



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Crislaine Zurilda Silveira

Publicação de dados de pesquisa: uma análise a partir dos *data papers* de
autores filiados às instituições brasileiras

FLORIANÓPOLIS

2026

Crislaine Zurilda Silveira

Publicação de dados de pesquisa: uma análise a partir dos *data papers* de autores filiados às instituições brasileiras

Tese submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Doutora em Ciência da Informação.

Linha de pesquisa: Informação, Comunicação Científica e Competência

Orientador: Prof. Thiago Magela Rodrigues Dias, Dr.

Coorientadora: Prof. Patricia da Silva Neubert, Dra

FLORIANÓPOLIS

2026

Silveira, Crislaine Zurilda

Publicação de dados de pesquisa : uma análise a partir dos data papers de autores filiados às instituições brasileiras / Crislaine Zurilda Silveira ; orientador, Thiago Magela Rodrigues Dias, coorientador, Patricia da Silva Neubert, 2026.

280 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2026.

Inclui referências.

1. Ciência da Informação. 2. Data papers. 3. Data journals. 4. Repositório de dados de pesquisa. I. Dias, Thiago Magela Rodrigues. II. Neubert, Patricia da Silva . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Informação. IV. Título.

Crislaine Zurilda Silveira

Publicação de dados de pesquisa: uma análise a partir dos *data papers* de autores filiados às instituições brasileiras

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 27 de fevereiro de 2026, pela banca examinadora composta pelos seguintes membros:

Prof.(a) Rosângela Schwarz Rodrigues, Dr.(a)
Universidade Federal de Santa Catarina

Prof.(a) Edgar Bisset Alvarez, Dr.(a)
Universidade Federal de Santa Catarina

Prof.(a) Elizabete Cristina de Souza de Aguiar Monteiro, Dr.(a)
Universidade Estadual Paulista

Prof.(a) Washington Luís Ribeiro de Carvalho Segundo, Dr.(a)
Instituto Brasileiro de Informação em Ciência e Tecnologia

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutora em Ciência da Informação.



Documento assinado digitalmente

Luciane Paula Vital

Data: 29/03/2026 16:37:47-0300

CPF: ***.658.789-**

Verifique as assinaturas em <https://v.ufsc.br>

Coordenação do Programa de Pós-Graduação



Documento assinado digitalmente

THIAGO MAGELA RODRIGUES DIAS

Data: 28/03/2026 19:38:35-0300

CPF: ***.987.396-**

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Thiago Magela Rodrigues Dias, Dr.
Orientador

Florianópolis, 2026

Dedico esta tese à educação pública, pois só ela é capaz de nos levar a lugares e a conquistas inimagináveis.

AGRADECIMENTOS

Iniciando os agradecimentos, tentando evitar os clichês, mas sendo quase impossível, quero agradecer primeiramente a Deus. Não que eu seja uma pessoa muito religiosa, mas por saber que temos uma força maior que vai além de qualquer religião e que de alguma maneira guia nossas decisões e protege nossos caminhos.

Agradeço à Universidade Federal de Santa Catarina (UFSC) como discente por ter tido a oportunidade de cursar uma pós-graduação de qualidade, pública e gratuita, com recursos que me permitiram ingressar e cursar o doutorado, com professores que possibilitaram expandir meus conhecimentos a partir das trocas em sala de aula, laboratórios, corredores, em especial a professora Heloísa Costa, por todo apoio e aprendizado ao longo do estágio de docência. Agradeço o apoio administrativo realizado pelos colegas técnicos administrativos. Também agradeço como servidora técnico-administrativa da mesma instituição pela oportunidade de gozar do afastamento para me dedicar integralmente ao doutorado. Agradeço aos meus amigos e colegas da Biblioteca Central, os quais não irei nomear sob o risco de esquecer alguém importante, por todo o apoio fornecido durante esse processo.

