistas científicas. Green (2017) aponta o projeto Gutenberg, de 1971, como marco inicial do movimento, pelo qual Michael S. Hart daria o pontapé inicial na digitalização de livros em domínio público para disponibilização em repositórios digitais. Duas décadas depois, outras organizações científicas passariam a seguir o modelo de repositórios de Hart, como viria a ser o caso do *arXiv.org*, fundado em 1991 pela *Cornell University*, um dos primeiros veículos de artigos científicos em acesso aberto.

A década de 1990 criou espaço para a emergência de novas tecnologias e projetos que fomentassem o acesso aberto. No fim da década, um projeto da Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP) em parceria com o Centro Latino-americano de Informação em Ciências da Saúde (BIREME) deu origem à *Scientific Electronic Library Online* (SciELO), disponível para o público a partir de 1998 (Fausto, 2013). A SciELO foi desenvolvida como uma biblioteca de acesso aberto, que realizava a indexação de revistas científicas publicadas por instituições de ensino superior e pesquisa, sociedades científicas e associações profissionais, com vista a promover o acesso da produção científica formal gerada nestes espaços (Packer, 2019).

O lançamento da SciELO e de outras bases de dados e periódicos científicos eletrônicos da época evidenciam que o movimento *open access* já vinha sendo defendido por iniciativas em áreas específicas, quando em 2002 foi lançada publicamente a *Budapest Open Access Initiative* (BOAI). A BOAI trouxe à tona o termo “acesso aberto”, como um conceito associado a todo tipo de literatura que estivesse disponível livremente nas redes públicas de internet, permitindo a leitura, download, cópia, distribuição, impressão, busca, conectividade por link, transferência em formato de dados através de softwares, sem quaisquer barreiras legais, financeiras ou técnicas (*Budapest Open Access Initiative*, 2002).

A declaração trazida por esta iniciativa, não somente dialogava com todas as áreas do conhecimento, como também convidava todos os atores envolvidos na comunicação científica - autores, editoras, bibliotecas, agências de fomento e universidades - a promover duas estratégias focais para o acesso aberto: 1) o autoarquivamento e 2) a publicação de revistas científicas abertas (*Budapest Open Access Initiative*, 2002). Com estas duas estratégias em mente, a BOAI formulava os parâmetros para as duas principais vias de *open access*, a verde e

a dourada, a primeira representada pelos repositórios temáticos e institucionais e a segunda pelas revistas científicas de editoras comerciais.

Os princípios da *Budapest Open Access Initiative* (2002) foram definidos sobre a crença de que a remoção das barreiras de acesso à literatura científica aceleraria os processos da pesquisa, enriqueceria a educação e aprendizagem, possibilitaria a democratização do conhecimento e tornaria todos os frutos da informação científica tão úteis quanto poderiam ser. Esta iniciativa tornou mais evidente a viabilidade econômica do acesso aberto, que deveria ser sustentado por agências de fomento e/ou instituições filantrópicas que pudessem investir em iniciativas de livre acesso ao conhecimento. Assim, em seu lançamento, a BOAI se comprometeu, com o auxílio do *Open Society Institute*⁵, a financiar outras iniciativas de acesso aberto e reafirmou a necessidade de que outras instituições investissem no movimento.

Pouco mais tarde, outras duas declarações surgiram para reforçar o que fora apresentado pela BOAI, a declaração de Bethesda (*Bethesda Statement on Open Access Publishing*), lançada em junho de 2003, e a declaração de Berlim (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*), assinada em outubro do mesmo ano. A primeira direcionada à renovação dos meios de publicação na Biomedicina, e a segunda voltada para as pesquisas em ciências gerais e humanidades. Juntas, a BOAI, a declaração de Bethesda e a de Berlim formulam os três Bs do acesso aberto e até hoje guiam as estratégias de consolidação da Ciência Aberta pelo mundo.

No Brasil, em 2005, foi lançado o Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Esta iniciativa reforçava o alcance do novo paradigma na comunicação científica nacional e abria as portas para a implementação das estratégias sugeridas pela Declaração de Berlim, com foco na criação de repositórios institucionais de acesso livre, no autoarquivamento e na publicação de revistas científicas abertas (Instituto Brasileiro de Informação em Ciência e Tecnologia, 2005).

No ano seguinte, o IBICT lançava o *Oasis.br* (*Open Access Scholarly Information System*), um diretório que centralizava o acesso a uma gama de revistas científicas e repositórios brasileiros. Quatro anos mais tarde, lançou a primeira iniciativa brasileira em nível governamental federal, o Edital de Chamada FINEP/PCAL/XBDB N°002/2009. Este tinha o objetivo de apoiar os projetos de implementação de repositórios em instituições públicas de

⁵ Atualmente *Open Society Foundations*, é uma rede internacional filantrópica fundada por George Soros em 1993 que até hoje apoia financeiramente iniciativas em prol da justiça, educação, saúde pública e a mídia independente.

ensino e pesquisa e sua integração ao Oasis.Br, com vistas a possibilitar o registro e proporcionar maior visibilidade da produção científica destas instituições.

Em 2012 foi lançada a BOAI10⁶, que apontava um novo panorama, no qual podia-se observar que o movimento *open access*, agora bem estabelecido, se expandia progressivamente em todas as áreas do conhecimento e ganhava cada vez mais visibilidade em instâncias técnicas, econômicas e legais. Nesse novo horizonte, foram estabelecidas quatro novas recomendações para a difusão do acesso aberto por mais 10 anos.

A primeira recomendação da BOAI10 teve como foco a criação de políticas em instituições de ensino superior, pesquisa e agências de fomento que garantissem que pelo menos uma versão dos futuros artigos científicos de seus membros fosse depositada em repositório. A segunda recomendação foi pautada no licenciamento das obras publicadas em acesso aberto, na qual se aconselha a licença *Creative Commons CC BY*⁷. A terceira recomendação, diretamente relacionada com a primeira, propunha que todas as instituições de ensino superior possuísem um repositório aberto, fizessem parte de um consórcio de repositório ou fizessem parceria com repositórios externos já consolidados. A quarta e última recomendação se preocupava com a conscientização dos atores envolvidos no processo de publicação quanto aos padrões de conduta para publicação em acesso aberto (*Budapest Open Access Initiative*, 2012).

O avolumamento de questões abrangidas na nova BOAI demonstravam não somente o avanço do paradigma da Ciência Aberta moderna e progresso da infraestrutura favorável ao acesso aberto, como também antecipava a expansão dos interesses do movimento. Dois anos depois desta atualização, Fecher e Friesike (2014) asseguravam que dificilmente haveria algum pesquisador que ainda não tivesse se deparado com o termo “Ciência Aberta”, e esclareceram que este se tratava de um termo guarda-chuva, tendo em vista a diversidade dos pressupostos que poderiam ser gerados a partir da ideia de “abertura” do conhecimento.

As recomendações da UNESCO sobre Ciência Aberta a definem como

um construto inclusivo que combina vários movimentos e práticas que têm o objetivo de disponibilizar abertamente conhecimento científico multilíngue, torná-lo acessível e reutilizável para todos, aumentar as colaborações científicas e o compartilhamento de informações para o benefício da ciência e da sociedade, e abrir os processos de criação, avaliação e comunicação do conhecimento científico a atores da sociedade, além da comunidade científica tradicional. Abrange todas as disciplinas científicas e todos os aspectos das práticas acadêmicas, incluindo ciências básicas e aplicadas, ciências naturais, sociais e humanas, e se baseia nos seguintes pilares-chave: conhecimento científico aberto, infraestrutura científica aberta, comunicação científica,

⁶ Atualização referente aos 10 anos do manifesto publicado em 2002.

⁷ Modelo de licenciamento de conteúdos abertos, que permite a distribuição, remixagem, adaptação e criação a partir de conteúdo já publicado, desde que se atribua o devido crédito ao original. (Creative Commons, [201-?]).

envolvimento aberto dos atores sociais e diálogo aberto com outros sistemas de conhecimento (Organização das Nações Unidas para a Educação, a Ciência e a Cultura, 2022, p. 7).

Nessa esfera, surgem cinco escolas de pensamento com perspectivas diferentes, sobre a Ciência Aberta:

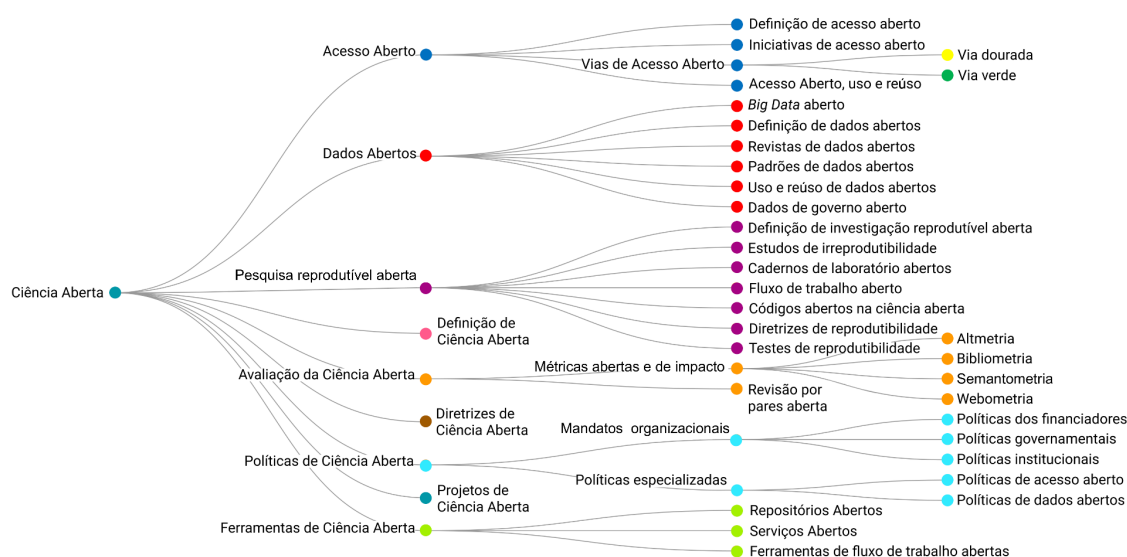
A escola pública (*public school*) [que] busca garantir não apenas a acessibilidade do processo de pesquisa, mas a compreensibilidade de seus resultados, recusando o hermetismo e promovendo a clareza e a comunicação do conhecimento científico; A escola democrática (*democratic school*) considera o acesso ao conhecimento um direito humano, condição que se torna ainda mais desejável quando a pesquisa científica conta com financiamento público; A escola pragmática (*pragmatic school*) trabalha com uma noção de aberto mais próxima da inovação aberta e vislumbra que o processo científico pode ser otimizado pela incorporação do conhecimento externo e a colaboração através de ferramentas on-line, o que permitiria acesso a diversos tipos de conhecimentos e expertises; A escola da infraestrutura (*infrastructure school*) foca nas possibilidades e nos desafios tecnológicos, especialmente os de infraestrutura, necessários às práticas emergentes da Ciência Aberta, com destaque para duas tendências: a computação distribuída através da conexão de diversos computadores para formar uma rede de alto desempenho no processamento de pesquisas com uso intensivo de dados; e a constituição de redes sociais de colaboração para promover maior interação e colaboração entre cientistas. Por fim, a escola das métricas (*measurement school*) busca criar novos modos de mensurar a produção científica, uma vez que esta tende a migrar para ambientes on-line e adotar novos formatos de publicação, para os quais tradicionalmente não se atribuía qualquer tipo de avaliação (Friesike, 2013 *apud* Albagli; Clinio; Raychtock, 2014, p. 437).

A divisão proposta por Fecher e Friesike (2014) atestou o crescimento do movimento de acesso aberto em diferentes instâncias do conhecimento. Nessa acepção, a escola de infraestrutura abarca o desenvolvimento de ferramentas tecnológicas (*softwares*) que facilitam o armazenamento e garantem a preservação do conhecimento científico em rede, bem como a criação de redes sociais e canais de comunicação interconectados, que possibilitem a sua tramitação. Enquanto isso, a escola pública busca envolver cidadãos comuns no processo científico, de modo que não sejam passivos no processo científico, mas que possam contribuir ativamente com ele. A escola de métricas, busca por novas maneiras de mensurar a dinâmica e os efeitos do conhecimento produzido na sociedade. Ao mesmo tempo, a escola democrática compreende que o acesso ao conhecimento é um direito igualitário de todas as pessoas, independentemente do status e da posição que ocupam dentro ou fora do meio acadêmico-científico. A escola pragmática, última desta composição, se solidifica sobre o ideal da cooperação, fortalecendo a ideia de que o conhecimento não se constrói sozinho.

Em suma, todas essas escolas carregam um pouco dos princípios da BOAI de 2002 que viriam a ser mais bem desenvolvidos a partir da explanação das vertentes descritas na

Taxonomia FOSTER. A FOSTER é uma plataforma criada em 2017 com apoio financeiro do projeto *Horizon 2020*, da União Europeia. O projeto *Horizon 2020*, lançado em 2014, direcionou cerca de 77 bilhões de euros em financiamento para produção científica e tecnológica. Este projeto buscava remover as barreiras do conhecimento em inovação e facilitar o trabalho conjunto dos setores público e privado (Horizon 2020, [2014]). A FOSTER, por sua vez, um dos produtos do projeto, reuniu uma série de recursos de aprendizado sobre Ciência Aberta, dentre as quais se destaca a taxonomia criada para evidenciar as subdivisões dos campos de atuação da Ciência Aberta (Figura 1).

Figura 1 - Taxonomia FOSTER para a Ciência Aberta



Fonte: Pontika *et al.* (2015). Traduzida por Ribeiro; Silveira; Santos (2020).

Como fora representado na Figura 1, os dados abertos e, consecutivamente, as revistas de dados abertos - objeto de interesse desta pesquisa - fazem parte do “guarda-chuva” conceitual da Ciência Aberta. Atualmente, muito se tem discutido sobre o valor dos conjuntos de dados produzidos ou coletados em pesquisas científicas e sua aplicabilidade em novos estudos, e cada vez mais os dados de pesquisa passam a ocupar um relevante espaço na produção do conhecimento científico.

2.3 DADOS ABERTOS DE PESQUISA

Considera-se como dado de pesquisa “todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica.” (Sales; Sayão, 2019, p. 36).

Colocando de outra forma, os dados de pesquisa são os insumos básicos utilizados para comprovar ou refutar a hipótese de estudos científicos, de modo que todo conhecimento científico é gerado a partir de um ou mais tipos de dados.

Como colocado no último parágrafo da seção 2.2, o valor dos conjuntos de dados produzidos ou coletados em pesquisas e sua capacidade de reúso, os tem colocado em alta conta ao se pensar sobre novas práticas e modelos que se adequem ao panorama crescente da Ciência Aberta. Mas, diferente do que se possa pensar, nem todo dado de pesquisa pode vir a se tornar um dado aberto. Isso ocorre porque, em algumas situações, os dados podem estar conectados a uma pesquisa privada ou a uma patente que demande um período de embargo ou até impeça sua divulgação.

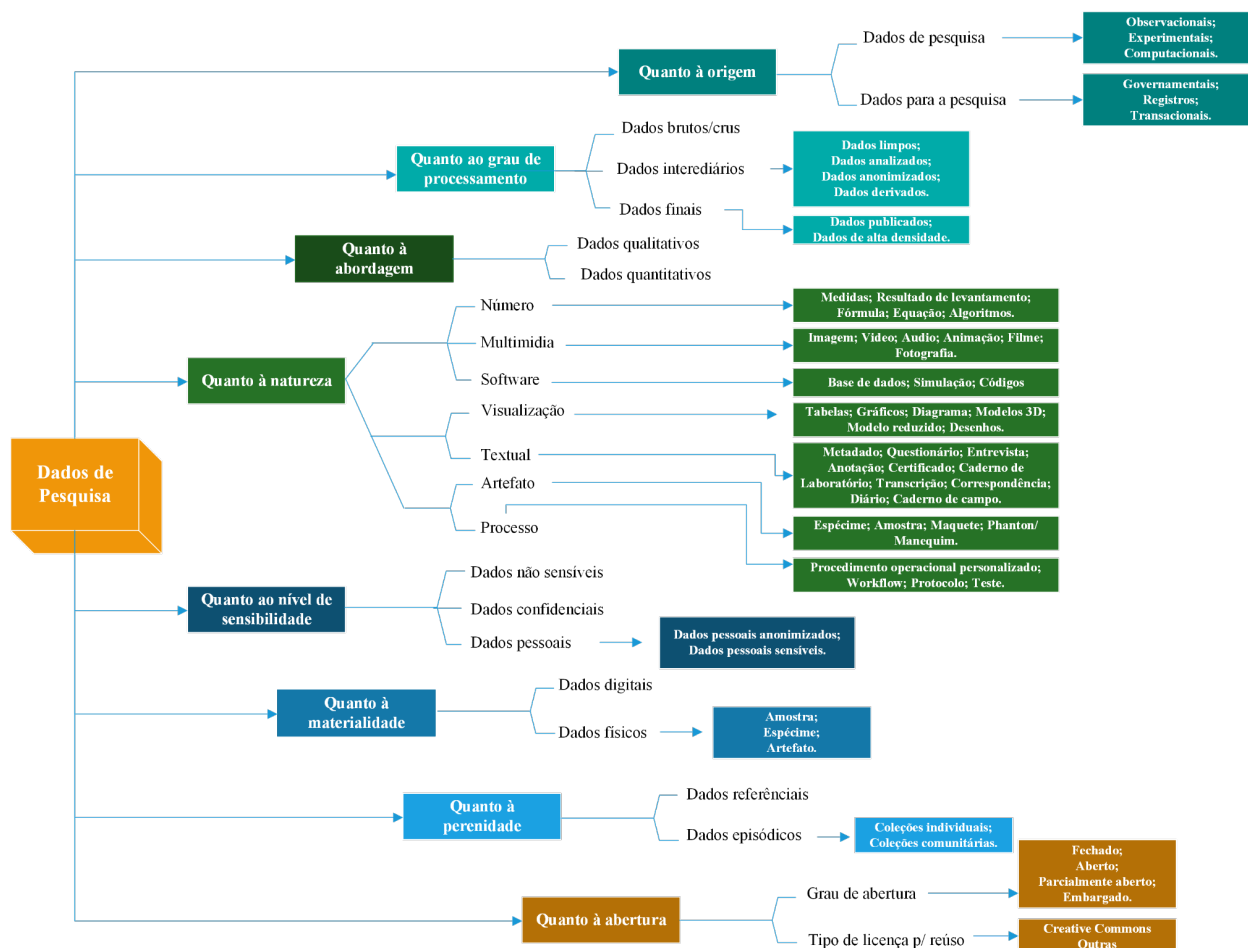
Em 2007, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), além de definir o que seriam dados de pesquisa, esclareceu o que seriam dados de pesquisa oriundos de financiamento público, obtido através de pesquisas conduzidas por instituições, órgãos ou departamentos governamentais, realizadas com fundos públicos fornecidos por qualquer nível de governo (municipal, estadual ou federal).

Nessa lógica, Sales e Sayão (2019) explicam que a heterogeneidade é intrínseca aos dados de pesquisa e reconhecer as diferenças entre os diversos tipos de dados se torna crucial para a gestão de dados de pesquisa e o funcionamento do ciclo de vida dos dados. Como resultado, para facilitar a compreensão dos diversos tipos de dados de pesquisa encontrados em investigações científicas, os autores desenvolveram uma taxonomia que os segrega segundo sete aspectos: origem, grau de processamento, abordagem, natureza, sensibilidade, materialidade, perenidade e abertura (Sales; Sayão, 2019).

Nessa acepção, a origem dos dados estabelece a profundidade e a abrangência da gestão que se deve aplicar para cada categoria subsequente dos dados abrangidos. O grau de processamento varia de acordo com o fluxo de pesquisa de cada área do conhecimento. A abordagem da pesquisa determina o tratamento distinto que os dados precisam receber de acordo com a metodologia escolhida para o desenvolvimento da pesquisa. A natureza dos dados elucida a diversidade e a heterogeneidade dos dados. O nível de sensibilidade define os níveis de abertura e compartilhamento possíveis para os dados na etapa de gestão. A materialidade se preocupa com a forma como os dados se materializam. A perenidade se inclina sobre a relevância dos dados dentro de sua área de origem e por quanto tempo estes podem ser preservados por sua utilidade. Por fim, a abertura aborda questões relacionadas ao nível de sensibilidade dos dados, propriedade intelectual, interesses comerciais, patentes, etc. (Sales; Sayão, 2019).

Uma adaptação da ilustração da Taxonomia de dados de pesquisa de Sales e Sayão é apresentada na Figura 2.

Figura 2 - Taxonomia dos dados de pesquisa



Fonte: Adaptado de Sales e Sayão (2019).

Os dados abertos de pesquisa, como apontado na figura acima, não precisam estar completamente abertos, mas para que o ciclo de vida dos dados se reproduza de forma eficaz é desejável que eles estejam. Os princípios da OCDE para o acesso aos dados de pesquisa produzidos ou coletados com auxílio de financiamento público listam como benefícios do livre compartilhamento de dados de pesquisa, dentre outros: 1) a boa administração do investimento público em informações factíveis; 2) o aumento da cooperação internacional; 3) o reforço da investigação científica aberta; 4) o encorajamento da diversidade de análises e opiniões; 5) a promoção de novas pesquisas; 6) a exploração de tópicos não previstos pelos estudos primários; e 7) a criação de novos conjuntos de dados quando os dados provenientes de múltiplas fontes são combinados (Organização para a Cooperação e Desenvolvimento Econômico, 2007).

O incentivo ao compartilhamento de dados em acesso aberto já podia ser observado desde a Declaração de Berlim (2003), que inclui a publicação de dados não processados e metadados, como contribuições necessárias para o movimento em prol do acesso livre ao conteúdo científico. Nesse sentido, Masuzzo e Martens (2017) definem dados abertos de pesquisa como aqueles que estão disponíveis na Internet, com licenciamento que permite a qualquer usuário baixar, copiar, tratar, reprocessar ou reusá-los para qualquer outra finalidade sem quaisquer barreiras de acesso.

Elemento essencial para a reformulação do *copyright* no contexto da Ciência Aberta, com impacto significativo na definição dos níveis de abertura dos dados de pesquisa, a *Creative Commons* (CC) foi concebida em 2001, por Lawrence Lessig, Hal Abelson e Eric Eldred como uma organização voltada para a concessão de licenças para uso legal de materiais em acesso aberto (Creative Commons, [201-?]). As licenças CC foram estruturadas em diferentes modelos de permissões, que poderiam ser combinadas e aplicadas de acordo com as necessidades de cada autor ou produtor de conteúdo aberto, conferindo-lhes a credibilidade e segurança de que precisavam para permitir o compartilhamento, uso e reúso de seu trabalho, em conformidade com princípios legais e éticos de pesquisa e a garantia dos direitos de propriedade intelectual sobre os trabalhos licenciados.

Inicialmente, o modelo criado por Abelson, Eldred e Lessig, possuía quatro modalidades, mas atualmente a organização conta com sete opções de licenças mais ou menos abrangentes (Quadro 1).