Quero agradecer, em especial, aos meus orientadores, professora Patrícia Neubert e o professor Thiago Magela, pelas revisões críticas do texto, pelas discussões, pelos incentivos para submeter trabalhos aos eventos e periódicos. Quero agradecer, sobretudo, pelo que muitas vezes fica nas entrelinhas das orientações, por acreditarem e confiarem no meu trabalho, por ouvirem minhas inseguranças, por serem exemplos, para acreditar que a “ciência presta” e que a educação e a pesquisa podem nos levar a lugares e conquistas inimagináveis. Ainda nas contribuições para a pesquisa, quero agradecer ao bibliotecário Fábio Lorensi do Canto, que teve uma contribuição fundamental na coleta dos dados.

Ainda nesse ínterim, quero agradecer meus amigos e colegas do grupo de pesquisa Informação Científica: métricas e publicações (INFOCIENT) pelas trocas, pelos aprendizados, por acolherem a “irmã mais velha”, nem sempre presente fisicamente, mas sempre disposta a ajudar. Meu muito obrigada, também, aos colegas do Programa de Pós-Graduação em Ciência da Informação que foram se juntando nessa caminhada e que, com as experiências compartilhadas foram tornando os caminhos menos árduos.

Quero agradecer aos membros da banca, professora Rosângela Schwarz Rodrigues, professor Edgar Bisset Alvarez, Elizabete Cristina de Souza de Aguiar Monteiro e Washington Luís Ribeiro de Carvalho Segundo, por dedicarem seu tempo, em geral, escasso e disputado, para realizar a leitura da tese e trazerem oportunidades de melhorias.

Quero agradecer meus amigos, parceiros ativos nessa jornada, com conselhos, trocas de experiências e ouvidos atentos, Ariana Corrêa, Elaine Monteiro, Luciana Bergamo, Mateus Rebouças, Sirlene Pinto, meu muito obrigada pelos “cafés com bobagens”.

Por último, não por ser menos importante, mas por ser a base que sustentou todo esse processo, agradeço aos meus pais, Zenir e Carlos, e ao meu noivo Raphael por terem me amparado ao longo desse tempo e por entenderem as minhas ausências. Amo muito vocês!

RESUMO

A Ciência Aberta vem demandando esforços voltados ao compartilhamento dos dados de pesquisa. No entanto, esse termo vem dando espaço para o termo publicação de dados de pesquisa que, associado ao reconhecimento ao esforço dos pesquisadores em disponibilizar seus dados e às práticas já consolidadas na comunicação científica, como a citação e revisão por pares, vem possibilitando a criação de novos canais voltados para a sua publicação como os *data papers* e a consolidação dos *data journals*. As iniciativas nacionais voltadas à consolidação da Ciência Aberta no Brasil e a publicação de dados de pesquisa ainda são incipientes. Considerando o contexto brasileiro e os benefícios inerentes aos *data papers*, essa tese tem como objetivo caracterizar a publicação de dados de pesquisa por meio de *data papers* por parte dos autores filiados às instituições brasileiras. Esse objetivo se materializa a partir da análise das características das instituições às quais os autores estão vinculados, agências de financiamentos, periódicos onde os *data papers* foram publicados e repositórios onde os dados associados foram depositados. É discutido como os *data papers*, os *data journals* e os repositórios formam o ecossistema da publicação dos dados de pesquisa. O universo da pesquisa compreendeu a identificação de 634 *data papers* indexados nas bases de dados Scopus e Web of Science, nos quais pelo menos um autor deveria estar filiado a uma instituição brasileira. Com o uso do OpenAlex foi possível obter um relatório com os metadados que foram analisados a partir da abordagem mista. Dentre as características temporais identificadas, destaca-se a jovialidade das publicações brasileiras, sendo os primeiros *data papers* publicados em 2014. Os temas de interesse global, como mudanças climáticas, conservação, genética e ecologia, sequenciamento de RNA e os temas que envolvem doenças arbovirais como Chagas e Leishmaniose, têm sido uma tendência nas publicações dos autores filiados às instituições brasileiras. A filiação dos autores se concentrou nas universidades, nacionais e internacionais, localizadas em países com maior capacidade científica, tecnológica e com incentivos voltados à Ciência Aberta, predominantemente localizados no Norte Global. Os financiamentos se caracterizam por aqueles advindos de instituições nacionais e estrangeiras, majoritariamente de fundo público. A publicação dos *data papers* se concentrou em *data journals*, mantidos por *publishers* comerciais e que cobram taxas de publicação. Nas características relacionadas ao depósito dos dados o destaque foi para o depósito em repositórios mantidos por *publishers* comerciais, localizados em países do Norte Global. A pesquisa chama a atenção para o cuidado com a soberania dos dados coletados e financiados por agências nacionais. A tese apresenta recomendações para estudos futuros e destaca o papel dos *data papers* como fontes de informação voltados à análise do comportamento das áreas quanto à publicação dos dados de pesquisa, o que pode contribuir para o desenvolvimento da Ciência Aberta.

Palavras-chave: *data papers*; *data journals*; repositório de dados de pesquisa; financiamento de pesquisa

ABSTRACT

Open Science has been demanding efforts focused on sharing research data. However, this term is giving way to the term "research data publication," which, associated with the recognition of researchers' efforts in making their data available and with already consolidated practices in scientific communication, such as citation and peer review, is enabling the creation of new channels for publication, such as data papers and the consolidation of data journals. National initiatives aimed at consolidating Open Science in Brazil and publishing research data are still incipient. Considering the Brazilian context and the inherent benefits of data papers, this thesis aims to characterize the publication of research data through data papers by authors affiliated with Brazilian institutions. This objective is achieved through the analysis of the characteristics of the institutions to which the authors are linked, funding agencies, journals where the data papers were published, and repositories where the associated data were deposited. The study discusses how data papers, data journals, and repositories form the ecosystem of research data publication. The research universe comprised the identification of 634 data papers indexed in the Scopus and Web of Science databases, in which at least one author had to be affiliated with a Brazilian institution. Using OpenAlex, it was possible to obtain a report with the metadata that was analyzed using a mixed-methods approach. Among the temporal characteristics identified, the relative youth of Brazilian publications stands out, with the earliest data papers published in 2014. Topics of global interest, such as climate change, conservation, genetics and ecology, RNA sequencing, and topics involving arboviral diseases like Chagas disease and Leishmaniasis, have been a trend in the publications of authors affiliated with Brazilian institutions. Institutional affiliation was concentrated in national and international universities located in countries with greater scientific and technological capacity and incentives geared towards Open Science, predominantly located in the Global North. Funding is characterized by sources from national and international institutions, mostly from public funds. The publication of data papers was concentrated in data journals maintained by commercial publishers that charge publication fees. Regarding data deposit characteristics, the focus was on deposits in repositories maintained by commercial publishers located in countries of the Global North. The research draws attention to the need for safeguarding the sovereignty of data collected and funded by national agencies. The thesis presents recommendations for future studies and highlights the role of data papers as information sources for analyzing the behavior of research areas regarding the publication of research data, which can contribute to the development of Open Science.

Keywords: data papers; data journals; research data repository; research funding

LISTA DE FIGURAS

Figura 1- Síntese das iniciativas voltadas à institucionalização	30
Figura 2 - Extrato taxonomia da Ciência Aberta de Silveira <i>et al.</i> (2023)	39
Figura 3 - Representação do ecossistema da publicação de dados de pesquisa	51
Figura 4 - Processo de publicação de um <i>data paper</i> num <i>data journal</i>	59
Figura 5 - Fluxo de identificação de estudos nas bases LISTA, Web of Science e Scopus	93
Figura 6 - Página da busca realizada na Scopus	96
Figura 7 - Página da busca realizada na Web of Science	96
Figura 8 - Página da busca por tipo de documento e filtro “Brazil” na Scopus	97
Figura 9 - Página da busca por tipo de documento e filtro “Brazil” na Web of Science	97
Figura 10 - Síntese da coleta de dados	98
Figura 11 - Estrutura de classificação de tópicos no OpenAlex	104
Figura 12 - Distribuição das instituições e sua tipologia predominante em cada país	125
Figura 13 - Distribuição das agências pelo mundo e sua tipologia predominante..	141
Figura 14 - Formas de publicação dos dados de pesquisa associados aos <i>data papers</i>	187