Quadro 1 - Licenças Creative Commons

CC BY	Permite que outros distribuam remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original.
CC BY-SA	Permite que outros remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem as novas criações sob termos idênticos.
CC BY-ND	Permite que outros reutilizem os trabalhos para quaisquer propósitos, incluindo uso comercial, desde que não sejam feitas adaptações e que lhe atribuam o devido crédito pela criação original.
CC BY-NC	Permite que outros remixem adaptem e criem a partir do seu trabalho para fins não comerciais, e embora os novos trabalhos tenham de lhe atribuir o devido crédito e não possam ser usados para fins comerciais, os usuários não têm de licenciar esses trabalhos derivados sob os mesmos termos.
CC BY-NC-SA	Permite que outros remixem adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam o devido crédito e que licenciem as novas criações sob termos idênticos.

CC BY-NC-ND	Esta é a mais restritiva das nossas seis licenças principais, só permitindo que outros façam download dos seus trabalhos e os compartilhem desde que atribuam crédito a você, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.
CC0	Dedicação ao domínio público. Permite copiar, modificar, distribuir e executar o trabalho, mesmo para fins comerciais, tudo sem pedir autorização.

Fonte: Adaptado de *Creative Commons* ([201-?]).

Considerando as modalidades de licenciamento expostas acima, espera-se que os dados abertos de pesquisa sejam veiculados com licença CC BY ou CC0. Diante do exposto, fica evidente que o compartilhamento e o reúso de dados de pesquisa são essenciais para se caminhar rumo ao estabelecimento da Ciência Aberta, mas para que os dados estejam prontos para serem divulgados e utilizados em outras circunstâncias que não a de origem, outras etapas do ciclo de vida dos dados precisam ser contempladas em seu tratamento.

O ciclo de vida dos dados de pesquisa pode variar em abrangência e conceito, a depender do cenário em que é abordado. De forma concisa, este se divide em seis etapas (Figura 3): 1) o planejamento da pesquisa; 2) a coleta de dados; 3) o processamento e análise dos dados; 4) a publicação e compartilhamento dos dados; 5) a preservação dos dados publicados ou compartilhados; e 6) o reúso dos dados (*United Kingdom Data Service*, [201-?] apud Aventurier, 2016).

Figura 3 - Ciclo de vida dos dados de pesquisa



Fonte: United Kingdom Data Service apud Aventurier (2016).

A primeira etapa do ciclo de vida dos dados de pesquisa cobre o início da pesquisa, onde é criado o Plano de Gestão de Dados (PGD), conforme detalhado por Aventurier (2016). Sayão e Sales (2015), por sua vez descrevem como os dados da pesquisa devem ser coletados ou gerados, processados, preservados, compartilhados ou publicados durante ou após a divulgação do produto final da pesquisa. Esta etapa estrutura e antecede a etapa seguinte, que é a coleta ou

produção dos dados de pesquisa. A etapa de coleta dos dados pode compreender a aplicação de instrumentos que auxiliem no recolhimento dos dados do campo a ser explorado. São exemplos de instrumentos utilizados com esta finalidade: questionários, entrevistas, registros bibliográficos, registros midiáticos, etc. Nesta etapa os dados ainda são organizados de modo a facilitar o andamento da terceira etapa, a análise e processamento dos dados.

A terceira etapa apontada aborda os métodos empregados no processamento e análise dos dados recuperados, a fim de interpretá-los dentro do contexto explorado e registrar os resultados que possam ser extraídos a partir de sua interpretação. Os resultados obtidos são então organizados em um documento para publicação, e após concluída esta etapa os dados utilizados devem ser armazenados em ambiente seguro para sua preservação. A quarta etapa, que trata da publicação e do compartilhamento dos dados, conversa com a quinta etapa, de preservação dos dados. Juntas, estas etapas recomendam a disponibilização dos dados em formato *open source*, que facilite sua reutilização, e prevê a criação de um *backup* dos dados e a elaboração de documentação (metadados) que auxiliem em sua recuperação.

A sexta e última etapa definida por é a do reúso dos dados, na qual os dados compartilhados são acompanhados através de citações e referências em outras pesquisas, dando continuidade ao seu ciclo de vida. Para que esta etapa ocorra, contudo, é preciso ter em conta outro importante elemento: o acesso aos dados. O acesso aos dados diz respeito ao nível de abertura, ao qual nos referimos ao abordarmos a taxonomia dos dados de pesquisa, que pode variar de acordo com sua origem e uso final.

Em 2015, Candela *et al.* indicavam que, embora houvesse um acordo sobre os benefícios do compartilhamento e do reúso de dados de pesquisa, existiam barreiras metodológicas, legais e técnicas que faziam com que pesquisadores não se sentissem inclinados a publicar seus dados abertamente. Assim, foi proposto que se criassem revistas científicas voltadas para a publicação de dados, *data journals*, na tentativa de romper as barreiras existentes.

Segundo Santos e Nassi-Calò (2020), *data journals* são:

[...] veículos que se destinam exclusivamente à publicação de artigos de dados. Normalmente, esses periódicos fornecem modelos que orientam os pesquisadores sobre como descrever adequadamente os dados e onde depositá-los. Esses periódicos podem utilizar guias próprios para a apresentação e descrição dos dados, mas podem também adotar critérios específicos de uma área do conhecimento ou disciplina (Santos; Nassi-Calò, 2020, p. 30).

Com o surgimento dos *data journals*, foram organizados os artigos de dados, *data papers*, orientados para apresentação dos metadados que descrevem a estrutura, o formato, os métodos utilizados na coleta, organização e processamento dos dados. Estes também incluem a

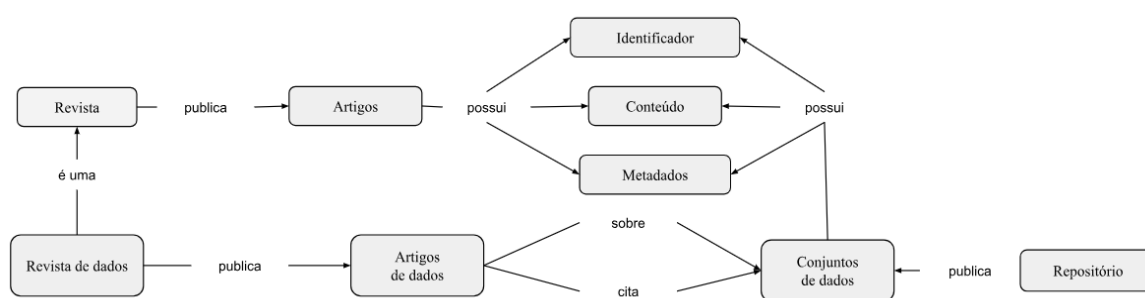
autoridade responsável pelos dados, o licenciamento sob o qual estão disponibilizados o identificador persistente e seu potencial de reutilização (Candela *et al.*, 2015).

É importante destacar que a criação de revistas e artigos científicos orientadas aos dados de pesquisa não invalida os repositórios institucionais, temáticos e de dados. De forma contrária, estes modelos de publicação e compartilhamento atuam em conjunto na promoção do acesso aberto aos dados de pesquisa. Essencialmente, a publicação de dados de pesquisa em periódicos científicos e sob a forma de artigos de dados demanda que os conjuntos de dados tenham sido previamente compartilhados em um repositório e possuam um identificador persistente que garanta sua fácil recuperação e acesso.

2.4 REVISTAS E ARTIGOS DE DADOS

Não se sabe exatamente quando as revistas de dados começaram a ser publicadas, mas a partir do estudo de Candela *et al.* (2015), podemos considerar que o desenvolvimento de um modelo orientado à disseminação dos dados abertos tenha ocorrido em período similar. Ao tentar resgatar uma definição do que vêm a ser revistas e artigos de dados, é possível observar um consenso na literatura, que concebe as revistas de dados ou *data journals* como periódicos dedicados parcial ou inteiramente à publicação de conjuntos de dados em documentos estruturados regularmente (artigos de dados ou *data papers*) e submetidos a processo de revisão por pares, que comprove seu rigor científico, vide Figura 4 (Candela *et al.*, 2015; Chavan; Penev, 2011; García-garcía *et al.*, 2015; Walters, 2020).

Figura 4 - Organograma do artigo de dados



Fonte: Adaptado de Candela *et al.* (2015).

Para fins desta pesquisa, compreende-se as revistas parcialmente dedicadas à publicação de dados como revistas “híbridas”, cuja abrangência do escopo também aceita artigos de dados, mas não se restringe a eles. Enquanto entende-se como revista de dados, especialmente, as que são orientadas unicamente à publicação de artigos de dados.

Segundo Santos e Nassi-Calò (2020, p. 30), *data journals*⁸ são:

[...] veículos que se destinam exclusivamente à publicação de artigos de dados. Normalmente, esses periódicos fornecem modelos que orientam os pesquisadores sobre como descrever adequadamente os dados e onde depositá-los. Esses periódicos podem utilizar guias próprios para a apresentação e descrição dos dados, mas podem também adotar critérios específicos de uma área do conhecimento ou disciplina.

Nesse encadeamento, os artigos de dados são similares aos artigos de pesquisa convencionais, e como tais, podem ser indexados e recuperados nas bases de dados tradicionais utilizadas para a disponibilização de periódicos de assuntos gerais ou especializados. O que os diferencia do artigo convencional é a intenção que os artigos de dados possuem de facilitar a reutilização e a reprodução de conjuntos de dados em novas pesquisas, promovendo sua livre exploração e citação, e descrevendo os métodos e protocolos adotados na curadoria dos dados a que se referem.

No que diz respeito ao procedimento de revisão por pares dos artigos de dados, além de avaliar os elementos básicos de um artigo científico: qualidade, consistência, utilidade, contribuição para o desenvolvimento científico etc. também devem ser consideradas: a metodologia empregada na coleta dos dados, a escolha da sistematização adotada em sua apresentação e a adequação dos metadados que os descrevem (Candela *et al.*, 2015; Torino; Roa-martínez; Vidotti, 2020). Uma vez aceitos pelos avaliadores, estes artigos poderão ser veiculados de acordo com as normas e padrões das revistas para as quais tenham sido submetidos.

Os artigos de dados funcionam como uma ferramenta para facilitar a recuperação e a compreensão do conteúdo de conjuntos de dados que estão disponíveis em repositórios e, consequentemente, motivar pesquisadores a compartilharem seus dados pelo reconhecimento dos artigos de dados como produção científica creditável e citável (Candela *et al.*, 2015; Kim, 2020). Contudo, para que isso ocorra, os conjuntos de dados compartilhados não necessariamente precisam estar completamente abertos. Assim, para garantir a recuperação destes dados e uso em sua integridade, foi estabelecido que, idealmente, os dados de pesquisa compartilhados deveriam estar de acordo com os princípios FAIR (*Findable, Accessible, Interoperable and Reusable*) que garantiria sua preservação e segurança, tornando-os recuperáveis, acessíveis, interoperáveis e reusáveis (Candela *et al.*, 2015; Torino; Roa-Martínez; Vidotti, 2020).

⁸ Os termos “revistas de dados” e “*data journals*” são empregados indistintamente nesta pesquisa.

O conteúdo da base dos artigos de dados é a apresentação dos metadados que descrevem os conjuntos de dados, bem como a descrição metódica de sua estrutura e formato, e a metodologia utilizada na coleta, organização e processamento dos dados. Estes também incluem a autoridade responsável pelos dados, o licenciamento sob o qual estão disponibilizados, um DOI e sugestão de potencial reuso (Candela *et al.*, 2015; Kim, 2020).

Um número crescente de artigos de dados já é encontrado em revistas híbridas ou em revistas de dados do exterior, como o *Data in Brief* e o *Scientific Data* que até o fim do ano de 2021 possuíam, respectivamente, um total de 6.009 e 1.478 artigos de dados indexados na *Web of Science* (Araujo; Principe; Alencar, 2021).

Em revisão anterior, realizada em maio de 2022, buscou-se identificar na literatura a existência de estudos que tratassem sobre a publicação de dados de pesquisa em periódicos científicos, suas principais características e o panorama dessas publicações no meio acadêmico-científico (Araujo; Principe; Alencar, 2021). Os resultados desta pesquisa evidenciaram a dificuldade de se mensurar o impacto da publicação de artigos de dados e periódicos de dados no panorama atual da comunicação científica aberta, uma vez que ainda não é possível encontrar (a) um padrão para a estrutura destes artigos, (b) um modelo de avaliação por pares, que comprove a qualidade dos dados compartilhados, ou ainda, (c) um programa estruturado para conscientizar pesquisadores quanto aos benefícios da produção de artigos de dados, para a divulgação e o reconhecimento de suas pesquisas (Chen, 2016; Jefferies *et al.*, 2019; Kim, 2020; Schöpfel *et al.*, 2019).

Nota-se a conformidade das conclusões obtidas pelos autores quanto à utilidade dos artigos de dados, que segundo os estudos podem: aumentar a confiança dos pesquisadores no compartilhamento de dados, devido ao aumento do reconhecimento que pode ser adquirido com o aumento de citações relacionadas aos seus trabalhos; ampliar a visibilidade dos aspectos metodológicos associados à pesquisa e à visibilidade de dados que não poderiam ser compartilhados no sistema tradicional de publicação (dados negativos e dados sobressalentes); e a percepção de que a avaliação por pares realizada nestas publicações pode validar a clareza e a credibilidade dos conjuntos de dados compartilhados (Chen, 2016; Jefferies *et al.*, 2019; Jiao; Darch, 2020; Kim, 2020; Li; Jiao, 2021; Schöpfel *et al.*, 2019).

Estes aspectos não só corroboram com as boas práticas de compartilhamento, como também favorecem a reutilização eficiente dos dados relacionados. Neste cenário, para que sejam evitados transtornos e más aplicações na publicação de artigos de dados, é preciso que os editores dos periódicos de dados levem em consideração, essencialmente, as informações contextuais necessárias para fomentar o reuso de dados, pensando sobre quais dados podem ter

seu compartilhamento atrelado a um artigo de dados e como se pode avaliar o reuso desses dados, rastreando seu impacto em publicações subjacentes (Jiao; Darch, 2020).

Verificou-se que novos aprofundamentos acerca da publicação de artigos de dados já vêm ganhando espaço nas publicações recentes sobre o tema (Jefferies *et al.*, 2019; Senderov; Georgiev; Penev, 2016), elaborando novas soluções e possibilidades que facilitam a interação de pesquisadores com o modelo de publicação. Dentre outras questões levantadas pelos autores dos estudos incluídos nesta revisão, a falta de padronização dos conteúdos dos artigos de dados pode gerar incertezas sobre os benefícios de sua publicação, que além de custosa aos autores, pode não lhes trazer o retorno esperado.

Diferenças encontradas no foco dos procedimentos de revisão por pares, no conjunto de metadados a ser publicado e no aprofundamento dos procedimentos metodológicos de coleta e tratamento dos dados e sugestões de estratégias de reuso dos mesmos, podem levantar questionamentos sobre a utilidade real destas publicações. Para contornar estes e outros problemas, é preciso que estas publicações sejam estruturadas de forma estratégica para que possam atingir toda a sua capacidade no fomento das práticas de compartilhamento e reuso de dados de pesquisa e no aumento do reconhecimento do trabalho de pesquisadores que se dedicam à coleta e tratamento de dados, conquistando, desta maneira, um público cada vez maior de autores e usuários interessados em vincular os dados que compartilham a um artigo de dados publicável.

Uma gama de estudos sobre o compartilhamento e a publicação de dados de pesquisa vem sendo desenvolvida no exterior (Callaghan *et al.*, 2012; Chavan; Penev, 2011; Organização para a Cooperação e Desenvolvimento Econômico, 2007). No campo científico, acredita-se que os artigos de dados podem promover o compartilhamento e o reuso de dados de pesquisa e aumentem a visibilidade e o reconhecimento dos coletores de dados científicos, protegendo seus direitos de autor e atestando a credibilidade da matéria prima de seus estudos. Aos editores científicos, a recepção de artigos de dados pode aumentar o número de materiais publicados e impactar na reputação e divulgação de suas revistas. No Brasil, principalmente, a existência de revistas híbridas que aceitem *data papers* ou de *data journals* pode reduzir o custo despendido pelos autores brasileiros para sua publicação em periódicos internacionais.

Além disso, o aumento de pesquisas que utilizem os dados disponibilizados em repositórios e descritos detalhadamente em *data papers* pode impactar direta ou indiretamente no desenvolvimento da sociedade, pelos efeitos gerados a partir dos conhecimentos desenvolvidos pelo reuso de dados. Um exemplo recente do impacto do compartilhamento e

reúso de dados na sociedade global foi a criação de vacinas em tempo recorde para o combate à COVID 19.

Os artigos de dados ainda podem facilitar a recuperação dos dados que anteriormente poderiam demandar maior esforço para serem localizados, como os dados de resultados negativos, tornando-os facilmente recuperáveis e, conseqüentemente, reutilizáveis pelas partes interessadas.

Em estudo sobre os motivos da não publicação de resultados negativos por pesquisadores, Sayão e Sales (2020) apontam que, apesar de serem parte comum do cotidiano dos laboratórios, e constituírem uma parte relevante da integralidade dos fluxos da investigação científica, os resultados negativos raramente são publicados em periódicos científicos e/ou outros meios de comunicação científica. Nesse quadro, os autores refletem seis barreiras comuns que impedem pesquisadores de publicar os dados/resultados negativos de seus estudos (Quadro 2).

Quadro 2 - Barreiras encontradas na publicação de resultados negativos

BARREIRAS	IMPLICAÇÕES
Contextuais	Impostas pelo próprio contexto da ciência como fonte de reconhecimento e bons resultados, no qual são valorizados principalmente resultados que confirmam uma tese ou hipótese.
Individuais e Cognitivas	Mesmo diante da confirmação de que os resultados negativos são uma ocorrência natural do cotidiano científico, os pesquisadores ainda se mantêm relutantes a publicar os resultados inesperados de seus estudos, observando esta prática como um desperdício de recursos e tempo.
Profissionais	Relacionadas à percepção real e imaginária de que a publicação de resultados negativos pode impactar de forma ruim sua carreira profissional e possíveis financiamentos de pesquisa.
Organizacionais	Dizem respeito às métricas das instituições de fomento, sociedades científicas, etc. que prestigiam publicações de alto impacto e com grande número de citações. Conseqüentemente reduzindo a chance de recompensar estudos de resultados negativos, que dificilmente conquistam visibilidade em publicações científicas de impacto.
Acadêmicas	Tratam do julgamento crítico dos processos de pesquisa que geram resultados negativos e seus efeitos na avaliação por pares, que geralmente observam esses resultados como insólitos e “mal-sucedidos”.
Editoriais	Relacionadas à escassez de infraestruturas editoriais para a publicação de resultados negativos e o baixo fator de impacto das publicações existentes que atendam a este escopo.

Fonte: Adaptado de Sayão e Sales (2020)

No Brasil, embora a literatura sobre estes temas esteja em construção, a publicação de artigos de dados ainda demanda uma infraestrutura que os periódicos nacionais ainda não detêm, do ponto de vista político, cultural e editorial, o que induz ao uso de periódicos do exterior para a publicação de dados de pesquisa. Deste modo, os autores brasileiros têm buscado periódicos híbridos, que também aceitam a publicação de artigos de dados, ou revistas de dados

internacionais como alternativa para a publicação dos seus artigos de dados (Araujo; Príncipe; Alencar, 2021).

Em trabalho anterior, verificou-se que as políticas de acesso aberto têm sido aplicadas progressivamente pelas editoras de revistas científicas brasileiras e que a infraestrutura para garantir seu cumprimento ainda está em construção (Araujo; Príncipe, 2021). Diante disso, entende-se porque a publicação de artigos de dados ainda não seja uma realidade consolidada nos periódicos científicos do Brasil.

Desse modo, a investigação da perspectiva dos editores científicos brasileiros quanto à publicação de artigos de dados pode oferecer um vislumbre de como estamos avançando em direção à sua difusão no Brasil e o aprofundamento da pesquisa sobre os formatos existentes destes artigos internacionalmente, pode orientar como melhor conduzir este processo.

Seguidamente, a seção 3 apresenta os procedimentos metodológicos adotados para o alcance e análise dos resultados da presente pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

Esta seção descreve os procedimentos metodológicos utilizados para alcançar os objetivos desta pesquisa. Nesse sentido, são esclarecidos o método, as técnicas e as fontes consultadas e aplicadas na coleta, tratamento e análise dos dados utilizados. Neste processo, foram adotadas de abordagem quantitativa e qualitativa para análise e exposição dos resultados obtidos.

A abordagem qualitativa é tida como uma maneira eficaz de se fazer compreender os fenômenos sociais de um dado contexto. Ao mesmo tempo que a abordagem quantitativa tem como foco a quantificação da coleta e da análise dos dados, sendo empregada no sentido de anunciar os percentuais observados nas amostras coletadas (Duarte *et al.*, 2009).

Para tanto, utilizou-se como abordagem quanti-qualitativa a análise de conteúdo de Bardin, em sua 70ª edição, pela adaptabilidade de seu emprego em estudos de natureza e formatos distintos. Este método utiliza de técnicas e procedimentos objetivos de análise a partir da descrição de informações não essencialmente quantificáveis para lhes atribuir sentido estruturado dentro de um contexto (Bardin, 2016).

A análise de conteúdo se divide em três etapas: 1) a pré-análise; 2) a exploração do material; 3) o desenvolvimento e interpretação dos resultados. A primeira etapa estabelece a delimitação da pesquisa em construção, os critérios de inclusão e exclusão que nortearão a seleção do conteúdo a ser explorado. A segunda etapa cuida da organização sistemática do conteúdo selecionado para facilitar o tratamento e a interpretação dos resultados, aplicáveis na última etapa da análise, que podem se apresentar em sínteses descritivas ou recortes estatísticos com base na aferição da frequência e uso de determinadas expressões.

3.1 PRÉ-ANÁLISE

Para dar início aos procedimentos utilizados e estabelecer os parâmetros da etapa de pré-análise desta pesquisa, foi redigido, em 22 de maio de 2022, um Plano de Gestão de Dados - PGD (Apêndice B) com o auxílio da ferramenta *DMPTool*⁹, serviço da *California Digital Library*. Trata-se de um aplicativo *online*, gratuito e de código aberto que facilita a criação de planos de gestão de dados para pesquisadores do mundo todo, em diversas áreas do conhecimento.

⁹ Disponível em: <https://dmptool.org/>. Acesso em: 10 dez. 2023.

Apesar de ainda não ser uma prática consolidada no Brasil, muitas agências de fomento à pesquisa têm solicitado os PGD como parte do processo de apresentação de propostas de financiamento¹⁰. A construção de um plano de gestão de dados orienta todas as etapas do ciclo de vida dos dados, desde a coleta até o armazenamento e publicação dos dados.

O plano de gestão de dados desta pesquisa busca não somente organizar, desde o princípio, os dados coletados para posterior compartilhamento, mas também registrar de forma transparente o comprometimento de preservá-los em sua integridade para potencial reúso. Como elemento vivo da pesquisa, o PGD criado foi atualizado de acordo com o desenvolvimento da mesma, e uma vez realizado seu preenchimento prévio, deu-se início à coleta de dados.