LISTA DE GRÁFICOS

Gráfico 1 - Publicação de <i>data papers</i> e as grandes áreas	115
Gráfico 2 - Variação no crescimento dos <i>data papers</i> nas grandes áreas do conhecimento.....	116
Gráfico 3 - Grandes áreas, áreas e os temas dos <i>data papers</i>	118
Gráfico 4 – Acesso aos <i>data papers</i> por meio dos periódicos	157
Gráfico 5 - Relação entre o Fator de Impacto (2023) e a quantidade de <i>data papers</i>	163
Gráfico 6 - Relação entre o SJR (2023) e a quantidade de <i>data papers</i>	163
Gráfico 7 - Relação entre o Fator de Impacto (2023) e a tipologia do periódico	166
Gráfico 8 - Relação entre o SJR (2023) e a tipologia do periódico	166
Gráfico 9 - Relação entre o Fator de Impacto (2023) e a APC	168
Gráfico 10 - Relação entre o SJR (2023) e a APC	168
Gráfico 11 – A disponibilidade dos dados de pesquisa nos periódicos	185
Gráfico 12 - Áreas cobertas pelos repositórios	204

LISTA DE QUADROS

Quadro 1 - Marcos mensuráveis dos planos brasileiros de ação do governo aberto	24
Quadro 2 - Definições de <i>data journal</i>	54
Quadro 3 - Comparação entre a revisão por pares em periódicos tradicionais e <i>data journals</i>	61
Quadro 4 - Definição de <i>data paper</i>	63
Quadro 5 - Benefícios associados a publicação de dados de pesquisa	65
Quadro 6 - Elementos comuns dos <i>data papers</i>	67
Quadro 7 - Variáveis relacionadas aos objetivos específicos	86
Quadro 8 - Estratégias de busca	91
Quadro 9 - Estratégias de buscas em bases adicionais.	92
Quadro 10 - Fontes de coleta de cada variável	99
Quadro 11 - Critérios de avaliação dos <i>data papers</i> e <i>datasets</i> informados pelos periódicos	109
Quadro 12 - Estratificação dos periódicos no Qualis (2017-2020)	171
Quadro 13 - Tipo de revisão por pares e aspectos revisados	175
Quadro 14 - Critérios identificados para avaliação dos <i>data papers</i> e <i>datasets</i>	179
Quadro 15 - Características dos repositórios	190
Quadro 16 - Lista de documentos consultados nas páginas dos periódicos e nos <i>data papers</i>	247
Quadro 17 - Repositórios e documentos sobre o depósito, cobrança e revisão dos dados	255
Quadro 18 - Grandes áreas, áreas e temas dos <i>data papers</i>	260
Quadro 19 - Indicadores e avaliações dos periódicos	272
Quadro 20 - Mantenedores dos repositórios	277
Quadro 21 - Áreas e subáreas da ASJC	280

LISTAS DE TABELAS

Tabela 1 - Assuntos com maior frequência nos <i>data papers</i>	120
Tabela 2 - Distribuição das instituições com maior número de autores filiados	129
Tabela 3 - Distribuição geográfica das instituições brasileiras e suas tipologias ...	131
Tabela 4 - Caracterização da tipologia e nacionalidade das agências de financiamentos dos <i>data papers</i>	136
Tabela 5 - Caracterização dos financiamentos a partir das grandes áreas e áreas do conhecimento	138
Tabela 6 - Distribuição das agências de financiamentos mais mencionadas.....	143
Tabela 7 - Características dos periódicos.....	151
Tabela 8 - Estimativa de gastos com APC por áreas, conforme esquema de classificação da All Science Journal Classification (ASJC) (USD)	160
Tabela 9 - Quantidade de repositórios mencionados nos <i>data papers</i>	199
Tabela 10 - Tipologia dos repositórios.....	200
Tabela 11 - Orientações dos periódicos e a adesão dos autores	208
Tabela 12 - Lista de países e tipologia das instituições de filiação	265
Tabela 13 - Lista de países e tipologia das instituições de financiamento	270