3.2 EXPLORAÇÃO DO MATERIAL: Coleta e organização dos dados

Considerando a segunda etapa da análise de conteúdo, a exploração do material utilizado nesta pesquisa teve início com a coleta e organização dos dados utilizados, realizada em duas fases. Na primeira, foi realizado o levantamento das revistas indexadas nas bases de dados *Scopus* e *Web of Science* que publicam artigos de dados, enquanto a segunda fase compreendeu a leitura das instruções aos autores ou sessões equivalentes das revistas coletadas na fase anterior, a fim de se obter os elementos que constituem os artigos de dados nelas publicados. Abaixo se encontra descrito o passo a passo dos processos realizados em ambas as fases.

3.2.1 Fase 1 - Levantamento das revistas que publicam artigos de dados

A busca ocorreu inicialmente em janeiro de 2022, sendo atualizada em maio de 2023, nas bases *Scopus* e *Web of Science*. A escolha destas bases se deu devido à ampla abrangência de cobertura em diferentes áreas do conhecimento e a possibilidade de recuperação direta de artigos de dados pelo filtro *data papers* existente em ambas as bases.

A recuperação dos dados na *Scopus* se deu através do uso da query DOCTYPE(dp) AND PUBYEAR <2023 no campo de pesquisa avançada da base, restringindo a busca a artigos de dados publicados antes de 2023. Os registros dos artigos recuperados na *Scopus* foram exportados em formato CSV para organização posterior.

Na *Web of Science* a busca foi realizada com o uso da query DT=(Data Paper) AND PY=(1900-2022), restringindo-se à recuperação de artigos de dados publicados desde o

¹⁰ Como a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por exemplo.

primeiro ano de indexação das fontes da base até o fim do ano de 2022. Os resultados obtidos nesta busca foram exportados em formato Excel e organizados em uma planilha única no *Google Sheets* junto aos resultados recuperados na *Scopus* para deduplicação, excluindo artigos com títulos e DOIs iguais.

Após a deduplicação dos registros recuperados, foi criada uma tabela dinâmica¹¹ para extração da lista das revistas nas quais foram publicados os artigos de dados recuperados.

Das 137 revistas excluídas por estarem fora do escopo, 128 não apresentavam nenhuma orientação relacionada à submissão de artigos de dados ou similar, enquanto 9 possuem algum tipo de similaridade, mas que não se relacionam com a definição de *data paper* utilizada como referencial nesta pesquisa. As observações específicas destes 9 títulos se encontram descritas no Apêndice C localizado no fim da dissertação.

3.2.2 Fase 2 - Características das revistas e elementos dos artigos dados

A partir da fase anterior foram exploradas as instruções aos autores de cada revista recuperada para a verificação da ocorrência dos elementos tratados pela literatura como essenciais para a construção de artigos de dados. Para facilitar a síntese comparativa das informações extraídas dessas orientações, foram criadas tabelas auxiliares na planilha *Excel* desenvolvida, na qual foi organizada em colunas a disposição dos dados a serem extraídos e analisados (Quadro 3).

Quadro 3 - Características das revistas e elementos essenciais dos artigos de dados

Categorias	Dados	Definição
Características da revista	Editora	Nome da editora que publica a revista
	País sede	País onde se localiza a sede da editora
	Tipo de acesso	Tipo de acesso sob o qual os artigos podem ser publicados (aberto x fechado x híbrido)
	Licenças de publicação	Licença sob as quais os artigos são publicados
	Idiomas de publicação	Idioma(s) aceito(s) para publicação
	Áreas do conhecimento ¹²	Principal área do conhecimento coberta pela revista
Elementos essenciais dos artigos de dados	Data de criação dos dados	Data de criação dos conjuntos de dados
	Formato dos dados	Formato de exposição dos dados (CSV, XLS, PDF, RTF, etc)
	Natureza dos dados	Tipologia dos dados (textuais, numéricos, multimídia, etc)
	Versão dos dados	Status de atualização dos dados (bruto, processados, limpos, etc)
	Nome/título do arquivo	Título do conjunto de dados descrito do <i>data paper</i>

¹¹ Tabela que pode ser criada nos aplicativos Excel ou Google Sheets a partir da seleção de duas ou mais colunas de uma planilha, intercalando dados específicos para gerar gráficos mais pontuais ou tabelas menores que facilitem a visualização dos resultados.

¹² De acordo com a tabela de áreas do conhecimento do CNPq.

	Tamanho do arquivo	Tamanho do arquivo
	Linguagem	Linguagem de exposição dos dados
	Método da coleta	Metodologia adotada na coleta/produção dos dados
	Espécimes e Artefatos	Cobertura espacial, temporal e/ou taxonômica dos dados
	Produtor dos dados	Responsável pela coleta/produção dos dados
	Análise dos dados	Metodologia adotada no processamento e tratamento dos dados para obtenção dos resultados
	Dados faltantes	Limitações dos dados (dados perdidos, excluídos ou não processáveis)
	Objetivos dos dados	Objetivo da coleta/produção dos dados
	Breve descrição do projeto	Descrição do projeto de pesquisa/contexto que motivou a coleta/produção dos dados
	Proveniência dos dados	Estudos/publicações relacionados
	Reputação e histórico do repositório	Repositórios confiáveis sugeridos pelas revistas
	Preservação dos dados	Política de preservação dos dados descritos no <i>data paper</i>
	Reúso anterior	Descrição do uso de dados de outra fonte na construção do conjunto de dados descrito
	Reúso potencial	Sugestão e orientações para o reúso dos dados descritos e compartilhados
	Termos e condições de uso	Condições de uso dos dados

Fonte: Adaptado de Kim (2020).

3.3 TRATAMENTO E INTERPRETAÇÃO DOS RESULTADOS: análise dos dados

A etapa final da análise de conteúdo desenvolvida contou com a identificação e sumarização da ocorrência dos elementos essenciais (Quadro 3) nas orientações para a submissão de artigos de dados nas revistas inseridas nesta pesquisa.

Os resultados foram organizados em três categorias adaptadas a partir da estrutura de análise proposta por Kim (2020): 1) Propriedades Gerais dos *Datasets*; 2) Contextualização da produção/coleta dos dados e; 3) Informações sobre o compartilhamento e reúso dos dados. A representação tabular e detalhada dos resultados observados, agrupados por área de conhecimento do CNPq, foi organizada em apêndices.

Para orientar a frequência dos elementos no conteúdo das revistas investigadas, nas tabelas dos apêndices, foram adotados os signos “S” para quando os elementos apareciam nas orientações ou templates para a elaboração de artigos de dados fornecidos pelas revistas e “N” para quando estes não foram encontrados nas seções de orientações aos autores, ou equivalente.

A análise de conteúdo transcorrida considerou o percentual da frequência dos elementos tidos como essenciais para a composição de artigos de dados em comparação com os resultados de estudos anteriores acerca do tema, com a intenção de confirmar seu estado de relevância e propor um modelo adaptável de artigo de dados que comporte adequadamente os elementos

verificados e esclareça o impacto do seu uso efetivo na promoção do compartilhamento e reuso de dados de pesquisa.

Cabe destacar que esta pesquisa se limitou às revistas indexadas nas bases de dados Scopus e Web of Science até o fim do ano de 2022 que possuísem *data papers* publicados e indexados nestas bases de dados.

A análise realizada levou em consideração somente as informações constantes nas sessões de orientações aos autores e equivalentes das revistas incluídas e *templates* disponibilizados nas mesmas. A síntese e representação gráfica dos dados extraídos foi inicialmente realizada em junho de 2022 e atualizada entre junho e novembro de 2023.

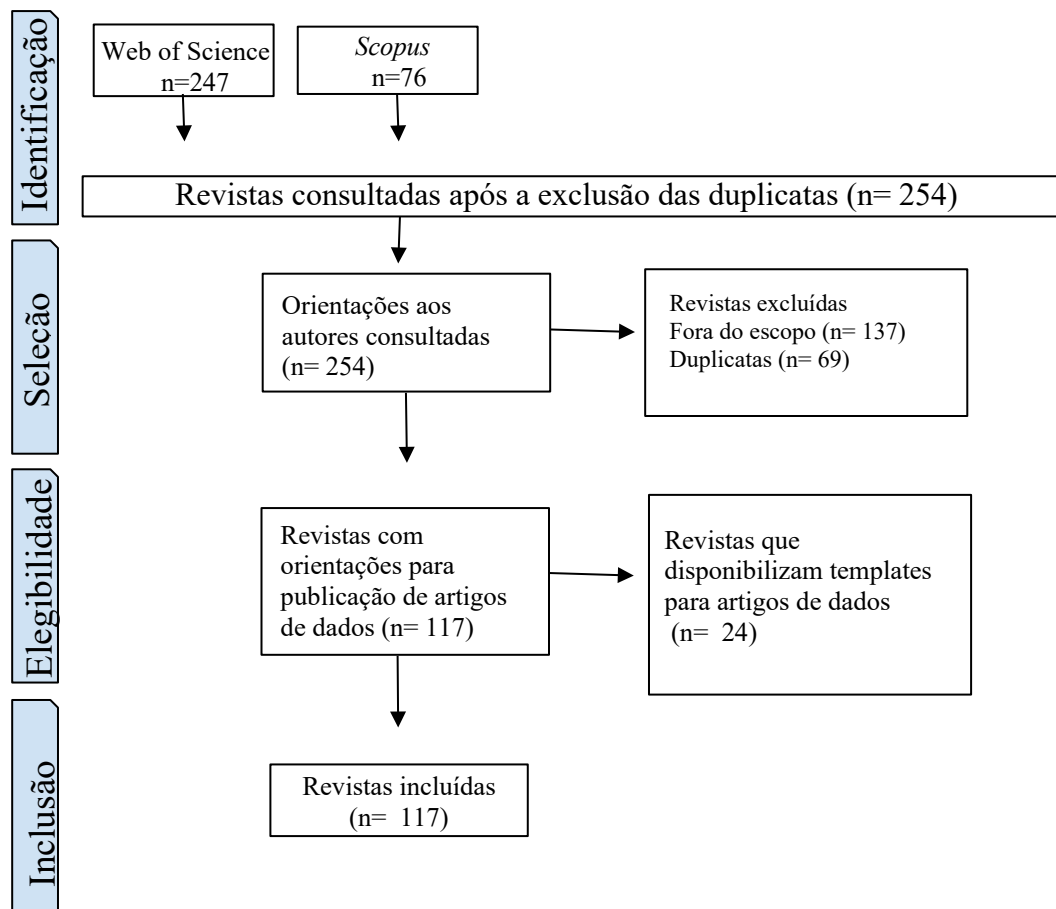
4 RESULTADOS

Foi recuperado um total de 9.698 documentos indexados como *data paper* na base de dados *Scopus* e 11.241 documentos indexados como *data paper* na *Web of Science*. A partir dos metadados destes documentos, foram obtidos 323 títulos de revistas científicas registradas como suas fontes de publicação. Destas, 69 eram duplicatas, estando indexadas em ambas as bases de dados.

Dos 254 títulos não duplicados restantes, em 137 não foram encontradas orientações específicas para a submissão de artigos de dados, sendo consideradas fora do escopo.

Assim, foram incluídas nesta pesquisa um total de 117 revistas (com orientações para publicação de artigos de dados), das quais 24 além de possuírem orientações direcionadas à submissão de artigos de dados, disponibilizavam *template* para sua elaboração. A Figura 5 representa a distribuição da quantidade de revistas por base de dados e ilustra o processo de filtragem e seleção dos títulos explorados nesta pesquisa.

Figura 5 - Fluxograma do processo de seleção das revistas incluídas na pesquisa



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Os resultados abaixo descritos se dividem em duas partes: a primeira orientada à caracterização das revistas utilizadas como fonte principal das informações investigadas nesta pesquisa e a segunda relacionada às suas orientações para elaboração e submissão de *data papers*.

4.1 CARACTERÍSTICAS DAS REVISTAS QUE PUBLICAM ARTIGOS DE DADOS

Tendo em vista a ambientação das revistas cujos dados são discutidos nesta pesquisa, são descritas algumas características relevantes para sua identificação.

Os títulos considerados se encontram dispersos por 26 editoras acadêmico-científicas, das quais se destacam com maior número de publicações: a *Frontiers Media S.A.*, com 42 revistas em diversas áreas do conhecimento; e a *Wiley*, com 17 revistas também em áreas distintas (Tabela 1).

Tabela 1 - Dispersão do número de revistas por editora científica

Editora	Revistas
<i>Frontiers Media S.A.</i>	42
<i>Wiley</i>	17
<i>Pensoft</i>	9
<i>Springer Nature</i>	7
<i>Elsevier</i>	6
<i>BMC - Springer Nature</i>	5
<i>Oxford University Press</i>	4
<i>Taylor & Francis</i>	4
<i>Ubiquity Press</i>	3
<i>EDP Sciences</i>	2
<i>MDPI</i>	2
<i>Sage Journals</i>	2
<i>Asociación Española de Ecología Terrestre</i>	1
<i>Cambridge University Press</i>	1
<i>Canadian Science Publishing</i>	1
<i>Copernicus Publications</i>	1
<i>F1000Research</i>	1
<i>Geochemical Society of Japan</i>	1
<i>Geological Survey of Denmark and Greenland (GEUS)</i>	1
<i>Gigvvy Science Open Access Publishing / Taiwan Association for Aerosol Research</i>	1
<i>GreyNet</i>	1
<i>International Union of Crystallography (IUCr)</i>	1
<i>Korean Society of Epidemiology</i>	1
<i>MIT Press</i>	1
<i>Regional Euro-Asian Biological Invasions Centre (REABIC)</i>	1
<i>Seismological Society of America</i>	1
Total	117

Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Concomitantemente, considerando a localização geográfica das editoras acima listadas, destacam-se com maior número de revistas que publicam artigos de dados: os Estados Unidos e a Suíça.

No que diz respeito ao idioma de publicação suportado por estas revistas, identificou-se que 114 das 117 revistas incluídas têm o inglês como idioma principal de publicação e ou outros três títulos, além de utilizarem o inglês como idioma preterível e publicação, aceitam também publicações em coreano (*Epidemiology and Health*, publicada pela *Korean Society of Epidemiology*), francês (*Artic Science*, publicada pela *Canadian Science Publishing*) e castelhano (*Ecosistemas*, publicada pela *Asociación Española de Ecología Terrestre*).

Outro ponto importante a ser descrito a respeito das revistas exploradas na presente pesquisa são as modalidades de acesso que suportam e as licenças aplicáveis a cada uma delas. A Tabela 2 destaca a dispersão das revistas em relação à modalidade de acesso que adota (acesso aberto, fechado ou híbrido) e as licenças sob as quais os materiais acessados podem ser utilizados.

Tabela 2 - Dispersão do número de revisas em relação ao acesso

Licença	Acesso			Total
	Aberto	Fechado	Híbrido	
Não encontrada	2	1	3	6
CC BY	74		14	88
CC BY / CC BY-NC / CC BY-NC-ND	4		7	11
CC BY / CC BY-NC-ND	2		1	3
CC BY / CC BY-NC	2			2
CC BY / CC BY-NC-ND	1		1	2
CC BY-NC	1			1
CC BY-NC-ND	1			1
CC BY-ND	1			1
CC BY / CC BY-NC / CC BY-ND / CC BY-NC-ND	1			1
CC BY / CC BY-ND	1			1
Total geral	91	1	25	117

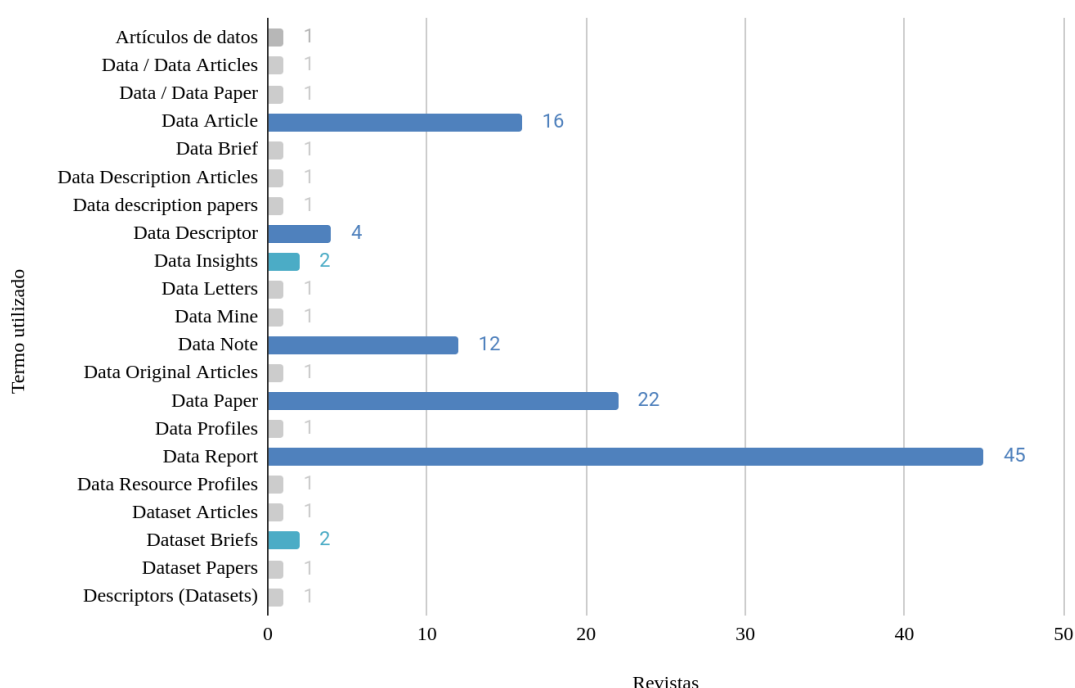
Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Verifica-se que a maior parte das revistas veicula seus artigos em acesso aberto, destacando-se o número elevado de publicações (74) que utiliza a segunda modalidade de licença *creative commons* mais permissiva (CC BY), que só não ultrapassa a modalidade CC0

(domínio público) em termos e condições de uso. Cabe destacar que nenhuma das revistas incluídas nesta pesquisa faz uso desta última.

Quanto à maneira como cada revista se refere aos artigos de dados em suas orientações, nota-se uma variedade de termos adotados, dentre os quais se destacam *Data Report*, *Data Paper*, *Data Article*, *Data Note* e *Data descriptor*. Todos os demais termos encontrados e a quantidade de revistas que apropria de cada um deles estão representados no gráfico abaixo (Gráfico 1).

Gráfico 1 - Termo adotado pelas revistas para se referir aos artigos de dados



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2 DAS ORIENTAÇÕES PARA A SUBMISSÃO DE ARTIGOS DE DADOS

Considerando a estrutura de análise proposta por Kim (2020), a síntese comparativa adotada nesta pesquisa se subdivide em três categorias: 1) Propriedades Gerais dos *Datasets*; 2) Contextualização da produção/coleta dos dados e; 3) Informações sobre o compartilhamento e reúso dos dados.

Para melhor entendimento, optou-se por apresentar, de forma tabular e detalhada, os resultados encontrados para cada revista, agrupados por área de conhecimento, nos Apêndices D, E e F, respectivamente para as categorias 1, 2 e 3 de Kim.

Abaixo são representados gráfica e individualmente os resultados obtidos para cada elemento presente nestas categorias.

4.2.1 Propriedades gerais dos *datasets*

A respeito das propriedades gerais dos *datasets*, Kim (2020) sugere sete (7) elementos a serem solicitados na composição de um artigo de dados: 1) data de criação dos *datasets*; 2) formato dos dados; 3) tipo dos dados; 4) versão dos dados; 5) título/nome dos conjuntos de dados; 6) tamanho dos arquivos de dados; e 7) linguagem dos dados.

4.2.1.1 Data de criação dos *datasets*

Está relacionada com a possibilidade de ajudar a quem possa reutilizar os conjuntos de dados descritos a identificar a base de amostragem dos dados e as alterações que podem ocorrer para além do contexto em que estes foram coletados/criados.

O Gráfico 2 representa a porcentagem de revistas que solicitam a descrição da data de criação dos *datasets* em suas orientações.

Gráfico 2 – Revista solicita descrição da data de criação dos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Das 117 revistas incluídas na pesquisa, pouco mais de 22% das revistas solicitaram a data de criação dos *datasets* em suas orientações ou *templates*. Em seu estudo, Kim (2020) havia identificado que esta era uma ocorrência rara na amostra utilizada por ele. Apesar da notória diferença entre o escopo da pesquisa em tela e a publicada pelo autor em 2020, Araujo e Alencar (2023, no prelo) perceberam um crescimento gradativo da inserção desta informação nos templates de revistas de dados.

Em Kim (2020) a diferença observada era de 14 *templates* sem a solicitação da data de criação dos *datasets* para um, enquanto em Araujo e Alencar (2023) a diferença observada em escopo similar foi de 13 *templates* sem a solicitação da data de criação dos *datasets* para cinco.

Contudo, em comparação com os demais elementos desta categoria, na presente pesquisa, esta informação não se destacou dos demais elementos devido a sua baixa porcentagem de adesão nas orientações selecionadas, pouco maior que 22% (Gráfico 2).

4.2.1.2 Formato dos dados

Conforme citado anteriormente, a pesquisa atual corrobora com o entendimento dos autores quanto ao que se concebe ao falar-se de formato dos dados. Nesse enquadramento foi observado menos adesão do elemento por parte das revistas em seus *templates* e/ou orientações, dos quais apenas cerca de 20% solicitam esta informação dos *conjuntos de dados* descritos (Gráfico 3).

O Gráfico 3 representa a porcentagem de revistas que solicitam a descrição do formato dos dados de pesquisa. Ao pensar este elemento como um dos elementos essenciais para a descrição efetiva de conjuntos de dados em artigos de dados, Kim (2020) considerou a visão de Chin e Lansing (2004) a seu respeito, que por sua vez compreendiam como formatos dados instrumentais, imagens, texto, tabelas, planilhas e publicações, bem como o que segundo os autores seriam suportes padrão do formato de dados, a saber: JPEG (*Joint Photographic Experts Group*), PDF (*Portable Document Format*), and CSV (*Comma-Separated Values*), etc. (Chin; Lansing, 2004).

Gráfico 3 - Revista solicita descrição do formato dos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Tendo em vista o exposto por Kim (2020), Candela *et al.* (2015) em seus estudos mais recentes acerca da composição dos *data papers*, acredita-se que a descrição do formato de apresentação dos conjuntos de dados, bem como do seu tipo ou natureza e de sua versão ou status de processamento possa contribuir com a melhor recuperação e leitura dos dados. Isso porque, uma vez que o(a) leitor(a) se depare com estas informações, ele/ela já pode se preparar

para calcular as possibilidades de recuperação, leitura e aplicabilidade destes dados diante de suas necessidades e aparatos técnicos que possibilitem seu melhor aproveitamento.

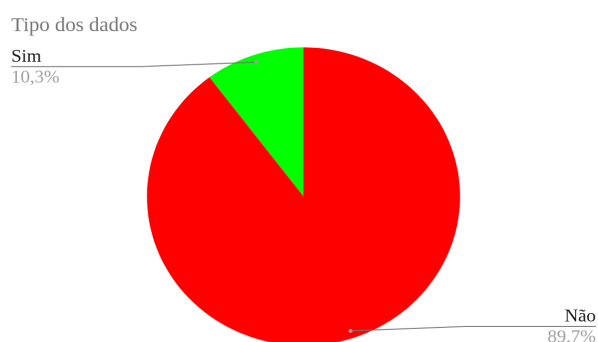
4.2.1.3 Tipo dos dados

A definição dos tipos de dados utilizada nesta pesquisa tem base no conceito de natureza dos dados conforme classificado na taxonomia proposta por Sales e Sayão (2019), na qual os autores explicam que a natureza ou o tipo dos dados que

retrata a grande diversidade e heterogeneidade de tipos de dados que podem ser originados no ambiente de pesquisa em termos de formatos, mídias, suportes, expressões, arcabouço tecnológico etc. (Sales; Sayão, 2019, p. 43).

Neste cenário, o Gráfico 4 representa a porcentagem de revistas que solicitam o esclarecimento da tipologia (ou natureza) dos dados apresentados, a partir das quais se observa que apenas 10% das orientações consultadas solicitam explicitamente o tipo dos dados de pesquisa compartilhados.

Gráfico 4 - Revista solicita descrição do tipo dos conjuntos de dados?



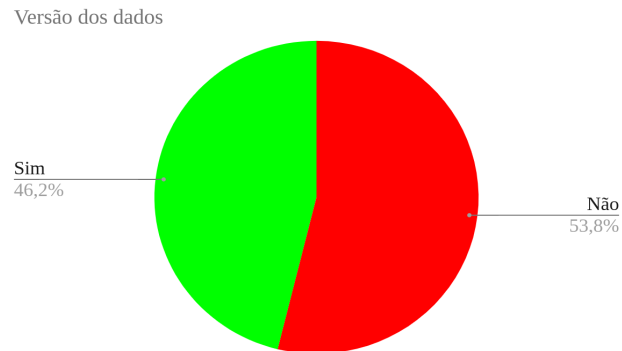
Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2.1.4 Versão dos dados

Se refere às etapas do processamento dos dados, se estes se tratam de: dados brutos, sem que tenham sido processados de nenhuma maneira; dados processados, que tenham passado por tratamento em alguma instância; dados reutilizados, caso provenientes de pesquisas anteriores; dados atualizados, em caso de atualizações de conjuntos de dados anteriormente publicados etc.

O Gráfico 5 representa a porcentagem de revistas que solicitam a versão dos dados no qual pode ser verificada a maior porcentagem de adesão pelas revistas em suas orientações (46%) na tríade de elementos que descrevem a tipologia, formato e status de processamento dos dados descritos.

Gráfico 5 - Revista solicita descrição da versão dos conjuntos de dados?



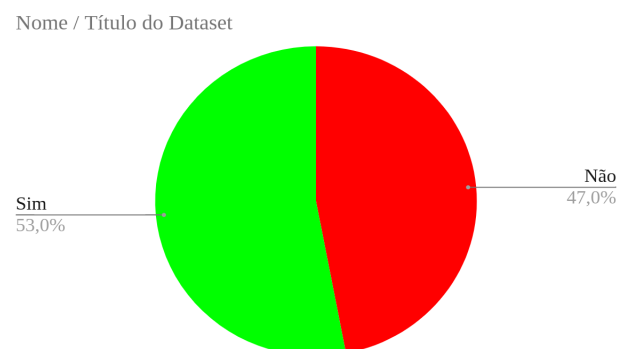
Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2.1.5 Título/nome dos conjuntos de dados

Na maioria das vezes, a nomeação dos conjuntos de dados, auxilia na distinção entre conjuntos de dados compostos e/ou arquivos submetidos em fontes externas que devem ser referenciados no corpo do texto dos artigos de dados. Não foram considerados para este elemento as citações a conjuntos de dados listadas na seção de referências dos artigos de dados.

O Gráfico 6, por sua vez, apresenta a segunda maior porcentagem de adesão por parte das revistas em suas orientações, na qual aproximadamente 53% requerem explicitamente a descrição do nome ou título dos conjuntos de dados representados nos artigos de dados submetidos. Este número representa, na categoria de propriedades gerais dos *datasets* o maior número de solicitações declaradas.

Gráfico 6 - Revista solicita descrição do título/nome dos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Foram considerados no cálculo de ocorrências deste elemento a solicitação dos nomes dos arquivos não apenas nos artigos de dados, mas também nas declarações de disponibilidade dos dados (*Data Availability Statement*), quando as orientações específicas para artigos de

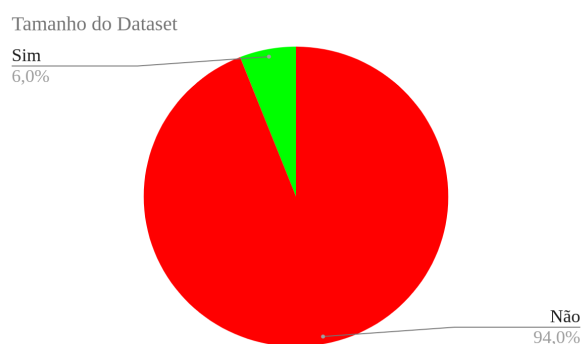
dados direcionavam para a mesma e requeriam sua submissão como parte do texto do *data paper*.

4.2.1.6 Tamanho dos arquivos de dados

A extensão dos dados a serem compartilhados ou submetidos junto aos artigos de dados para armazenamento em repositório ou arquivo específico sugerido pelas revistas.

O Gráfico 7 representa a porcentagem de revistas que solicitam a descrição do tamanho dos *datasets* em suas orientações que, em contraposição com o anterior, posiciona seu elemento como segundo menos requerido nas orientações e *templates* consultados, com apenas 6% de ocorrências (Gráfico 6).

Gráfico 7 - Revista solicita descrição do tamanho dos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2.1.7 Linguagem dos dados

Este elemento descreve a linguagem na qual os dados estão expostos. Ex.: terminologia técnica, alfanumérica, de programação etc.

Encerrando a primeira categoria explorada, o Gráfico 8 evidencia a porcentagem de revistas que solicitam a descrição da linguagem dos *datasets* em suas orientações. Neste se observa o menor número de solicitações, com pouco mais de 4% das revistas aderindo à sua citação direta em suas orientações.

Gráfico 8 - Revista solicita descrição da linguagem dos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Apesar do exposto por Candela *et al.* em 2015 se referir mais à descrição dos formatos, tipos e versões dos dados compartilhados, geralmente mais bem discutidos ao se pensar na aplicação de um ou mais conjuntos de dados compartilhados, a partir da pesquisa em tela, acredita-se que os elementos que ressaltam o tamanho dos conjuntos de dados disponibilizados e sua linguagem, natural ou programada, podem inferir no tempo despendido na compreensão dos dados de pesquisa que se pretende reutilizar.

Questões como esta são discutidas mais adiante no texto, após dar-se seguimento à exposição dos resultados das demais categorias. Os resultados encontrados nas orientações de cada revista incluída nesta pesquisa para a categoria de propriedade dos *datasets* estão expostos no Apêndice D.

4.2.2 Contextualização da produção dos *datasets*/coleta dos dados

No que se refere à categoria de produção dos conjuntos de dados ou coleta dos dados, Kim (2020) registra em sua pesquisa que os elementos presentes na 2ª categoria explorada tendem a ser solicitados com mais frequência pelas revistas de dados em comparação com os elementos das demais categorias. Esse fenômeno se consolida na observância das porcentagens de ocorrência de cada elemento destacado a seguir.

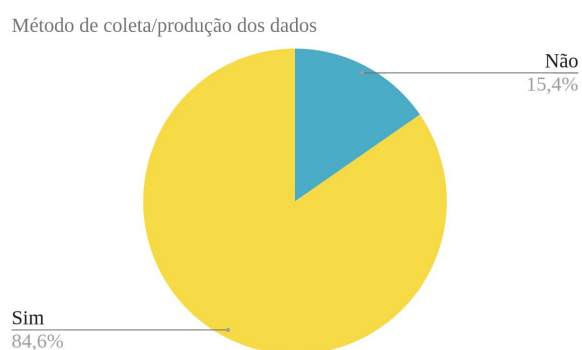
Para maior entendimento, são apresentados o significado de cada elemento desta categoria no contexto da atual pesquisa: 1) método de coleta/produção dos *datasets*; 2) cobertura dos dados; 3) produtor dos dados; 4) análise ou processamento dos dados; 5) limitações; 6) objetivo da coleta/produção dos dados; 7) contexto.

4.2.2.1 Método de coleta/produção dos *datasets*

Retrata a metodologia conduzida na produção ou coleta dos dados que compõem o conjunto de dados descrito no *data paper*. Ao que Kim (2020) relaciona com a explanação dos passos da coleta de dados, estratégias de seleção de amostras e mecanismos de controle de qualidade.

O Gráfico 9 representa a porcentagem de revistas que solicitam a descrição da metodologia utilizada na coleta dos dados/produção dos conjuntos de dados em suas orientações.

Gráfico 9 - Revista solicita descrição do método de produção/coleta dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Diferente dos resultados observados por Kim (2020), no qual ocorreu a totalidade do requerimento deste elemento pelas fontes consultadas, na presente pesquisa identificou-se que cerca de 85% das revistas consultadas solicita este elemento. Um número que ultrapassa o resultado obtido pelo elemento “título/nome do(s) *datasets*”, o mais solicitado pelas revistas na categoria anterior. Isso porque, em amplitude, foram incluídas nesta pesquisa todas as revistas que possuem orientações relacionadas aos *data papers*, independentemente da profundidade com que descrevem seu formato.

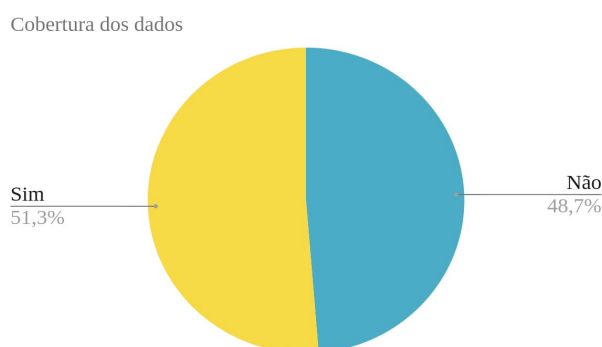
Em um panorama geral, verificou-se que existe uma variedade de formas com as quais as revistas podem descrever suas instruções para publicação de artigos de dados, fazendo com que alguns títulos especifiquem os detalhes deste modelo de publicação em *templates*, de modo mais direto e objetivo, enquanto outros títulos podem esclarecer em pouco mais de um parágrafo o que vem a ser um artigo de dados e de que informações ele é composto.

4.2.2.2 Cobertura dos dados

A cobertura dos dados, ou espécimes e artefatos (Kim, 2020), diz respeito às delimitações espaciais, temporais e taxonômicas dos conjuntos de dados.

O Gráfico 10 representa a porcentagem de orientações que solicita a descrição da cobertura dos dados representados nos *data papers*. Nesse cenário são compreendidos como delimitadores de cobertura todo e qualquer limite estabelecido na seleção da amostra dos dados.

Gráfico 10 - Revista solicita descrição da cobertura dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

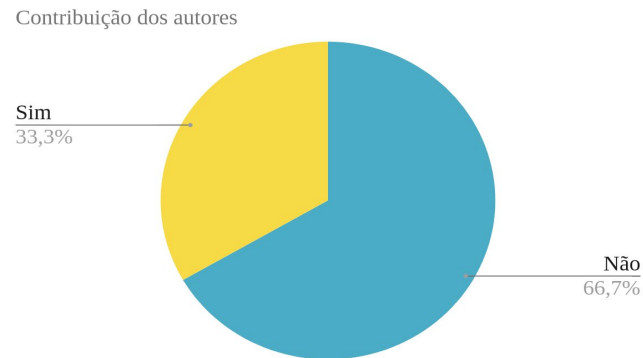
Neste quadro nota-se que pouco mais da metade das revistas consultadas (51%) declaradamente solicita o preenchimento da cobertura dos dados nos artigos de dados que publica.

4.2.2.3 Produtor dos dados

Indivíduo(s) responsável(eis) pela coleta ou produção dos conjuntos de dados. Para fins desta pesquisa, além da presença explícita desse elemento nos *templates* e orientações analisados, foram consideradas ocorrências deste elemento as solicitações da descrição da contribuição dos autores, co-autores e pessoas reconhecidas (*Acknowledgement*).

Pouco menos encontrado de forma evidente nas orientações e modelos examinados foi a apresentação do produtor dos dados, isto é, o registro da(s) pessoa(s) responsável(eis) pela coleta ou produção dos dados. Para fins de reconhecimento, na presente pesquisa foram consideradas ocorrências deste elemento a solicitação da descrição da contribuição dos autores nos artigos de dados, para o qual se observa que 33% (1/3) das revistas aqui exploradas demanda tal informação (Gráfico 11).

Gráfico 11 - Revista solicita descrição da contribuição dos autores?



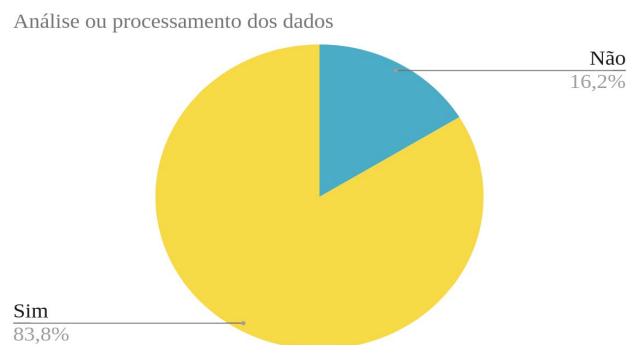
Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2.2.4 Método do processamento dos dados

Refere-se à forma com que os dados foram processados, caso tenham sido. Levou-se em consideração nesta pesquisa não somente as ocorrências explícitas deste elemento, em que as orientações ou *templates* requerem esta exata informação, como também aquelas em que se demanda a especificação de softwares utilizados no tratamento dos dados e outras indicações de descrição relacionadas ao processamento e tratamento dos dados.

O Gráfico 12 representa a porcentagem de revistas que solicitam a descrição da análise, processamento ou tratamento aplicados aos dados, quando estes ocorrem. Ao comparar a ocorrência deste elemento com o valor obtido nas solicitações da descrição da metodologia de coleta dos dados (Gráfico 9), é percebida uma singela diferença entre os resultados de ambas.

Gráfico 12 - Revista solicita descrição do método do processamento dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Mais uma vez destoando-se dos resultados observados por Kim (2020), cujo objeto *data analysis* surge em pouco menos de $\frac{1}{4}$ dos *templates* consultados pelo autor. Na presente

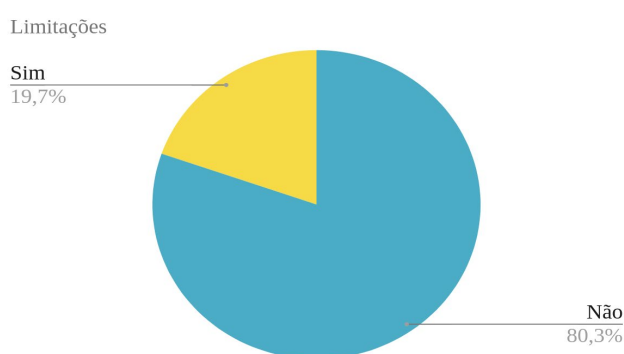
pesquisa identificou-se que cerca de 84% das revistas consultadas solicita a descrição desses processos e das ferramentas nele envolvidas (Gráfico 12).

4.2.2.5 Dados faltantes

Limitações ou *missing data* (Kim, 2020) são entendidas como eventuais anomalias, ruídos ou vieses dentro dos conjuntos de dados que tenham levado à sua exclusão ou redução na disponibilização dos dados ou no produto compartilhado.

Com o menor resultado obtido na categoria de contextualização da coleta/produção dos dados, o Gráfico 13 apresenta que apenas cerca de 20% das revistas solicitam a descrição dos dados faltantes ou quaisquer limitações que tenham sido encontrados durante a coleta/produção dos dados.

Gráfico 13 - Revista solicita descrição dos dados faltantes nos conjuntos de dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Não fosse por sua relevância e implicações no reúso dos dados compartilhados, a baixa adesão das revistas à solicitação de informações sobre as limitações dos conjuntos de dados compartilhados poderia não ser observada como um problema, mas ao pensarmos na problemática abordada por Sayão e Sales (2020) a respeito da não publicação de resultados negativos, sobre o qual os autores esclarecem que

a invisibilidade sobre os resultados negativos tem efeito crucial sobre os princípios de autocorreção da ciência, da reprodutibilidade dos experimentos científicos e da transparência de seus processos. Isso sem falar na duplicação de esforços financeiros e intelectuais desperdiçados na repetição de experimentos que são becos sem saída; nos desvios éticos causados pelo chamado “viés de publicação”, que cria uma inclinação artificial para resultados positivos que são mais facilmente publicáveis [...]. (Sayão; Sales, 2020, p. 99).

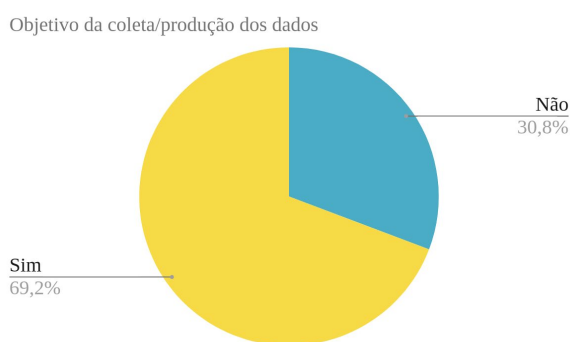
É de se esperar e entender que ele possa suceder, em nível diferente, com a anulação da descrição dos dados que são subnotificados, excluídos ou ilegíveis nos conjuntos de dados que são compartilhados e publicados, inferindo em viés de pesquisas futuras que possam se beneficiar de seu reuso e aplicação.

4.2.2.6 Objetivo da coleta/produção dos dados

Trata sobre as motivações que levaram à coleta ou produção do conjunto de dados.

O penúltimo gráfico desta categoria representa a porcentagem de revistas que solicitam a descrição do objetivo da produção ou coleta dos dados em suas orientações que detém cerca de 70% de ocorrências (Gráfico 14).

Gráfico 14 - Revista solicita descrição do objetivo da produção/coleta dos dados?



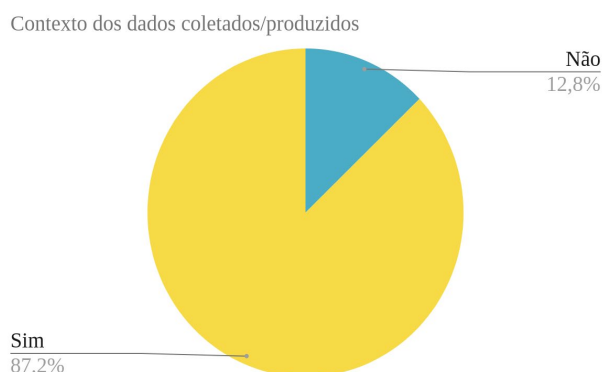
Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

4.2.2.7 Contexto

O contexto ou *project* (Kim, 2020) trata justamente do contexto a partir do qual e para o qual os dados foram coletados ou produzidos originalmente. Ex.: um projeto de pesquisa, financiamento etc.

Encerrando a segunda categoria, o Gráfico 15 reflete o maior número valor obtido dentre os resultados desta pesquisa, no qual se observa que aproximadamente 88% das revistas investigadas solicitam a descrição do contexto ou projeto que deu origem à produção dos dados em suas orientações.

Gráfico 15 - Revista solicita descrição do contexto dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Nota-se uma diferença um tanto quanto significativa entre este e o Gráfico 14, isso porque apesar de a maioria das revistas indicarem a necessidade de uma breve descrição do contexto em que os dados foram obtidos e para o qual os dados serão empregados, uma parcela menor delas esclarece diretamente a necessidade de descrição do objetivo dos dados.

É válido ressaltar que uma informação não necessariamente anula ou garante a existência da outra, por isso estes elementos são tratados por Kim (2020) separadamente, porém com mesma valia em termos de relevância.

Os resultados encontrados nas orientações de cada revista incluída nesta pesquisa para a categoria de propriedade dos *datasets* estão expostos no Apêndice E.

4.2.3 Informações sobre o compartilhamento e reuso dos dados

A categoria de informações sobre o compartilhamento e reuso dos dados, carrega em si os últimos 6 elementos apontados por Kim (2020) em seu estudo, três dos menos citados nos modelos de artigos de dados analisados pelo autor, a saber: 1) reputação e histórico dos repositórios; 2) curadoria e digitalização, “preservação” para fins da pesquisa em curso e 3) Reuso prévio, ou reuso de dados anteriores.

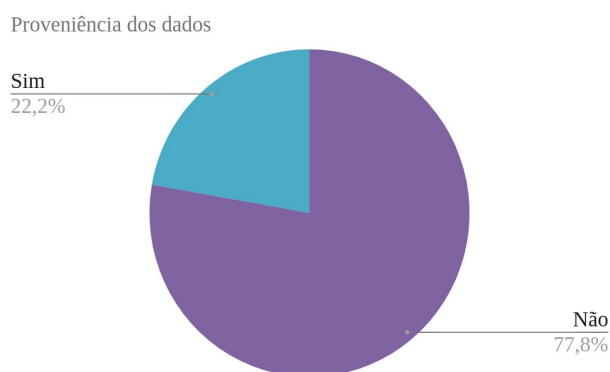
Para fins de adaptação e retratação mais fiel da realidade das orientações para publicação de artigos de dados consultadas, algumas das definições dos elementos destas categorias foram revistos e ressignificados conforme abaixo.

4.2.3.1 Proveniência dos dados

Em seu estudo, Kim (2020) aborda a proveniência dos dados sob óticas distintas, na qual uma delas retrata relação dos dados com outros materiais e a outra, mais comumente abordada, retrata a relação dos conjuntos de dados com seu local de depósito. A partir da pré-chegagem dos dados coletados na pesquisa em tela, foi possível verificar que 100% das revistas incluídas solicitam a descrição do identificador persistente dos conjuntos de dados relacionados aos artigos de dados nelas publicados e/ou solicitados que estes estejam publicados em repositórios de dados. Assim, decidiu-se por adotar para fins desta pesquisa a primeira perspectiva de proveniência apresentada por Kim (2020), sendo consideradas ocorrências deste elemento as orientações que solicitasse a listagem ou citação de artigos, projetos, protocolos e demais materiais que tivessem como objeto de estudo os mesmos dados de pesquisa.

O Gráfico 16 representa a porcentagem de revistas que solicitam a descrição da proveniência dos dados, a partir da perspectiva que demonstra a relação dos conjuntos de dados com outros artigos, projetos, protocolos e demais materiais que os utilize como objeto de estudo e investigação.

Gráfico 16 - Revista solicita descrição da proveniência dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

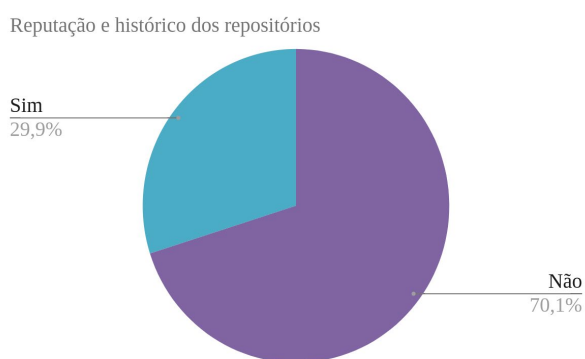
Nesse sentido, observa-se que apenas pouco mais de 22% das revistas exploradas requer esta informação em suas orientações para publicação de artigos de dados. Um número relativamente baixo de ocorrência, mas que em comparação com os resultados obtidos por Kim (2020), em que apenas 1(um) dos 15 templates consultados pelo autor requeriam esta informação, pode-se prever um crescimento de pelo menos 2(duas) ocorrências deste elemento a cada 10 orientações analisadas.

4.2.3.2 Reputação e histórico dos repositórios

Assim como observado por Kim (2020), a primeira averiguação dos dados obtidos na presente pesquisa constatou que nenhuma das revistas incluídas solicita diretamente qualquer informação a respeito da reputação ou histórico dos repositórios em que os dados estão compartilhados. Contudo, uma parte significativa destas revistas disponibiliza em suas orientações para publicação de *data papers* uma lista de repositórios confiáveis recomendados para fins de compartilhamento dos dados. Desse modo, foram consideradas ocorrências deste elemento as revistas que em suas orientações sugerissem repositórios confiáveis aos autores.

O Gráfico 17 representa a porcentagem de revistas que solicitam a descrição da reputação e histórico dos repositórios em que os conjuntos de dados relacionados aos *data papers* publicados estão disponibilizados, considerando ocorrências deste elemento a sugestão de repositórios confiáveis por parte das revistas. Isto porque, conforme supracitado, identificou-se na presente pesquisa que nenhuma das orientações consultadas trazia em si nenhuma solicitação do esclarecimento da reputação ou histórico dos repositórios.

Gráfico 17 - Revista solicita descrição da reputação e histórico dos repositórios?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Ao que Kim (2020), citando Faniel, Frank e Yakel (2019), declara que a subtração desta informação ocorre devido a subjetividade da mesma, que evoca a visão pessoal do autor a respeito dos repositórios que utiliza, ao invés de considerar fatores mais objetivos de confiabilidade para os mesmos. O autor segue, contudo, afirmando que a descrição de informações sobre as condições de acesso e disponibilidade a longo-prazo dos dados depositados poderia contribuir para a validação da confiabilidade dos repositórios.

Apesar disso, os autores reconhecem e destacam positivamente as revistas que ao menos direcionam os autores a repositórios confiáveis em suas orientações. Desse modo, verifica-se que aproximadamente 30% das revistas incluídas nesta pesquisa recomendam um ou mais

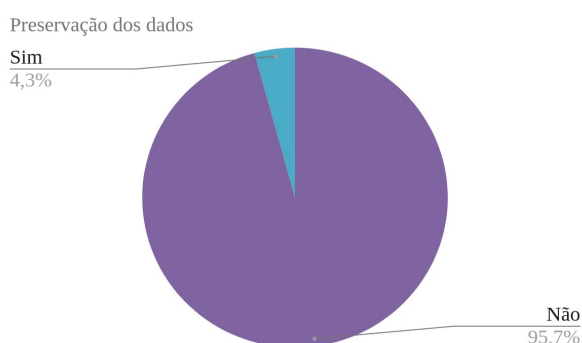
repositórios, multidisciplinares ou temáticos para depósito dos dados descritos em seus artigos de dados (Gráfico 17).

4.2.3.3 Preservação dos dados

Refere-se à forma com que se pretende preservar os dados a curto, médio ou longo prazo, a depender das revistas e das necessidades específicas de cada área do conhecimento.

O Gráfico 18 representa a porcentagem de revistas que solicitam a descrição da preservação dos conjuntos de dados a curto, médio ou longo prazo nos repositórios em que foram armazenados de acordo com as políticas pré-estabelecidas pelas revistas ou por seus repositórios sugeridos. Kim (2020) registra este elemento em seu estudo combinando os termos “curadoria e digitalização”, alterados na presente pesquisa por sugestão de pares após apresentação dos resultados parciais da mesma no 23º Encontro Nacional de Pós-Graduação em Ciência da Informação, ocorrido em novembro de 2023, em Aracaju - Sergipe, Brasil.

Gráfico 18 - Revista solicita descrição da preservação dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

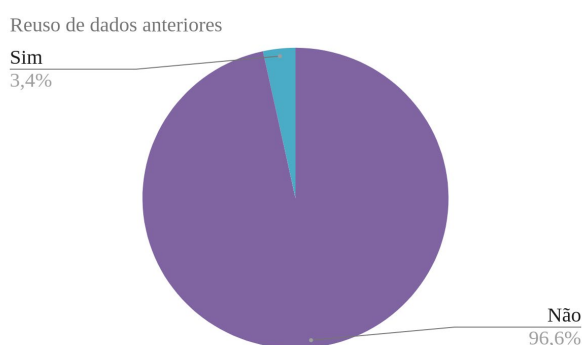
Em seu trabalho, o autor relacionou este elemento com a exigência de informações sobre os procedimentos de arquivamento dos dados, incluindo uma descrição de como os dados são arquivados para armazenamento de longo prazo e como lhes é garantido o acesso durante o período de sua disponibilização. Na ocasião, Kim (2020) identificou que apenas uma das fontes consultadas em sua pesquisa requeria estas informações em sua composição. Na presente pesquisa é percebido um singelo aumento nos casos, em que é possível observar que pouco mais de 4% (5) das revistas consultadas solicitam detalhes acerca do período de preservação dos dados ou requerem que estes sejam preservados durante um período específico (Gráfico 18).

4.2.3.4 Reúso de dados primários ou anteriores

É entendido como um registro do reúso de dados coletado/produzidos anteriormente com finalidade similar ou não.

O Gráfico 19 representa as revistas que solicitam a descrição da do reúso de dados anteriores, ou seja a utilização de dados ou conjuntos de dados coletados ou produzidos anteriormente por terceiros ou em distinta ocasião com finalidade similar ou não a do *dataset* relacionado ao artigo de dados em construção.

Gráfico 19 - Revista solicita descrição do reúso de dados anteriores?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

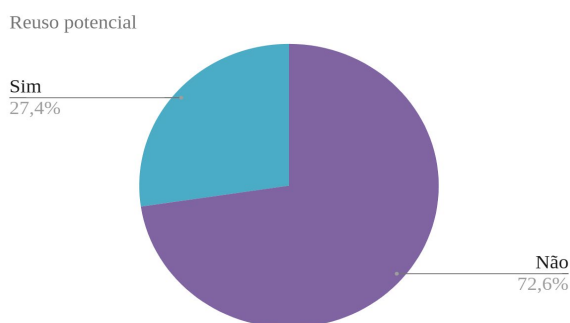
Os resultados observados por Kim (2020) apontavam que apenas três dos templates investigados pelo autor requeriam esta informação, ao que a presente pesquisa corrobora ao demonstrar que apenas pouco mais de 3% (4) das revistas consultadas solicitam explicitamente esta informação em suas orientações (Gráfico 19).

4.2.3.5 Reúso potencial dos dados

Versa sobre as diferentes possibilidades de reúso e aplicação dos dados abertos compartilhados.

O gráfico Gráfico 20 da terceira categoria explorada representa a porcentagem de revistas que solicitam a descrição do potencial de reúso dos conjuntos de dados descritos nos artigos de dados em construção. Informação de extrema relevância para o cumprimento das funções idealizadas para o modelo de publicação que intende fomentar o reúso e aplicação de dados abertos de pesquisa em estudos diversos através da melhor compreensão de suas propriedades e conteúdo.

Gráfico 20 - Revista solicita descrição do reúso potencial dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

No entanto, os resultados demonstram que a exigência ou evidência deste elemento como parte essencial da estrutura do artigo de dados ainda não é uma realidade aplicada de modo geral pelas revistas que recebem submissões deste tipo de publicação. Em seu estudo, Kim (2020) aponta que aproximadamente metade dos templates incluídos em sua pesquisa não possuíam a descrição do reúso potencial dos dados como parte da estrutura de seus modelos.

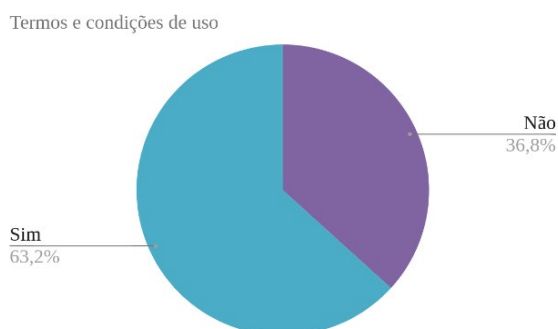
Os dados obtidos na pesquisa em tela retratam que apenas cerca de 28% das revistas incluídas solicitam de alguma maneira a descrição do potencial reúso dos dados compartilhados, dentro ou fora do seu contexto inicial de criação/coleta (Gráfico 20).

4.2.3.5 Termos e condições de uso dos dados

Trata justamente sobre a delimitação de uso e aplicação dos dados compartilhados. Ex.: a licença sob a qual os dados estão disponibilizados

Por fim, o Gráfico 21 representa a porcentagem de revistas que indicam ou solicitam a indicação dos termos e condições de uso dos dados de pesquisa disponibilizados e representados em seus artigos de dados.

Gráfico 21 - Revista solicita descrição dos termos e condições de uso dos dados?



Fonte: Elaborado pela autora a partir dos dados da pesquisa (2023).

Refletindo em maior proporção o que fora obtido no estudo de Kim (2020), este elemento carrega o maior número de ocorrências de sua categoria, que demonstra que mais de 63% das revistas incluídas nesta pesquisa descreve ou solicita a descrição dos termos condições sob os quais os conjuntos de dados compartilhados dever estar disponibilizados ao público.

Os resultados encontrados nas orientações de cada revista incluída nesta pesquisa para a categoria de informações sobre o compartilhamento e reuso dos dados estão expostos no Apêndice F.

A discussão mais aprofundada dos resultados obtidos segue na seção seguinte.

5 DISCUSSÃO

A investigação de modelos descritivos para representação de dados não é um assunto novo, e com o passar dos anos percebe-se que notórios avanços foram alcançados nesta instância, principalmente no que figura o desenvolvimento de padrões de metadados mais ou menos específicos com a intenção de tornar dados encontráveis e reproduzíveis.

Desde 2009 modelos para a descrição contextual de dados de pesquisa vêm sendo discutidos. Apesar das tentativas promissoras de se estabelecer *frameworks* descritivos efetivos que possibilitasse o melhor aproveitamento dos dados compartilhados e publicados, poucos esforços envolveram a perspectiva de pesquisadores interessados na utilização de dados coletados/produzidos para o desenvolvimento de pesquisas posteriores que servissem de matéria prima para novos estudos (Faniel; Frank; Yakel, 2019).

Faniel, Frank e Yakel (2019) em seu estudo apontam esta realidade de forma crítica ao considerar que o custo de coletar do zero ou reproduzir um método de coleta ou produção de dados é geralmente muito maior do que o despendido para documentar os dados de forma extensiva e clara, que torne a replicabilidade da pesquisa ou reuso eficiente dos dados mais fácil. Embora a documentação de dados consuma tempo e também não seja desprovida de custo.

No entanto, ao tomar-se em conta que alguns dados, principalmente dados observacionais, não podem ser recoletados ou reproduzidos, fica evidente que dentre as opções acima ponderadas a menos onerosa ainda seria a documentação apropriada dos dados de pesquisa passíveis de reaproveitamento.

Em contrapartida, a relevância da documentação dos dados para além do que consta nos metadados, já amplamente divulgados e adotados em bases de dados, arquivos e repositórios, com uma gama de padrões variados mais ou menos detalhados passa a ser questionada. Ao que Costello (2009) prontamente contestaria com sua colocação de os metadados por si só podem não ser suficientes para possibilitar o reuso e adequação dos dados compartilhados à investigações futuras que não as que lhes deram origem.

Nesta mesma linha de pensamento, Chao (2015) esclarece que isso se dá porque a descrição detalhada dos métodos adotados na coleta e no tratamento dos dados não está incluída na maioria dos padrões de metadados existentes. Nesse sentido, os argumentos apresentados pelos autores supracitados somam-se às intenções, definidas por Chavan e Penev (2011), dos artigos de dados de atuarem como uma publicação orientada a dados, com o objetivo descrever o como, por quê, onde e quando dos dados de pesquisa em um modelo creditável e citável,

corroboradas por Callaghan *et al.* (2012), Candela *et al.* (2015) e mais recentemente reforçadas por Kim (2020).

Sobressaindo a outros veículos de divulgação e publicação de dados, Silva, Stueber e Witt (2022) expõe como principais vantagens dos artigos de dados: 1) a maior visibilidade de *datasets* compartilhados, através da indexação dos artigos em bases de dados de ampla cobertura e da citação dupla do artigo de dados e do conjunto de dados compartilhado em repositório; 2) reuso dos dados facilitado pelo detalhamento da descrição dos conjuntos de dados; 3) menor ou nenhuma restrição no tamanho e extensão dos *datasets* compartilhados, uma vez que estes estão armazenados em repositórios, externos na maioria das vezes; e 4) qualidade comprovada por revisão dos pares e direitos de autoria garantidos pelo uso e citação do *paper* por terceiros.

Os autores não isentam as desvantagens que se apresentam atualmente contra este modelo de publicação em sua deliberação e apresentam como principais empecilhos para a adoção deste meio de comunicação orientada à dados de pesquisa: 1) a necessidade de compartilhamento dos conjuntos de dados em repositórios sem que se tenha conhecimento dos mesmos; 2) o tempo a ser despendido para a produção de mais uma documentação para os dados; e 3) a cobrança de taxas elevadas para sua publicação (Silva; Stueber; Witt, 2022).

De modo particular, ao levar esta discussão a termo com pares científicos em eventos da área da Ciência da Informação no Brasil e Conferências internacionais sobre o Acesso Aberto, percebeu-se que os argumentos desfavoráveis em relação à produção e publicação de artigos de dados não costumam variar para além das problematizações apontadas acima. A questão dos *data papers* quando levantada em meio à outras questões que permeiam a área e o movimento de abertura dos moldes científicos não se apresenta como um desafio muito grande para além do desconforto natural com a ideia de desperdício do tempo com uma matéria de menor importância numa realidade em que o uso de identificadores persistentes e metadados específicos podem contribuir com o reuso dos dados de pesquisa de modo geral.

Nesse sentido, é raro observar uma inclinação em favor da construção de uma visão mais favorável sobre esse modelo de publicação visando seu melhor aproveitamento nas etapas do processo científico. Apesar disso, nos últimos cinco anos têm se observado nesses mesmos cenários o aumento tímido porém progressivo de pesquisas que tenham como objeto principal de suas indagações as revistas e os artigos de dados, isto é, os *data journals* e *data papers*.

Tornando à questão de interesse nesta pesquisa, é recobrada a colocação de que não há um padrão comum para os artigos de dados nas diferentes áreas do conhecimento (Callaghan *et al.*, 2012; Candela *et al.*, 2015; Chavan; Penev, 2011). Apesar da realidade observada na

presente pesquisa demonstrar não ter havido uma mudança definitiva nesta instância, percebe-se que a estruturação dos artigos de dados tem base em pelo menos 4 pontos comuns: a) descrição da autoria relacionada ao artigo de dados e à produção dos conjuntos de dados; b) descrição técnica e contextual do conjunto de dados; c) descrição da metodologia de criação/coleta e análise/tratamento dos dados; e d) descrição do potencial de reúso e termos e condições de uso dos dados.

O que se reflete nos resultados desta pesquisa são as ocorrências de elementos apontados pela literatura mais recente sobre o tema como essenciais na descrição destes pontos (Quadro 3, seção 3.2.2).

Novamente, conforme apontado no início desta discussão, poucos esforços envolveram a perspectiva de pesquisadores interessados no reúso de dados de pesquisa no desenvolvimento de modelos ideais para a descrição satisfatória de conjuntos de dados com a intenção de fomentar seu reaproveitamento e aplicação em novos estudos, dentre estes, destaca-se a iniciativa de Faniel, Frank e Yakel (2019) que utilizaram os dados qualitativos de sua pesquisa produzida entre os anos de 2012 e 2013 com o objetivo de capturar a opinião de pesquisadores das Ciências Sociais, Arqueologia e Zoologia sobre os elementos que lhes são essenciais para a reutilização de dados publicados por outros pesquisadores.

É a partir deste estudo que elementos como “Reputação e histórico do repositório” e “Reúso anterior”, dentre outros surgem na listas dos elementos defendidos por Kim (2020) na lista de descritores reprodutíveis e aplicáveis essencialmente na estrutura de artigos de dados eficazes, independentemente da área do conhecimento em que se inserem.

Considerando a perspectiva dos pesquisadores quanto às informações que podem determinar se eles escolheriam ou não reutilizar um conjunto de dados com base em sua descrição, Faniel, Frank e Yakel (2019) destacam como categoria de maior interesse a relacionada à Contextualização da produção dos *datasets*/coleta dos dados, com quatro dos seus sete elementos estando entre os mais buscados pelos pesquisadores ao considerar um *dataset* para reúso em suas investigações, a saber: a metodologia de coleta/produção dos dados, a cobertura dos dados, a identidade do responsável pela produção coleta do dados e a metodologia de análise e tratamento dos dados.

Refletindo a ordem de apresentação dos resultados obtidos na presente pesquisa e os resultados observados por Faniel, Frank e Yakel (2019), temos que a data de criação dos dados é utilizada por pesquisadores como base para o desenvolvimento os recortes e amostras dos seus estudos, bem como para identificar pontos de mudança no contexto da produção dos dados primários diante da influência de fatores internos e externos, que podem influenciar a

interpretação dos dados. Observa-se um número maior de não ocorrências explícitas deste elemento nas orientações das revistas exploradas na atual pesquisa (pouco mais de 77% das revistas não solicitam esta informação).

Em contrapartida, mais de 84% das revistas consultadas solicita a descrição dos processos de coleta/produção dos conjuntos de dados em suas orientações, que para os pesquisadores consultados por Faniel, Frank e Yakel (2019) é uma informação determinante na ponderação da aplicabilidade da combinação de diferentes conjuntos de dados de uma mesma natureza.

Quanto à descrição do formato dos dados representados nos artigos de dados, pouco mais de 80% das orientações investigadas não requer essa informação de forma direta. Enquanto na perspectiva dos pesquisadores que reusam dados, esta informação pode indicar o nível de acessibilidade dos dados, considerando as ferramentas que estão à sua disposição para execução de seus estudos (Faniel; Frank; Yakel, 2019).

Já o detalhamento da cobertura dos dados pode ter maior ou menor importância de acordo com a área de estudo dos pesquisadores interessados em reutilizar um determinado conjunto de dados. A pesquisa em tela reflete que pouco mais de 51% das revistas consultadas, mais da metade da amostra selecionada, solicita informações sobre a cobertura espacial, temporal e taxonômica dos dados compartilhados. Quanto ao interesse dos pesquisadores nesta informação, Faniel, Frank e Yakel (2019) descrevem que principalmente os zoologistas utilizam esta informação para tomada de decisão sobre a relevância e acurácia dos dados disponíveis em sua área.

Sobre a identificação do produtor dos dados de pesquisa nos artigos de dados, observa-se que mais de 66% das revistas apenas solicita os dados de autoria dos autores do artigo de dados, sem que necessariamente precisem identificar o(s) responsável(eis) pela coleta e/ou produção dos dados e sua afiliação. O que na visão dos pesquisadores tratados por Faniel, Frank e Yakel (2019) pode dificultar a avaliação das competências e da reputação da(s) pessoa(s) que geraram os dados que se pretende reutilizar que, ainda segundo estes, é frequentemente o primeiro passo para avaliar a confiabilidade de um conjunto de dados.

Outros elementos relevantes de acordo com as autoras são a descrição dos métodos empregados na análise e tratamento dos dados disponibilizados e de seus objetivos. Este primeiro obteve 83% de ocorrências na atual pesquisa e segundo os resultados obtidos por Faniel, Frank e Yakel (2019) ajudam os pesquisadores a identificar os possíveis impactos desse processo na representação dos dados disponíveis para reuso. Já no que tange aos objetivos dos dados, com ocorrência superior a 69% nas orientações de publicação das revistas consultadas,

as autoras explicam que, junto à consulta dos trabalhos que citam o conjunto de dados e sua metodologia de coleta e análise, tornam-se importante para o entendimento de quais questões de pesquisa os dados compartilhados podem responder e se estes se encaixam nas necessidades de reuso dos pesquisadores interessados.

Em paralelo, os dados faltantes são apontados pelos pesquisadores consultados por Faniel, Frank e Yakel (2019) como imprescindíveis para a antecipação da capacidade de alteração dos seus objetivos de reuso e análise de um determinado conjunto de dados. Isso porque um dado faltante pode significar que o dado não foi encontrado, que o dado foi anulado ou excluído objetivamente do conjunto obtido, o que pode acarretar em viés e distorcer os resultados da amostra. De forma preocupante se observa que 80% das orientações analisadas nesta pesquisa não solicita a descrição de dados faltantes nos artigos de dados.

Dentre os elementos mais valorosos para a determinação do potencial reuso de um conjunto de dados na perspectiva dos pesquisadores que por ele possam se interessar, se inserem os elementos relacionados à categoria 3 de análise, apresentada na seção 4.2.3 desta pesquisa, que versa sobre as informações de compartilhamento e reuso de dados.

Quanto à proveniência dos dados, cujo resultado obtido demonstra que mais de 77% das orientações consultadas não requer informações quanto aos trabalhos relacionados ao conjunto de dados e sua origem, o estudo de Faniel, Frank e Yakel (2019) aponta que esta é considerada pelos pesquisadores como a descrição clara da relação do conjunto de dados com suas versões anteriores e outras fontes utilizadas. Os pesquisadores consultados pelas autoras sugerem que a possibilidade de traçar um histórico linear do ciclo de vida dos dados nos permite julgar melhor sua credibilidade e confiabilidade.

No que se refere à descrição do reuso primário ou anterior dos dados por outros autores ou da combinação de dados anteriores com dados recentes, os resultados obtidos por Faniel, Frank e Yakel (2019) demonstram que os pesquisadores utilizam esta informação para determinar a quantas vezes um conjunto de dados já foi reutilizado, já que dados que já tenham sido aplicados em outras publicações indicam um nível elevado de aceitação pela comunidade científica e revisão de integridade por pares. Este elemento foi um dos menos solicitados pelas revistas, com pouco menos de 4% de ocorrências observadas.

Por fim, no que diz respeito à descrição do reuso potencial dos dados nos artigos de dados, cuja ocorrência nas orientações para a submissão de artigos de dados investigada na presente pesquisa se limitou à pouco mais de 27%, Faniel, Frank e Yakel (2019) destacaram que, especialmente os pesquisadores em Ciências Sociais demonstraram apreço pelo

fornecimento desta informação, principalmente na explanação sobre como trabalhar com os dados, melhor combiná-los, medi-los, etc.

Ao comparar-se as informações dispostas nos modelos explorados de *templates* e orientações para publicação de artigos de dados com a perspectiva de pesquisadores que reutilizam dados de pesquisa ou têm interesse em fazê-lo, fica evidente que o que os usuários consideram relevante para sua decisão de reutilizar um conjunto de dados ou não, podem não ser os mesmos elementos que as revistas refletem na maioria das estruturas atuais propostas para a submissão e publicação de artigos de dados.

Tendo isso em vista, propõem-se na seção seguinte um modelo adaptável de artigo de dados que reúne os principais elementos gráficos essenciais para publicações científicas (metadados de autoria, contexto e assunto) e outros aspectos citados por pesquisadores, conforme demonstrado na Literatura, como decisórios para o reúso adequado de dados de pesquisa.

6 PRIMEIROS PASSOS PARA A CONSTRUÇÃO DE UM PADRÃO ADAPTÁVEL

Levando em consideração os resultados apresentados e discutidos acima, foi elaborada uma proposta de modelo padrão para elaboração de artigos de dados adaptável às diferentes áreas do conhecimento. O modelo pensado se aproxima da estrutura comum a *data papers* multidisciplinares, como o *Bmc Research notes*, *Data*, *Data In Brief*, *F1000research*, etc. explorados nesta pesquisa, com atenção a todos os elementos compilados por Kim (2020) a partir de estudos relacionados à investigação do conteúdo que compõe um artigo de dados.

Desse modo, o seguinte modelo é proposto:

Folha de Identificação
Título do Artigo de Dados <i>*obrigatório</i> [Este elemento pode seguir à proposição de cada revista quanto à quantidade de caracteres e a soma da palavra “dataset” ou “conjunto de dados” a fim de esclarecer o tipo de publicação a que se refere]
Autores do data paper <i>*obrigatório</i> [Neste campo devem ser incluídos o nome e sobrenome de todos os autores envolvidos no processo de redação do artigo, sua filiação e ORCID. Os detalhes da descrição da afiliação compreendem: nome da instituição acadêmico-científica com a qual os autores possuem vínculo, cidade, estado e país de localização. Para a destinação do envolvimento de cada um dos autores listados, recomenda-se o uso do CRediT.]
Informações de contato do autor correspondente [Pessoa responsável pela submissão do artigo e gestão da negociação de publicação com a editora. Devem ser incluídos neste campo o nome e sobrenome do autor correspondente, e-mail de contato.]
Produtor dos dados <i>*obrigatório</i> [Pessoa(s) físicas ou entidades coletivas responsável(eis) pela produção do conjunto de dados ou coleta dos dados de pesquisa aos quais o artigo submetido se refere. Devem ser incluídos neste campo o nome e sobrenome do produtor dos dados, título acadêmico-científico (caso possua), afiliação, e-mail de contato (para atender a dúvidas e solicitações acerca do conjunto de dados organizado/produzido) e ORCID. Os detalhes da descrição da afiliação compreendem: nome da instituição acadêmico-científica com a qual o produtor possui vínculo, cidade, estado e país de localização.]
Palavras-chave <i>*obrigatório</i> [Termos adotados para indexação e recuperação do artigo de dados em bases de dados. Podem ser incluídas de 3 a 5 palavras-chave ou mais, relacionadas ao assunto do(s) conjunto(s) de dados descrito(s) no artigo a depender da política de indexação adotada em cada revista. As palavras-chave devem iniciar com a descrição da construção do assunto partindo de uma visão mais ampla para o assunto específico abordado pelos dados. Ex.: Ciência Aberta; Ciência da Informação; Comunicação Científica; Revistas de dados; Artigos de dados.]
Contexto <i>*obrigatório</i> [Breve descrição dos objetivos que motivaram a coleta ou produção dos dados. Caso se trate dos dados oriundos de um projeto de pesquisa, forneça também uma breve descrição do projeto e dos objetivos.]
Descrição dos dados

[Estruturada como seção específica, a descrição dos dados reúne todos os elementos relacionados à identificação e caracterização dos dados de pesquisa descritos]

Título do conjunto de dados **obrigatório*

[O título do conjunto de dados não pode ser o mesmo do *data paper*, pelo contrário, este deve ser idêntico ao título que foi atribuído no compartilhamento do mesmo em repositório. Caso o título do conjunto de dados no repositório seja o mesmo utilizado no artigo de dados, é preciso alterar um dos dois, mas que se possa distinguir nos buscadores um material do outro.]

Identificador persistente do conjunto de dados **obrigatório*

[O Identificador Persistente (DOI ou PURL) atribuído ao *dataset* durante o armazenamento dos mesmos em repositório.]

Data de criação/coleta dos dados representados no artigo **obrigatório*

[A data de criação dos conjuntos de dados facilita a leitura do contexto temporal em que os dados foram produzidos e coletados.]

Formato de exposição dos dados **obrigatório*

[O formato de exposição dos dados evidencia as necessidades de adaptação de softwares e hardware levando em conta as ferramentas de que pesquisadores interessados em reutilizar um conjunto de dados pode precisar.]

Natureza/tipo dos dados disponíveis

[A natureza ou tipo de dados disponíveis evidenciam se os dados descritos se tratam de arquivos numéricos, textuais, multimídia, etc.]

Versão dos dados **obrigatório*

[A versão dos dados representados pelo artigo no que se refere ao seu nível de processamento ou status de atualização (dados brutos, dados limpos, dados processados, etc. - variando de acordo com a área do conhecimento).]

Cobertura dos dados **obrigatório*

[A cobertura dos dados deve destacar a cobertura espacial ou geográfica, cobertura temporal e/ou cobertura taxonômica dos dados, a depender da natureza dos dados descritos.]

Tamanho do arquivo de dados **obrigatório*

[O tamanho do arquivo ou volume de dados contido no *dataset* representado pelo artigo deve ser exposto para fins de compreensão da demanda do espaço de armazenamento e reprocessamento, caso seja necessário o download do arquivo e a capacidade das ferramentas utilizadas seja limitada.]

Linguagem dos dados

[A linguagem dos dados compreende a descrição da linguagem de processamento na qual os dados se encontram. Este elemento pode ser melhor empregado na representação de dados computacionais.]

Método empregado na coleta/produção dos dados **obrigatório*

[Este elemento demanda a descrição detalhada do processo adotado na coleta/produção dos dados descritos. Agrega informações sobre os instrumentos usados (Ex.: hardware ou software específico, incluindo marca/modelo, parâmetros da coleta, questionário, entrevista, etc.), bem como os critérios de inclusão/exclusão adotados na seleção dos dados e a forma de organização/preparação dos dados para análise.]

Método empregado na análise dos dados **obrigatório se os dados tiverem sido analisados*

[Este elemento demanda a descrição detalhada do processo adotado no processamento e tratamento dos dados para obtenção de resultados, esclarecendo a natureza da pesquisa (Ex.: qualitativa, quantitativa, experimental, etc.).]

Dados faltantes **obrigatório*

[Os dados faltantes apresentam as limitações do conjunto de dados descrito (Ex.: dados perdidos, excluídos ou não processáveis).]

Compartilhamento e reuso dos dados

[Esta seção foi pensada como complemento para orientar os leitores quanto ao compartilhamento e reuso dos dados descritos no artigo]

Proveniência dos dados **obrigatório*

[Este elemento se refere ao histórico de origem dos dados, em especial à publicação que deu origem ao conjunto de dados descrito, considerando: título, autoria, data de publicação (se houver), status de publicação (caso ainda esteja no prelo ou em construção), e localização (onde se encontra armazenada). Caso existam mais de uma publicações originalmente relacionadas ao conjunto de dados, liste todas elas.]

Repositório na qual os dados estão disponibilizados

[Este elemento faz alusão a apresentação de uma breve descrição do repositório, arquivo ou base de dados em que o conjunto de dados referido no artigo se encontra, tendo em conta: o nome do repositório, sua organização/instituição mantenedora, a natureza do repositório (institucional, de dados, temático, etc.) e se este possui política de preservação (fornecer link).]

Preservação dos dados **obrigatório*

[Este elemento evidencia se o conjunto de dados se encontra sob a proteção de alguma política de preservação de dados (da própria revista, do repositório, estabelecida pelos autores, etc.) e quais são os termos ou as diretrizes nela incluídos.]

Reuso anterior **obrigatório*

[O elemento de reuso anterior demanda o esclarecimento da combinação de dados de pesquisa de outras fontes na construção do conjunto de dados descrito, ou ainda de outras investigações que já tenham reusado direta ou indiretamente o conjunto de dados representado.]

Reuso potencial **obrigatório*

[Este elemento se encarrega das orientações de reuso sugeridas pelos autores do artigo de dados para melhor aproveitamento dos dados descritos, podendo incluir mas não se limitar a: descrição de outras análises que possam ser empregadas dentro de um mesmo contexto, uso dos dados para responder a outras questões de pesquisa pertinentes, com sugestão de questões a serem tratadas, indicação de como os dados podem ser combinados com outros dados ou interpretados em outro contexto que não o de origem, etc.]

Termos e condições de uso **obrigatório*

[Os termos e condições de uso do conjunto de dados se refere ao tipo de licença sob a qual estes foram compartilhados/publicados e sob quais condições estes podem ser aproveitados em outras investigações.]

Referências

[Esta seção prevê a descrição ordenada da lista de referências utilizadas na elaboração do artigo de dados, incluindo trabalhos publicados nos quais o conjunto de dados descrito tenha sido utilizado.]

Embora tenham sido retratados no modelo proposto alguns elementos opcionais, isto é, que podem ser mais utilizados em áreas do conhecimento específicas, compreende-se que a presença destes amplia o conhecimento dos autores interessados em submeter artigos de dados

a respeito das informações que podem ser relevantes para a valorização e credibilidade de um conjunto de dados e a promoção do reuso consciente e eficaz dos dados de pesquisa.

Foram subtraídos desta proposta elementos compreendidos como parte integrante do padrão de publicação das revistas, como a sessão de Reconhecimento e Financiamento.

Como limitação desta etapa é importante destacar que modelo de artigo de dados proposto não foi testado.

Por fim, na seção 7 desta Dissertação apresentam-se as conclusões alcançadas nesta pesquisa.

7 CONCLUSÃO

A presente pesquisa comparou as características dos modelos de artigos de dados orientados e fornecidos pelas revistas indexadas na *Scopus* e na *Web of Science*, com o propósito de elaborar uma proposta de diretrizes para a implementação de um modelo de artigo de dados adaptável que sirva de apoio aos periódicos científicos brasileiros.

O processo teve como principal referencial o trabalho de Kim (2020), enriquecido pelos resultados obtidos no estudo de Faniel, Frank e Yakel (2019), com objetivo de identificar os componentes primordiais de um artigo de dados, de acordo com a literatura e a perspectiva de pesquisadores que aplicam o reúso de dados em suas investigações, buscando facilitar a compreensão das melhorias necessárias às orientações oferecidas por revistas de dados para a produção de *data papers*.

Das 117 revistas incluídas e exploradas nesta pesquisa, 20 estiveram presentes direta ou indiretamente na pesquisa de Kim (2020): *Scientific Data*, *Ecological Research*, *Ecology*, *Geoscience Data Journal*, *Earth System Science Data*, *Data in Brief*, *Frontiers in Marine Science*, *GigaScience*, *BioInvasion Records*, *International Journal of Robotics Research*, *Journal of Open Archaeology Data*, *Biodiversity Data Journal*, *BMC Genomics*, *Comparative Cytogenetics*, *Journal of Hymenoptera Research*, *Neobiota*, *Phytokeys*, e *Zookeys*. A análise atual destas revistas demonstrou diferenças sutis nos moldes disponibilizados para a elaboração de artigos de dados. Para fins de exemplificação, foi percebida a solicitação de novos elementos nos templates das revistas *Data*, *Data In Brief*, *Ecology* e *Scientific data*, indicando melhorias. Enquanto isso, na revista *Geoscience Data Journal* a ausência de alguns dos elementos identificados anteriormente por Kim (2020) se destacam.

As diferenças encontradas demonstram o avanço e adaptabilidade dos moldes fornecidos pelas revistas para a elaboração de artigos de dados nos últimos anos. Avanços como esses possibilitam a abertura para indagações futuras acerca das motivações de editores científicos ao planejarem tais mudanças, ou ainda à exploração da opinião dos autores quanto ao detalhamento das informações dos seus conjuntos de dados nestes modelos de publicação, como abordado por Faniel, Frank e Yakel (2019).

Nesse sentido, Kim (2020) expressa que a principal motivação para a publicação de artigos de dados é fazer com que os dados de pesquisa se tornem facilmente reutilizáveis e reproduzíveis. Para tanto, estes artigos devem fornecer uma descrição extensa dos dados que ofereçam informações contextuais e aplicáveis o suficiente para o reaproveitamento dos dados fornecidos. Contudo, ao comparar os resultados da presente pesquisa com o panorama

apresentado na pesquisa de Faniel, Frank e Yakel (2019) observa-se que dos nove elementos mais destacados pelos pesquisadores das áreas de arqueologia, ciências sociais e zoologia como facilitadores do seu processo de decisão sobre o reúso ou não de um conjunto de dados, apenas três estão presentes em mais de 50% das revistas consultadas.

Ao se ponderar a efetividade dos artigos de dados em seu objetivo principal, conforme estabelecido e reforçado por Kim (2020) de fomentar o compartilhamento e reúso de dados nas mais diversas áreas do conhecimento, é importante levar em conta até que ponto as orientações e formatos disponíveis para construção e submissão de artigos de dados contemplam as necessidades informacionais para tomada de decisão dos pesquisadores sobre o reúso de dados em suas investigações. Essa visão torna possível a reflexão dos moldes atuais e a formulação de propostas para a construção de um padrão adaptável de *data paper* cujas orientações sejam claras e concisas o suficiente para garantir o melhor entendimento das propriedades dos conjuntos de dados compartilhados e a garantia da qualidade e credibilidade dos mesmos.

Apesar disso, entende-se que os primeiros passos para o surgimento de um padrão de campos e recomendações para a publicação de artigos de dados já vêm sendo fundamentados. Deste modo, a presente pesquisa sugere investigações mais apuradas e aprofundadas em contexto acadêmico-científico para melhor compreender as motivações que norteiam a construção das diretrizes para publicação de artigos de dados. Nesse cenário, ainda é possível indagar a grupos de pesquisadores que já tenham reusado dados ou que tenham publicado artigos de dados sobre suas experiências com seus processos relacionados, desde o compartilhamento dos dados em repositório até a submissão, revisão e publicação dos artigos.

Ao buscar-se observar a visão de editores científicos sobre os elementos que compõem os modelos desenvolvidos em suas revistas, e de pesquisadores-autores a respeito de sua vivência no processo de publicação de artigos de dados, espera-se facilitar a identificação de melhorias possíveis e padrões que possam ser efetivamente aplicáveis. De modo geral, entende-se que o processo de comunicação científica se torna eficaz na medida em que os objetivos dos meios de comunicação e o conteúdo disseminado por eles conseguem ser claros e úteis àqueles que os consomem.

REFERÊNCIAS

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, [s. l.], v. 10, n. 2, 2014. Disponível em: <https://doi.org/10.18617/liinc.v10i2.749>. Acesso em: 13 mar. 2024.

ALBAGLI, Sarita. Ciência Aberta: movimento de movimentos. *In*: SHINTAKU, Milton; SALES, Luana Farias (Org.) **Ciência aberta para editores científicos**. Botucatu: ABEC Brasil, 2019. p. 15-20. Disponível em: <http://dx.doi.org/10.21452/978-85-93910-02-9.cap2>. Acesso em: 13 mar. 2024.

ARAUJO, L.; PRÍNCIPE, E. Adesão das revistas brasileiras de saúde às práticas de ciência aberta pelos critérios SciELO Brasil. **Abec Meeting**, [s. l.], p. 1-8, 2021. Disponível em: <https://doi.org/10.21452/abecmeeting2021.7>. Acesso em: 13 mar. 2024.

ARAUJO, Lyvia Rocha de Jesus; PRÍNCIPE, Eloisa; ALENCAR, Maria Simone de Menezes. Data papers brasileiros: análise da publicação de artigos de dados indexados na web of science e scopus. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: IBICT/UFRJ, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/177/479>. Acesso em: 11 jun. 2022.

AVENTURIER, Pascal. Dados de pesquisa: descrição do ciclo de vida de gestão de dados. **A publicação científica**, [s. l.], 2016. Disponível em: <https://publicient.hypotheses.org/1359>. Acesso em 10 maio 2024.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016

BARTLING, S.; FRIESIKE, S. Towards Another Scientific Revolution. *In*: BARTLING, S.; FRIESIKE, S. (eds). **Opening Science: the Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing**. New York: Springer, 2014. p. 3-15. Disponível em: https://doi.org/10.1007/978-3-319-00026-8_1. Acesso em: 26 jun. 2024.

BOOTH, Andrew; PAPAIOANNOU, Diana; SUTTON, Anthea. The literature review: its role within research. *In*: BOOTH, Andrew; PAPAIOANNOU, Diana; SUTTON, Anthea. **Systematic approaches to a successful literature review**. [Internet], 2012. Disponível em: <https://uk.sagepub.com/en-gb/eur/systematic-approaches-to-a-successful-literature-review/book244586>. Acesso em: 10 out. 2021.

BETHESDA Statement On Open Access Publishing. 2003. Disponível em: <https://www.ouvrirlascience.fr/bethesda-statement-on-open-access-publishing/>. Acesso em: 04 set. 2022.

BERLIN Declaration On Open Access To Knowledge In The Sciences And Humanities. 2003. Disponível em: https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf. Acesso em: 04 set. 2022

BUDAPEST OPEN ACCESS INITIATIVE. **Declaração da Iniciativa de Acesso Aberto de Budapeste**. 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/read/>. Acesso em 13 jul. 2022.

BUDAPEST OPEN ACCESS INITIATIVE. **Dez anos da iniciativa de Budapeste em acesso aberto**: a abertura como caminho a seguir. Tradução Carolina Rossini. BOAI – Budapest Open Access Initiative. 2012. Disponível em: <https://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation>. Acesso em: 04 set. 2022.

CALLAGHAN, Sarah. et al. Making data a first class scientific output: data citation and publication by NERC's environmental data centres. **Int J Digit Curation**, v. 7, n. 1, p. 107-113, 2012. Disponível em: <http://www.ijdc.net/article/view/208>. Acesso em: 13 fev. 2022.

CANDELA, Leonardo. *et al.* **Data journals**: a survey. **Journal of the Association of Information Science and Technology**, v. 66, p.1747-1762, 2015. Disponível em: <https://doi.org/10.1002/asi.23358>. Acesso em: 26 jun. 2024.

CHAO, T. C. Mapping methods metadata for research data. **International Journal of Digital Curation**, v. 10, n. 1, p. 82-94, 2015.

CHAVAN, Vishwas; PENEV, Lyubomir. The data paper: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, v. 12, Supl. 2, 2011. Disponível em: <https://doi.org/10.1186/1471-2105-12-S15-S2>. Acesso em: 26 jun. 2024.

CHEN, Ya-Ning. An analysis of characteristics and structures embedded in data papers: a preliminary study. **Libellarium**: journal for the research of writing, books, and cultural heritage institutions, v. 9, n. 2, 2016. Disponível em: <http://dx.doi.org/10.15291/libellarium.v9i2.266>. Acesso em: 26 jun. 2024.

CHIN, G; LANSING, C. S. Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 2004, Chicago. **Anais [...]**. Chicago:ACM, 2004. Disponível em: <https://doi.org/10.1145/1031607.1031677>. Acesso em: 26 jun. 2024.

COSTELLO, M. J. Motivating online publication of data. **BioScience**, v. 59, p. 5418-5427, 2009. Disponível em: <https://doi.org/10.1525/bio.2009.59.5.9>. Acesso em: 26 jun. 2024.

CREATIVE COMMONS. **About CC Licenses**. [201-?]. Disponível em: <https://creativecommons.org/about/cclicenses/>. Acesso em: 04 set. 2022.

DUARTE, Emeide Nóbrega. *et al.* Estratégias metodológicas adotadas nas pesquisas de iniciação científica premiadas na UFPB: em foco a série “iniciados”. **Encontros Bibli**, Florianópolis, v. 14, n. 27, p. 170-190, 2009. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2009v14n27p170>. Acesso em: 22 maio 2022.

FANIEL, I. M.; FRANK, R. D.; YAKEL, E. Context from the data reuser's point of view. **J Doc**, v. 75, n. 6, p. 1274-1297, 2019. Disponível em: <https://doi.org/10.1108/JD-08-2018-0133>. Acesso em: 26 jun. 2024.

FAUSTO, Sibeles. Evolução do acesso aberto: breve histórico. **SciELO em Perspectiva**, 2013. Disponível em: <https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/#.YxUKsXvMK01>. Acesso em: 04 set. 2022.

FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. In: BARTLING, S., FRIESIKE, S. (org.) **Opening Science**. New York: Springer, 2014. p. 17-48. Disponível em: https://doi.org/10.1007/978-3-319-00026-8_2. Acesso em: 26 jun. 2024.

FJÄLLBRANT, Nancy. Scholarly communication: historical development and new possibilities. In: INTERNATIONAL ASSOCIATION OF SCIENTIFIC AND TECHNOLOGICAL UNIVERSITY LIBRARIES, 18., 1997, Gothenburg. **Anais [...]**. Gothenburg: IATUL, 1997. Disponível em: <https://docs.lib.purdue.edu/iatul/1997/papers/5>. Acesso em: 28 ago. 2022.

FREITAS, Marília Augusta de; LEITE, Fernando César Lima. Atores do sistema de comunicação científica: apontamentos para discussão de suas funções. **Informação & Informação**, v. 24, n. 1, p. 273-299, 2019. Disponível em: <https://doi.org/10.5433/1981-8920.2019v24n1p273>. Acesso em: 26 jun. 2024.

GARCÍA-GARCÍA, Alicia; LÓPEZ-BORRULL, Alexandre; PESET, Fernanda. Data journals: eclosión de nuevas revistas especializadas en datos. **El Profesional de la Información**, v. 24, n. 6, p. 845-854, 2015. Disponível em: <http://dx.doi.org/10.3145/epi.2015.nov.17>. Acesso em: 26 jun. 2024.

GREEN, Samantha. An Illustrated History of Open Science. **Society leaders**, 2017. Disponível em: <https://www.wiley.com/network/societyleaders/open-science/an-illustrated-history-of-open-science>. Acesso em: 04 set. 2022.

HORIZON 2020. **What is Horizon 2020?**. [201-?]. Disponível em: <https://wayback.archive-it.org/12090/20220124080448/https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Acesso em: 04 set. 2022.

JEFFERIES, Neil *et al.* Data2paper: Giving Researchers Credit for Their Data. **Publications**, v. 7, n. 2, p. 1-6, 2019. Disponível em: <https://www.mdpi.com/2304-6775/7/2/36/htm>. Acesso em: 29 maio 2022.

JIAO, Chenyue; DARCH, Peter T. The role of the data paper in scholarly communication. **Proceedings of the Association for Information Science and Technology**, v. 57, n. 1, p. 1-6, 2020. Disponível em: <https://doi.org/10.1002/pa2.316>. Acesso em: 26 jun. 2024.

MANIFESTO de Apoio ao Acesso Livre à Informação Científica. IBICT, 2005. Disponível em: <https://livroaberto.ibict.br/Manifesto.pdf>. Acesso em: 04 set. 2022.

KIM, Jihyun. An analysis of data paper templates and guidelines: types of contextual information described by *data journals*. **Sci Ed.**, v. 7, n. 1, p. 16-23, 2020. Disponível em: <https://www.escienceediting.org/upload/kcse-185.pdf>. Acesso em: 13 fev. 2022.

LASSERSON, Toby J.; THOMAS, James; HIGGINS, Julian P. T. Starting a review. In: HIGGINS, Julian; THOMAS, James. **Cochrane handbook for systematic reviews of**

interventions-Wiley-Blackwell. Reino Unido: BMJ, 2020. p. 1010-1012. Disponível em: <https://training.cochrane.org/handbook/current>. Acesso em: 10 out. 2021.

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

LI, Kai; JIAO, Chenyue. The data paper as a socio-linguistic epistemic object: a content analysis on the rhetorical moves used in data paper abstracts. **Preprint**, [s. l.], p. 1-21, 2021. Disponível em: <https://arxiv.org/ftp/arxiv/papers/2106/2106.01083.pdf>. Acesso em: 29 maio 2022.

MASUZZO, Paola; MARTENS, Lennart. Do you speak open science? Resources and tips to learn the language. **PeerJ Preprints**, [s. l.], v. 5, p. 1-22, 2017. Disponível em: <https://doi.org/10.7287/peerj.preprints.2689v1>. Acesso em: 26 jun. 2024.

MEADOWS, A.J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. [s. l.]: OECD, 2007. *E-book*. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 05 jul. 2022.

OLIVEIRA, Adriana Carla Silva de. Publicações científicas ampliadas. *In*: SHINTAKU, Milton; SALES, Luana Farias (org.) **Ciência aberta para editores científicos**. São Paulo: ABEC Brasil, 2019. p. 103-107. Disponível em: <http://dx.doi.org/10.21452/978-85-93910-02-9.cap15>. Acesso em: 26 jun. 2024.

PACKER, Abel L. O modelo SciELO de publicação como política pública de acesso aberto. **SciELO em perspectiva**, 2019. Disponível em: <https://blog.scielo.org/blog/2019/12/18/o-modelo-scielo-de-publicacao-como-politica-publica-de-acesso-aberto/#.YxULdnvMK01>. Acesso em: 04 set. 2022.

PINHEIRO, L. V. R.; LOUREIRO, J. M. M. Traçados e limites da ciência da informação. **Ciência da Informação**, v. 24, n. 1, p. 1-19, 1995. Disponível em: <https://revista.ibict.br/ciinf/article/view/609>. Acesso em: 26 jun. 2024.

PONTIKA, Nancy *et al.* Fostering open science to research using a taxonomy and an elearning portal. IKNOW: INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA DRIVEN BUSINESS, 15., 2015, Austria. **Anais [...]**. Austria, 2015. Disponível em: <https://doi.org/10.1145/2809563.2809571>. Acesso em: 26 jun. 2024.

RIBEIRO, Nivaldo Calixto; SILVEIRA, Lúcia da; SANTOS, Sarah Rúbia de Oliveira. **Taxonomia da ciência aberta**. Figshare. 2020. Disponível em: <https://doi.org/10.6084/m9.figshare.12124002.v4>. Acesso em: 26 jun. 2024.

SALES, Luana Farias; SAYÃO, Luís Fernando. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, jan./jun. 2019. Disponível em: <https://revistas.ufrrj.br/index.php/rca/article/view/26337/14573>. Acesso em: 1 jul. 2022.

SANTOS, Solange Maria dos; NASSI-CALÒ, Lilian. Gestão editorial: tendências e desafios na transição para a ciência aberta. In: SILVEIRA, Lúcia da; SILVA, Fabiano Couto Corrêa da (org.). **Gestão editorial de periódicos científicos: tendências e boas práticas**. Florianópolis: BU Publicações /UFSC : Edições do Bosque/UFSC, 2020. p. 17-55. Disponível em: https://seer.ufrgs.br/wp/wp-content/uploads/2020/06/Gest%C3%A3o-Editorial_v06.pdf. Acesso em: 14 nov. 2021.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, 1996.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. *E-book*. Disponível em: <http://www.cnen.gov.br/component/content/article/75-cin/material-didatico-cnen/160-guia-de-gestao-de-dados-de-pesquisa>. Acesso em: 04 set. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. A ciência invisível: por que os pesquisadores não publicam seus resultados negativos? **Informação & Informação**, v. 25, n. 4, p. 98-116, 2020. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/40016>. Acesso em: 22 maio 2022.

SCHÖPFEL, Joachim *et al.* Data papers as a new form of knowledge organization in the field of research data. In: **Colloque international d'isko-france**, 12. Montpellier: ISKO, 2019. Disponível em: <https://halshs.archives-ouvertes.fr/halshs-02284548/document>. Acesso em: 05 jun. 2022.

SENDEROV, V.; GEORGIEV, T.; PENEV, L. Online direct import of specimen records into manuscripts and automatic creation of data papers from biological databases . **Research Ideas and Outcomes**, v. 2, p. e10617, 2016. Disponível em: <https://doi.org/10.3897/rio.2.e10617>. Acesso em: 26 jun. 2024.

SILVA, F. C. C. da .; STUEBER, K.; WITT, A. S. . Análise das publicações de Data Papers sobre Biodiversidade: considerações acerca da preservação digital . **Revista Brasileira de Preservação Digital, Campinas**, v. 3, n. 00, p. e022012, 2022. Disponível em: <https://doi.org/10.20396/rebpred.v3i00.16556>. Acesso em 26 jun. 2024.

TORINO, Emanuelle; ROA-MARTÍNEZ, Sandra Milena; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dados de pesquisa: disponibilização ou publicação?. In: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201. Disponível em: <http://doi.org/10.21452/978-85-93910-04-3.cap15>. Acesso em: 26 jun. 2024.

UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. Brasília: UNESCO Office, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 11 set 2022.

WALTERS, William H. Data journals: incentivizing data access and documentation within the scholarly communication system. **Insights**, [s. l.], v. 33, n. 18, p. 1–20, 2020. Disponível em: <https://doi.org/10.1629/uksg.510>. Acesso em: 26 jun. 2024.

APÊNDICES

APÊNDICE A - A autora



Lyvia Rocha de Jesus Araujo, 26 anos.

Concluiu Bacharelado Biblioteconomia na Universidade Federal do Estado do Rio de Janeiro (UNIRIO) em janeiro de 2021, tendo como produto final de sua formação o Trabalho de Conclusão de Curso “REAS em saúde: boas práticas para sua aplicação e o papel das bibliotecas”, gerado a partir do projeto de pesquisa “Recursos educacionais abertos em Biblioteconomia”, coordenado e orientado pela Profa. Dra. Dayanne da Silva Prudencio.

Entre 2019 e 2020, esteve envolvida nos projetos de iniciação científica da Profa. Dra. Maria Simone de Menezes Alencar, nos quais pode debater e pesquisar a respeito da atuação das práticas de gestão, compartilhamento e reúso de dados de pesquisa na ciência aberta e, de que maneira os dados de pesquisa podem se tornar mais acessíveis e melhor aproveitados. O que impulsionou o interesse pelo formato de publicação de dados de pesquisa em periódicos científicos tornando possível a comunicação dos insumos básicos de investigações científicas (dados de pesquisa) de modo íntegro, seguro e creditável.

Na ocasião, teve a oportunidade de desenvolver os projetos “Construção de questionário sobre práticas e percepções de pesquisadores no uso e compartilhamento de dados de pesquisa (2019)”, no qual colaborou com a Dra. Solange Campos na construção da ferramenta de coleta de dados utilizada em sua pesquisa de Doutorado; e “Gestão de dados de pesquisa em Enfermagem: um projeto piloto no

Website do LabINOVA (2020)”, onde colaborou com os membros do LabINOVA¹³ no desenvolvimento do piloto da estrutura do site principal do laboratório e seu banco de dados.

Em maio de 2021, ingressou no curso de Mestrado do Programa de Pós-Graduação em Ciência da Informação (PPGCI), administrado pelo convênio estabelecido entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Escola de Comunicação da Universidade Federal do Rio de Janeiro (ECO/UFRJ). Onde desenvolveu a presente pesquisa em sua área interesse, como parte do projeto de pesquisa “Comunicação científica: percursos, desafios e perspectivas”, coordenado pela Profa. Dra. Eloisa Príncipe, com o objetivo de analisar do grau de adesão e absorção das novas tecnologias e práticas da ciência aberta pelas revistas científicas brasileiras e pelos atores atuantes no processo de comunicação científica.

Esteve presente na edição de 2020 da *Open Education Conference*, na edição de 2021 do *ABEC Meetings*, da Associação Brasileira de Editores Científicos, nas edições de 2021, 2022 e 2023 da Conferência Lusófona de Ciência Aberta (CONFOA), e nas edições de 2021, 2022 e 2023 do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB). Nestes eventos apresentou os resultados das pesquisas em andamento e participou como ouvinte, integrando-se à comunidade científica e absorvendo seus aprendizados.

Dentre a produção científica gerada neste período, destacam-se as seguintes referências:

- ARAUJO, L. R. J. ; ALENCAR, M. Simone de M. . Construção de questionário sobre práticas e percepções de pesquisadores no uso e compartilhamento de dados de pesquisa. In: 19a Jornada de Iniciação Científica da Unirio, 2020, Rio de Janeiro. Livro de Resumos, 2020. p. 66-68.
- PRUDENCIO, D. S.; Andre Luis Ferreira; ARAUJO, L. R. J. Use of Open Educational Resources in Health Sciences Programs Libraries. 2020.
- ARAUJO, Lyvia Rocha de Jesus; PRUDENCIO, D. S. . REA em bibliotecas de ciências da saúde: uso, gerenciamento e disseminação. In: XXI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação, 2021, Rio de Janeiro. **Anais ...** Rio de Janeiro, 2021.
- ARAUJO, L. R. J.; PRÍNCIPE, E.; ALENCAR, M. Simone de M. Data Papers Brasileiros: análise da publicação de artigos de dados indexados na Web of Science e Scopus. In: XXI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação, 2021, Rio de Janeiro. **Anais ...** Rio de Janeiro, 2021.
- ARAUJO, Lyvia; PRÍNCIPE, Eloísa . Adesão das revistas brasileiras de saúde às práticas de ciência aberta pelos critérios SciELO Brasil. In: Abec Meeting, 2021. **Anais ...** [online], 2021.
- PRUDENCIO, Dayanne da Silva; ARAUJO, Lyvia Rocha de Jesus. REA e bibliotecas: discussões sobre a importância da classificação e representação. **Informação@Profissões**, v. 10, p. 173-188, 2021.

¹³ Laboratório de Ciência Aberta e Dados de Pesquisa para apoio à Inovação, sediado na Universidade Federal do Estado do Rio de Janeiro – Unirio. O LabInova é um *lôcus* de capacitação para que o pesquisador se conscientize da importância de uma ciência mais acessível, transparente, disponível e gratuita. O objetivo geral da Ciência Aberta é aumentar a qualidade, o progresso e o impacto científico e social de pesquisa e do conhecimento, estimulando a inovação. Membro do LabInova desde sua criação, a autora desta pesquisa vem desenvolvendo sua pesquisa com vistas a contribuir ativamente com o *corpus* informacional produzido pelo projeto. Conheça mais em: <https://labinova.tec.br/>

- ARAUJO, L. R. J. ; PRINCIPE, E. ; ALENCAR, M. Simone de M. . Panorama da publicação de artigos de dados em periódicos científicos: uma revisão sistemática. *In: XXII ENANCIB - Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação*, 2022. **Anais ...** Porto Alegre: Ancib, 2022.
- ARAUJO, L. R. J.; PRÍNCIPE, ELOÍSA ; ALENCAR, M. S. M. . Repositórios de dados de pesquisa: critérios core CoreTrustSeal. *In: Conferência Lusófona de Ciência Aberta (ConfOA)*, 2022, Maputo, Moçambique. **Revista Científica da Universidade Eduardo Mondlane**. Maputo, Moçambique: Universidade Eduardo Mondlane, 2022. v. 4.
- ARAUJO, L. R. J. ; PRINCIPE, E. ; ALENCAR, M. Simone de M. Repositórios de dados de pesquisa: critérios core CoreTrustSeal. **Revista Científica da UEM: Série Letras e Ciências Sociais**, [s. l.], v. 4, p. 124-128, 2023.
- ARAUJO, Lyvia; PRINCIPE, Eloísa . Diretrizes editoriais para a publicação de artigos de dados em Ciências da Saúde. **BiblioCanto**, v. 9, p. 114-122, 2023.
- ARAUJO, Lyvia; ALENCAR, M. S. M. Orientações para a formatação de artigos de dados: primeiros passos para a construção de um padrão adaptável. Encontro Nacional de Pesquisa em Ciência da Informação, 23., 2023, Rio de Janeiro. **Anais [...]**. Aracaju: PPGCI/UFSE, 2023.

APÊNDICE B - Plano de Gestão de Dados (PGD)

QUE DADOS VOCÊ IRÁ COLETAR OU CRIAR?

Será realizada uma coleta de dados textuais em formato de planilha XLS e CSV em 2 etapas distintas.

A primeira etapa será realizada para identificar dos títulos dos periódicos indexados na *Scopus* e na *Web of Science* que publicam artigos de dados; A segunda realizará a extração de dados das “orientações aos autores” para a publicação de artigos de dados dos periódicos listados na coleta anterior.

Serão reutilizados nesta pesquisa os dados dos estudos de Kim (2020) e Candela *et al* (2015) sobre as características de artigos de dados (*data papers*) e revistas de dados (*data journals*), estruturados em tabelas (docx) e planilha Excel (xls).

COMO OS DADOS SERÃO COLETADOS OU CRIADOS?

Na primeira etapa realizada, referente à identificação dos títulos de periódicos que publicam artigos de dados, serão conduzidas buscas avançadas nas bases *Scopus* e *Web of Science* pelos artigos de dados publicados entre 1990 e 2022. Os dados desta busca serão exportados em formato de planilha Excel/CSV, e organizados em uma única planilha XLS, onde será aplicada a deduplicação pelos títulos de cada registro recuperado.

Na segunda etapa serão exploradas as orientações aos autores de cada revista listada para a coleta das características essenciais exigidas para a submissão de artigos de dados. Os dados extraídos serão organizados em uma terceira planilha XLS.

Os títulos dos arquivos de dados extraídos em cada coleta seguirão o padrão "Col X_dados_2022-MM-DD" em que deverão ser alterados o número da coleta (X) correspondente ao mês e o dia da última alteração realizada. Durante a realização da pesquisa as planilhas deverão ser armazenadas no drive particular da autora deste plano para que o versionamento dos documentos seja administrado através do *Google Sheets*.

QUE DOCUMENTAÇÃO E METADADOS ACOMPANHARÃO OS DADOS?

Para a melhor organização dos registros recuperados nas coletas citadas anteriormente, serão adotados e, quando necessário, adaptados, os 15 elementos principais do padrão de metadados *Dublin Core*: 1) título; 2) autor; 3) palavras-chave; 4) resumo; 5) editor/periódico/instituição; 6) colaboradores; 7) data de publicação; 8) tipo de recurso; 9) formato; 10) DOI ou *link*; 11) base de dados; 12) idioma; 13) referências; 14) áreas do conhecimento; 15) direitos autorais (licença/acesso aberto ou fechado etc.).

O detalhamento dos procedimentos metodológicos e resultados de todo o ciclo dos dados utilizados nesta pesquisa estará disponível na dissertação inicialmente intitulada “Práticas editoriais para publicação de artigos de dados”, com previsão de defesa para fevereiro de 2024.

COMO VOCÊ ADMINISTRARÁ QUAISQUER QUESTÕES ÉTICAS?

As etapas de coleta de dados conduzidas nesta pesquisa são baseadas em documentos publicados, sem a existência de dados sensíveis que comprometam quaisquer indivíduos, não sendo necessária a aprovação prévia do comitê de ética para sua realização.

COMO VOCÊS IRÃO ADMINISTRAR AS QUESTÕES DE DIREITOS AUTORAIS E DE PROPRIEDADE INTELECTUAL (IP/IPR)?

Os dados coletados para pesquisa advêm dos registros de metadados das bases de dados consultadas e orientações aos autores fornecidas pelos periódicos investigados, estando todos disponíveis para consulta em suas fontes originais.

COMO OS DADOS SERÃO ARMAZENADOS E TERÃO BACKUP DURANTE A PESQUISA?

Durante a pesquisa os dados coletados e utilizados nesta pesquisa serão armazenados no *drive* na nuvem (*Google drive*) e no computador pessoal da autora deste projeto, sendo compartilhado somente com suas orientadoras e possuindo dois *backups*, um armazenado em outro *drive* na nuvem (*One drive*) e outro em mídia externa (*pendrive*). A atualização periódica do *backup* dos dados é de responsabilidade da autora e será realizada de acordo com a atualização do versionamento (dados brutos, dados intermediários e dados finais) dos dados durante a pesquisa.

COMO VOCÊ IRÁ GERENCIAR O ACESSO E A SEGURANÇA?

Durante a realização da pesquisa, a proteção dos dados será garantida pela autora e o acesso destes será restrito apenas à mesma e suas orientadoras. Após a pesquisa, os dados serão publicados em um repositório de dados e estarão submetidos às políticas de preservação do mesmo. Os dados recuperados e utilizados nesta pesquisa estarão em acesso aberto, estando livres para ser adaptados e reutilizados em quaisquer outros trabalhos, desde que sejam corretamente citados, garantindo os direitos autorais cabíveis.

QUAIS DADOS SÃO DE VALOR A LONGO PRAZO E DEVEM SER RETIDOS, COMPARTILHADOS E/OU PRESERVADOS?

Todos os dados recuperados nesta pesquisa podem ser compartilhados e preservados pelo tempo que for necessário. Em caso de acidentes, perda de acesso, espaço, ou quaisquer inconvenientes relacionados à tecnologia, a autora deste projeto se compromete em manter uma cópia reserva dos dados em outro repositório pelo período de 5 anos após o fim da pesquisa.

QUAL É O PLANO DE PRESERVAÇÃO A LONGO PRAZO PARA OS CONJUNTO DE DADOS?

O plano de preservação a longo prazo dos dados desta pesquisa serão estabelecidos de acordo com a política de preservação do repositório selecionado para o seu armazenamento.

COMO VOCÊ COMPARTILHARÁ OS DADOS?

Os dados coletados e utilizados nesta pesquisa serão publicados imediatamente após a conclusão e defesa da Dissertação, produto desta pesquisa, com atribuição de identificador persistente para que se tornem citáveis e reusáveis. Os identificadores persistentes de todos os quatro conjuntos de dados usados nesta pesquisa serão citados em todos os documentos em que forem utilizados.

É NECESSÁRIA ALGUMA RESTRIÇÃO AO COMPARTILHAMENTO DE DADOS?

Os dados serão compartilhados sem nenhuma restrição ou período de embargo.

QUEM SERÁ RESPONSÁVEL PELO GERENCIAMENTO DOS DADOS?

A autora desta pesquisa é a principal responsável pela implementação das atividades previstas neste Plano de Gestão de Dados e a manutenção dos dados em geral.

DE QUE RECURSOS VOCÊ PRECISARÁ PARA ENTREGAR SEU PLANO?

O Plano de Gestão de Dados e os dados desta pesquisa serão publicados em repositórios adequados de forma gratuita, sem que seja necessário o pagamento de nenhuma taxa adicional.

APÊNDICE C - Revistas excluídas

Revista	Critério de exclusão	Observações
Annals of Silvicultural Research	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	O Historical Notes cita um resumo conciso de dados históricos mas não dá maiores informações e detalhes. O link "Dataset Papers" nas orientações para publicação direciona para um link com erro.
Arquivos de Neuropsiquiatria	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	Cita o tipo de publicação "historical note - data on discription", mas não fornece informação específica a respeito, nem um template.
Atmospheric and oceanic science letters	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	Cita "Data description", mas não fornece informação específica a respeito, nem template.
Genomics Data	Revista descontinuada	Descontinuada
Health Affairs	Foge ao escopo adotado nesta pesquisa	Os artigos DataWatch são documentos curtos (2000 palavras ou menos) que destacam dados que "falam por si" relativamente a questões ou tópicos políticos importantes. Devem lançar luz sobre alguma questão importante e ser "dignos de conhecimento". Normalmente, não testam hipóteses, não se baseiam em métodos estatísticos sofisticados, nem incluem longos debates políticos. O nosso objetivo é apresentar novos dados ou novas análises de dados existentes que sejam fiáveis e credíveis e que promovam a compreensão entre não especialistas em tópicos importantes e relevantes para as políticas. Incentivamos os trabalhos baseados em fontes de dados novas ou pouco utilizadas.
Journal Of Bridge Engineering	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	"Data papers" são aceitos em outros títulos da editora, mas nada específico foi visto nas instruções desta revista.
Journal Of Statistics And Data Science Education	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	Cita "Data Sets and Stories" mas não fornece nenhuma informação específica a respeito desse modelo de publicação.
Journal Of Structural Engineering	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	Cita "Data Sets and Stories", mas não fornece nenhuma informação específica a respeito desse modelo de publicação.
Nuclear Data sheets	Não fornece orientações específicas ou gerais para a submissão de data papers, embora os cite de alguma forma	Não possui orientação aos autores, os artigos são gerados a partir dos dados nucleares publicados na plataforma "National Nuclear Data Center" do Departamento de Energia dos EUA.

Fonte: Dados da Pesquisa (2023).

APÊNDICE D - Síntese comparativa dos elementos relacionados às propriedades gerais do conjunto de dados

Propriedades gerais do conjunto de dados								
	Revistas	Data de criação dos dados	Formato dos dados	Tipo de dados	Versão dos dados	Nome/título do arquivo	Tamanho do arquivo	Linguagem
Ciências Agrárias	CABI Agriculture Bioscience	N	N	S	N	N	N	N
	European Journal of Soil Science	N	N	N	N	N	N	N
	Frontiers in Veterinary Science	N	N	N	S	S	N	N
	Knowledge and Management of Aquatic Ecosystems	N	N	N	N	N	N	N
	Limnology And Oceanography Letters	S	N	S	N	N	N	N
	Marine Genomics	N	N	N	N	N	N	N
Ciências Biológicas	Biodiversity Data Journal	S	N	N	N	N	N	N
	Bioinvasions Records	N	N	N	N	N	N	N
	Bmc Genomic Data	N	S	N	N	S	N	N
	Comparative Cytogenetics	S	N	N	N	N	N	N
	Ecological Research	N	S	N	N	S	S	N
	Ecological Solutions And Evidence	N	N	N	N	N	N	N
	Ecology	S	S	N	S	S	S	N
	Ecosistemas	N	N	N	N	N	N	N
	Frontiers In Bioengineering And Biotechnology	N	N	N	S	S	N	N
	Frontiers In Cell And Developmental Biology	N	N	N	S	S	N	N
	Frontiers In Climate	N	N	N	S	S	N	N
	Frontiers In Ecology And Evolution	N	N	N	S	S	N	N
	Frontiers In Environmental Science	N	N	N	S	S	N	N
	Frontiers In Forests And Global Change	N	N	N	S	S	N	N
	Frontiers in Fungal Biology	N	N	N	S	S	N	N
	Frontiers In Genetics	N	N	N	S	S	N	N
	Frontiers In Molecular Biosciences	N	N	N	S	S	N	N
	Frontiers In Pharmacology	N	N	N	S	S	N	N
	Frontiers In Physiology	N	N	N	S	S	N	N
	Frontiers In Plant Science	N	N	N	S	S	N	N
	Frontiers In Sustainable Food Systems	N	N	N	S	S	N	N
	Functional Integrative Genomics	N	N	N	N	N	N	N
	Gigascience	N	N	S	N	N	S	N
	Global Ecology And Biogeography	S	S	N	N	N	N	N
	Human Genome Variation	N	N	N	N	N	N	N
	Journal Of Biogeography	N	N	N	N	N	N	N
	Journal Of Cheminformatics	N	N	N	N	N	N	N

	Journal Of Hymenoptera Research	S	N	N	N	N	N	N
	Journal Of Open Archaeology Data	S	S	N	S	S	N	S
	Journal Of Plant Ecology	N	N	N	N	N	S	N
	Metabarcoding And Metagenomics	S	N	N	N	N	N	N
	Metabolites	N	S	N	N	S	N	N
	Nature Conservation	S	N	N	N	N	N	N
	Neobiota	S	N	N	N	N	N	N
	Neuroinformatics	N	N	N	N	N	N	N
	Ocl Oilseeds And Fats Crops And Lipids	N	N	N	N	N	N	N
	One Ecosystem	S	N	N	N	N	N	N
	Open Quaternary	S	S	N	S	S	N	S
	Patterns	N	N	N	N	N	N	N
	Phytokeys	S	N	N	N	N	N	N
	Plant, Cell & Environment	N	N	N	N	N	N	N
	Political Research Exchange	N	N	N	N	N	N	N
	Proteomics Clinical Applications	N	N	N	N	N	N	N
	Scientific data	N	N	N	N	S	N	N
	Zookeys	S	N	N	N	N	N	N
Ciências da Saúde	Brain & Behavior	N	N	N	N	N	N	N
	Epideiology and Health	S	S	N	N	N	N	N
	Frontiers In Aging Neuroscience	N	N	N	S	S	N	N
	Frontiers In Cellular Neuroscience	N	N	N	S	S	N	N
	Frontiers In Human Neuroscience	N	N	N	S	S	N	N
	Frontiers In Medicine	N	N	N	S	S	N	N
	Frontiers In Molecular Neuroscience	N	N	N	S	S	N	N
	Frontiers In Neuroanatomy	N	N	N	S	S	N	N
	Frontiers In Neuroinformatics	N	N	N	S	S	N	N
	Frontiers In Neurorobotics	N	N	N	S	S	N	N
	Frontiers In Neuroscience	N	N	N	S	S	N	N
	Frontiers In Nutrition	N	N	N	S	S	N	N
	Frontiers In Oncology	N	N	N	S	S	N	N
	Frontiers In Psychiatry	N	N	N	S	S	N	N
	Frontiers In Public Health	N	N	N	S	S	N	N
	Frontiers in Reproductive Health	N	N	N	S	S	N	N
	Frontiers In Sports And Active Living	N	N	N	S	S	N	N
	International Journal Of Epidemiology	S	N	S	N	N	N	N
Ciências Exatas e da Terra	Advances In Atmospheric Sciences	N	S	S	N	S	S	N
	Big Earth Data	N	N	N	N	N	N	N
	Chemical Data Collections	N	N	S	S	S	N	N
	Computer Networks	N	S	N	S	N	N	N

	Earth System Science Data	N	N	N	N	N	N	N
	Earthquake Spectra	N	N	N	N	N	N	N
	Frontiers In Applied Mathematics And Statistics	N	N	N	S	S	N	N
	Frontiers In Big Data	N	N	N	S	S	N	N
	Frontiers In Chemistry	N	N	N	S	S	N	N
	Frontiers In Computer Science	N	N	N	S	S	N	N
	Frontiers In Earth Science	N	N	N	S	S	N	N
	Frontiers In Energy Research	N	N	N	S	S	N	N
	Frontiers In Marine Science	N	N	N	S	S	N	N
	Frontiers In Robotics And Ai	N	N	N	S	S	N	N
	Geochemical Journal	N	S	N	N	N	N	N
	Geoscience Data Journal	N	N	S	S	S	N	N
	Geus Bulletin	S	S	S	S	S	N	N
	Journal Of Arrhythmia	N	S	N	N	N	N	N
	Seismological Research Letters	S	N	N	N	N	N	N
Ciências Humanas	Studies In Family Planning	S	N	S	S	N	S	N
	Frontiers In Education	N	N	N	S	S	N	N
	Frontiers In Political Science	N	N	N	S	S	N	N
	Frontiers In Psychology	N	N	N	S	S	N	N
Ciências Sociais Aplicadas	Journal Of Open Humanities Data	S	S	N	N	S	N	S
	European Sociological Review	N	N	N	N	N	N	N
	Frontiers In Communication	N	N	N	S	S	N	N
Engenharias	Grey Journal	S	S	N	N	S	N	S
	Frontiers in Medical Technology	N	N	N	S	S	N	N
	Hydrological Processes	N	N	N	N	S	N	N
	Integrating Materials And Manufacturing Innovation	N	N	N	N	N	N	N
	International Journal of Naval Architecture and Ocean Engineering	N	N	N	N	N	N	N
	IUCrData	N	S	S	N	S	N	N
	Journal Of Environmental Quality	S	N	N	N	S	S	N
	Powder Diffraction	N	N	N	N	N	N	N
	Proteomics	N	N	N	N	N	N	N
	The International Journal Of Robotics Research	S	S	N	N	N	N	N
	Tribology Letters	N	N	N	N	N	N	N
Multidisciplinar	Addiction	N	N	N	N	N	N	N
	Aerosol and Air Quality Research	S	S	S	S	N	N	N
	All Life	N	N	N	N	N	N	N
	Annals Of Forest Science	S	N	S	S	S	N	S

	Arctic Science	N	N	N	N	N	N	N
	Bmc Research notes	N	S	N	N	S	N	N
	Cogent Education	S	S	N	N	N	N	N
	Data	N	S	N	N	N	N	N
	Data In Brief	N	S	N	S	S	N	N
	Data Intelligence	N	N	N	S	N	N	N
	F1000research	N	S	N	N	S	N	N

Legenda: Foram adotados os signos “S” para quando os elementos apareciam nas orientações ou *templates* para a elaboração de artigos de dados fornecidos pelas revistas e “N” para quando estes não foram encontrados nas seções de orientações aos autores, ou equivalente.

Fonte: Dados da Pesquisa (2023).

APÊNDICE E - Síntese comparativa dos elementos relacionados à contextualização da produção dos *datasets*/coleta dos dados

Contextualização da produção dos <i>datasets</i> /coleta dos dados								
	Revistas	Método da coleta	Cobertura dos dados	Produtor dos dados (Contribuição)	Análise dos dados	Dados faltantes (Limitações)	Objetivos dos dados	Breve descrição (Contexto)
Ciências Agrárias	CABI Agriculture Bioscience	S	N	N	N	N	N	N
	European Journal of Soil Science	S	N	N	S	N	N	S
	Frontiers in Veterinary Science	S	S	N	S	N	S	S
	Knowledge and Management of Aquatic Ecosystems	S	N	N	S	N	N	S
	Limnology And Oceanography Letters	N	S	S	S	N	S	S
	Marine Genomics	N	N	N	N	N	S	S
Ciências Biológicas	Biodiversity Data Journal	S	N	S	S	N	S	S
	Bioinvasions Records	S	N	N	S	N	N	S
	Bmc Genomic Data	S	N	S	S	S	S	S
	Comparative Cytogenetics	S	N	N	N	N	N	N
	Ecological Research	S	S	S	S	S	N	S
	Ecological Solutions And Evidence	S	S	N	S	N	S	S
	Ecology	S	S	S	S	S	S	S
	Ecosistemas	N	S	S	N	N	S	S
	Frontiers In Bioengineering And Biotechnology	S	S	N	S	N	S	S
	Frontiers In Cell And Developmental Biology	S	S	N	S	N	S	S
	Frontiers In Climate	S	S	N	S	N	S	S
	Frontiers In Ecology And Evolution	S	S	N	S	N	S	S
	Frontiers In Environmental Science	S	S	N	S	N	S	S
	Frontiers In Forests And Global Change	S	S	N	S	N	S	S
	Frontiers in Fungal Biology	S	S	N	S	N	S	S
	Frontiers In Genetics	S	S	N	S	N	S	S
	Frontiers In Molecular Biosciences	S	S	N	S	N	S	S
	Frontiers In Pharmacology	S	S	N	S	N	S	S
	Frontiers In Physiology	S	S	N	S	N	S	S
	Frontiers In Plant Science	S	S	N	S	N	S	S
	Frontiers In Sustainable Food Systems	S	S	N	S	N	S	S
	Functional Integrative Genomics	N	N	N	N	N	N	N
	Gigascience	S	N	S	S	N	N	S
	Global Ecology And Biogeography	S	S	N	S	N	S	S

	Human Genome Variation	N	N	N	N	N	N	S
	Journal Of Biogeography	S	S	S	S	N	S	S
	Journal Of Cheminformatics	N	N	N	S	S	S	S
	Journal Of Hymenoptera Research	S	N	S	S	N	S	S
	Journal Of Open Archaeology Data	S	S	S	S	S	S	S
	Journal Of Plant Ecology	N	N	N	N	N	N	S
	Metabarcoding And Metagenomics	S	N	S	S	N	S	S
	Metabolites	S	N	S	S	N	S	S
	Nature Conservation	S	N	S	S	N	S	S
	Neobiota	S	N	S	S	N	S	S
	Neuroinformatics	N	N	N	S	N	N	S
	Ocl Oilseeds And Fats Crops And Lipids	N	N	N	N	N	N	N
	One Ecosystem	S	N	S	S	N	S	S
	Open Quaternary	S	S	S	S	S	S	S
	Patterns	S	N	N	N	N	S	S
	Phytokeys	S	N	S	S	N	S	S
	Plant, Cell & Environment	S	N	N	S	N	S	S
	Political Research Exchange	N	N	S	N	N	N	S
	Proteomics Clinical Applications	N	N	N	N	N	N	N
	Scientific data	S	N	S	S	S	S	S
	Zookeys	S	N	S	S	N	S	S
Ciências da Saúde	Brain & Behavior	N	N	N	N	N	N	N
	Epidemiology and Health	S	S	S	S	S	N	S
	Frontiers In Aging Neuroscience	S	S	N	S	N	S	S
	Frontiers In Cellular Neuroscience	S	S	N	S	N	S	S
	Frontiers In Human Neuroscience	S	S	N	S	N	S	S
	Frontiers In Medicine	S	S	N	S	N	S	S
	Frontiers In Molecular Neuroscience	S	S	N	S	N	S	S
	Frontiers In Neuroanatomy	S	S	N	S	N	S	S
	Frontiers In Neuroinformatics	S	S	N	S	N	S	S
	Frontiers In Neurorobotics	S	S	N	S	N	S	S
	Frontiers In Neuroscience	S	S	N	S	N	S	S
	Frontiers In Nutrition	S	S	N	S	N	S	S
	Frontiers In Oncology	S	S	N	S	N	S	S
	Frontiers In Psychiatry	S	S	N	S	N	S	S
	Frontiers In Public Health	S	S	N	S	N	S	S
	Frontiers in Reproductive Health	S	S	N	S	N	S	S
	Frontiers In Sports And Active Living	S	S	N	S	N	S	S
	International Journal Of Epidemiology	S	S	N	S	S	S	S
	Advances In Atmospheric Sciences	S	N	S	S	N	S	S

Ciências Exatas e da Terra	Big Earth Data	S	N	N	S	N	S	S
	Chemical Data Collections	S	N	S	S	N	N	S
	Computer Networks	S	N	S	S	N	N	N
	Earth System Science Data	N	N	N	N	N	N	N
	Earthquake Spectra	S	N	N	S	N	N	S
	Frontiers In Applied Mathematics And Statistics	S	S	N	S	N	S	S
	Frontiers In Big Data	S	S	N	S	N	S	S
	Frontiers In Chemistry	S	S	N	S	N	S	S
	Frontiers In Computer Science	S	S	N	S	N	S	S
	Frontiers In Earth Science	S	S	N	S	N	S	S
	Frontiers In Energy Research	S	S	N	S	N	S	S
	Frontiers In Marine Science	S	S	N	S	N	S	S
	Frontiers In Robotics And Ai	S	S	N	S	N	S	S
	Geochemical Journal	S	N	S	S	N	N	S
	Geoscience Data Journal	S	N	S	S	S	S	S
	Geus Bulletin	S	S	S	S	N	N	S
	Journal Of Arrhythmia	S	N	N	S	N	S	S
	Seismological Research Letters	S	S	S	S	S	S	S
	Studies In Family Planning	S	S	S	S	S	S	S
Ciências Humanas	Frontiers In Education	S	S	N	S	N	S	S
	Frontiers In Political Science	S	S	N	S	N	S	S
	Frontiers In Psychology	S	S	N	S	N	S	S
	Journal Of Open Humanities Data	S	S	S	S	S	N	S
Ciências Sociais Aplicadas	European Sociological Review	N	N	N	N	N	N	N
	Frontiers In Communication	S	S	N	S	N	S	S
	Grey Journal	S	S	S	S	S	N	S
Engenharias	Frontiers in Medical Technology	S	S	N	S	N	S	S
	Hydrological Processes	S	N	S	S	N	S	S
	Integrating Materials And Manufacturing Innovation	S	N	N	S	N	N	S
	International Journal of Naval Architecture and Ocean Engineering	N	N	N	N	N	N	N
	IUCrData	N	N	S	N	N	N	S
	Journal Of Environmental Quality	S	N	N	S	S	N	S
	Powder Diffraction	S	N	N	S	N	N	S
	Proteomics	N	N	N	N	N	N	N
	The International Journal Of Robotics Research	S	N	S	S	S	N	S
	Tribology Letters	S	N	N	S	S	S	S
	Addiction	S	N	N	S	S	S	S

Multidisciplinar	Aerosol and Air Quality Research	N	N	N	S	N	N	N
	All Life	S	N	N	S	S	S	S
	Annals Of Forest Science	S	S	S	S	S	S	S
	Arctic Science	S	N	N	S	N	S	S
	Bmc Research notes	S	N	S	S	S	S	S
	Cogent Education	S	N	N	N	N	N	N
	Data	S	N	S	S	N	S	S
	Data In Brief	S	N	S	S	S	N	N
	Data Intelligence	S	N	N	N	N	N	N
	F1000research	S	S	S	S	S	S	S

Legenda: Foram adotados os signos "S" para quando os elementos apareciam nas orientações ou *templates* para a elaboração de artigos de dados fornecidos pelas revistas e "N" para quando estes não foram encontrados nas seções de orientações aos autores, ou equivalente.

Fonte: Dados da Pesquisa (2023).

APÊNDICE F - Síntese comparativa dos elementos relacionados às informações sobre o compartilhamento e reuso dos dados

Informações sobre o compartilhamento e reuso dos dados							
	Revistas	Proveniência dos dados	Reputação e histórico do repositório	Preservação dos dados	Reuso Anterior	Reuso Potencial	Termos e condições de uso
Ciências Agrárias	CABI Agriculture Bioscience	S	N	N	N	N	N
	European Journal of Soil Science	N	N	N	N	S	S
	Frontiers in Veterinary Science	N	N	N	N	N	S
	Knowledge and Management of Aquatic Ecosystems	N	N	N	N	N	N
	Limnology And Oceanography Letters	S	N	N	S	S	N
	Marine Genomics	N	N	N	N	N	N
Ciências Biológicas	Biodiversity Data Journal	N	S	N	N	N	S
	Bioinvasions Records	N	N	N	N	N	N
	Bmc Genomic Data	S	S	N	N	N	N
	Comparative Cytogenetics	N	N	N	N	N	N
	Ecological Research	S	S	N	N	N	S
	Ecological Solutions And Evidence	S	N	N	N	N	N
	Ecology	S	N	N	S	S	S
	Ecosistemas	N	S	N	N	S	N
	Frontiers In Bioengineering And Biotechnology	N	N	N	N	N	S
	Frontiers In Cell And Developmental Biology	N	N	N	N	N	S
	Frontiers In Climate	N	N	N	N	N	S
	Frontiers In Ecology And Evolution	N	N	N	N	N	S
	Frontiers In Environmental Science	N	N	N	N	N	S
	Frontiers In Forests And Global Change	N	N	N	N	N	S
	Frontiers In Fungal Biology	N	N	N	N	N	S
	Frontiers In Genetics	N	N	N	N	N	S
	Frontiers In Molecular Biosciences	N	N	N	N	N	S
	Frontiers In Pharmacology	N	N	N	N	N	S
	Frontiers In Physiology	N	S	N	N	N	S
	Frontiers In Plant Science	N	N	N	N	N	S
	Frontiers In Sustainable Food Systems	N	N	N	N	N	S
	Functional Integrative Genomics	N	N	N	N	N	N
	Gigascience	N	S	N	N	S	N
	Global Ecology And Biogeography	N	N	N	N	N	S
	Human Genome Variation	N	N	N	N	N	N
	Journal Of Biogeography	N	N	N	N	S	N
	Journal Of Cheminformatics	N	S	S	N	S	S

	Journal Of Hymenoptera Research	N	S	N	N	N	S
	Journal Of Open Archaeology Data	S	S	N	N	S	S
	Journal Of Plant Ecology	N	S	N	N	N	N
	Metabarcoding And Metagenomics	N	S	N	N	N	S
	Metabolites	N	N	N	N	S	S
	Nature Conservation	N	S	N	N	N	S
	Neobiota	N	S	N	N	N	S
	Neuroinformatics	N	N	N	N	N	N
	Ocl Oilseeds And Fats Crops And Lipids	N	N	N	N	N	N
	One Ecosystem	N	S	N	N	N	S
	Open Quaternary	S	S	N	S	S	S
	Patterns	N	N	N	N	S	N
	Phytokeys	N	S	N	N	N	S
	Plant, Cell & Environment	N	N	N	N	N	N
	Political Research Exchange	N	S	N	N	N	N
	Proteomics Clinical Applications	N	S	N	N	N	S
	Scientific data	S	S	N	N	S	S
	Zookeys	N	S	N	N	N	S
Ciências da Saúde	Brain & Behavior	N	N	N	N	N	N
	Epideiology and Health	S	N	N	N	S	N
	Frontiers In Aging Neuroscience	N	N	N	N	N	S
	Frontiers In Cellular Neuroscience	N	N	N	N	N	S
	Frontiers In Human Neuroscience	N	N	N	N	N	S
	Frontiers In Medicine	N	N	N	N	N	S
	Frontiers In Molecular Neuroscience	N	N	N	N	N	S
	Frontiers In Neuroanatomy	N	N	N	N	N	S
	Frontiers In Neuroinformatics	N	N	N	N	N	S
	Frontiers In Neurorobotics	N	N	N	N	N	S
	Frontiers In Neuroscience	N	N	N	N	N	S
	Frontiers In Nutrition	N	N	N	N	N	S
	Frontiers In Oncology	N	N	N	N	N	S
	Frontiers In Psychiatry	N	N	N	N	N	S
	Frontiers In Public Health	N	N	N	N	N	S
	Frontiers in Reproductive Health	N	N	N	N	N	S
	Frontiers In Sports And Active Living	N	N	N	N	N	S
	International Journal Of Epidemiology	N	N	S	S	S	S
	Advances In Atmospheric Sciences	N	S	N	N	S	S
Ciências Exatas e da Terra	Big Earth Data	S	S	S	N	S	N
	Chemical Data Collections	S	S	N	N	S	N
	Computer Networks	S	S	N	N	S	N

	Earth System Science Data	N	S	N	N	N	S
	Earthquake Spectra	N	S	N	N	S	N
	Frontiers In Applied Mathematics And Statistics	N	N	N	N	N	S
	Frontiers In Big Data	N	N	N	N	N	S
	Frontiers In Chemistry	N	N	N	N	N	S
	Frontiers In Computer Science	N	N	N	N	N	S
	Frontiers In Earth Science	N	N	N	N	N	S
	Frontiers In Energy Research	N	N	N	N	N	S
	Frontiers In Marine Science	N	N	N	N	N	S
	Frontiers In Robotics And Ai	N	N	N	N	N	S
	Geochemical Journal	N	N	N	N	N	N
	Geoscience Data Journal	S	N	N	N	S	S
	Geus Bulletin	S	S	N	N	S	N
	Journal Of Arrhythmia	N	N	N	N	N	N
	Seismological Research Letters	S	N	N	N	N	N
	Studies In Family Planning	N	N	N	N	N	N
Ciências Humanas	Frontiers In Education	N	N	N	N	N	S
	Frontiers In Political Science	N	N	N	N	N	S
	Frontiers In Psychology	N	N	N	N	N	S
	Journal Of Open Humanities Data	N	S	N	N	S	S
Ciências Aplicadas	Sociais	European Sociological Review	N	N	N	N	N
		Frontiers In Communication	N	N	N	N	S
		Grey Journal	N	S	N	N	S
Engenharias	Frontiers in Medical Technology	N	N	N	N	N	S
	Hydrological Processes	S	N	N	N	S	S
	Integrating Materials And Manufacturing Innovation	S	N	N	N	S	N
	International Journal of Naval Architecture and Ocean Engineering	N	N	N	N	N	N
	IUCrData	N	N	N	N	N	N
	Journal Of Environmental Quality	S	S	N	N	S	N
	Powder Diffraction	N	N	N	N	N	N
	Proteomics	N	S	N	N	N	S
	The International Journal Of Robotics Research	N	N	S	N	S	S
	Tribology Letters	N	S	N	N	S	N
Multidisciplinar	Addiction	S	N	N	N	N	N
	Aerosol and Air Quality Research	N	S	S	N	N	S
	All Life	S	N	N	N	N	N
	Annals Of Forest Science	S	N	N	N	S	S
	Arctic Science	N	N	N	N	S	S
	Bmc Research notes	S	S	N	N	N	N

	Cogent Education	N	N	N	N	N	N
	Data	S	N	N	N	S	S
	Data In Brief	S	N	N	N	S	N
	Data Intelligence	N	N	N	N	N	N
	F1000research	S	S	N	N	N	S

Legenda: Foram adotados os signos "S" para quando os elementos apareciam nas orientações ou *templates* para a elaboração de artigos de dados fornecidos pelas revistas e "N" para quando estes não foram encontrados nas seções de orientações aos autores, ou equivalente.

Fonte: Dados da Pesquisa (2023).