LISTA DE ABREVIATURAS E SIGLAS

ABEC	Associação Brasileira de Editores Científicos
ANCIB	Associação de Pesquisa e Pós-graduação em Ciência da Informação
APC	Article Publishing Charges
API	Application Programming Interface
BASE	Bielefeld Academic Search Engine
BRAPCI	Base de Dados em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CGU	Controladoria Geral da União
CNEN	Comissão Nacional de Energia Nuclear
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CoNCiênciaA	Consórcio Nacional de Ciência Aberta
CONFAP	Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa
CONFOA	Conferência Lusófona de Ciência Aberta
CRedit	Contributor Roles Taxonomy
DAS	Declaração de disponibilidade de dados
DFG	Deutsche Forschungsgemeinschaft
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
DORA	Declaração de São Francisco
DPCs	data publishing charges
DUAs	Acordos de Uso de Dados
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ENA	European Nucleotide Archive
ENANCIB	Encontro Nacional de Pesquisa em Ciência da Informação
ENCTI	Estratégia Nacional De Ciência e Tecnologia
ERC	Conselho Europeu de Investigação
FAIR	Findable, Accessible, Interoperable, Reusable
FAPEG	Fundação de Amparo à Pesquisa do Estado de Goiás
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FGV	Fundação Getúlio Vargas
FIOCRUZ	Fundação Oswaldo Cruz

G7	Grupo dos Sete
GBIF	Global Biodiversity Information Facility
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ITA	Instituto Tecnológico de Aeronáutica
JIF	Journal Impact Factor
JPN	Japan Society for the Promotion of Science
MCTI	Ministério da Ciência, Tecnologia e Inovação
NCBI	National Center for Biotechnology Information
OCABr	Observatório de Ciência Aberta do Brasil
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ODC	Open Data Commons
OKBR	Open Knowledge Brasil
ONG	Organização Não Governamental
ORCID	Open Researcher and Contributor ID
OSTP	Escritório de Ciência e Tecnologia da Casa Branca
PGCIN	Programa de Pós-Graduação em Ciência da Informação
RBR	Rede Brasileira de Reprodutibilidade
REDAPE	Repositório de Dados de Pesquisa da Embrapa
RNP	Rede Brasileira para Educação e Pesquisa
ROR	Research Organization Registry
SBPC	Sociedade Brasileira para o Progresso da Ciência
SCI	Science Citation Index
SciELO	Scientific Electronic Library Online
SJR	SCImago Journal Rank
TOP	Transparency and Openness Promotion
UFABC	Universidade Federal do ABC
UFAL	Universidade Federal de Alagoas
UFC	Universidade Federal do Ceará
UFG	Universidade Federal de Goiás
UFRR	Universidade Federal de Roraima
UFSC	Universidade Federal de Santa Catarina
UFSCAR	Universidade Federal de São Carlos
UNB	Universidade de Brasília

UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UNESP	Universidade Estadual Paulista
UNICAMP	Universidade Estadual de Campinas
UNIFESP	Universidade Federal de São Paulo
UNIRIO	Universidade Federal do Estado do Rio de Janeiro
USP	Universidade de São Paulo
VPN	Virtual Private Network
WoS	Web of Science

SUMÁRIO

1	INTRODUÇÃO	20
1.1	DELIMITAÇÃO DO PROBLEMA DE PESQUISA	22
1.2	OBJETIVO GERAL E ESPECÍFICOS.....	32
1.3	JUSTIFICATIVA	33
1.4	ABORDAGEM EPISTEMOLÓGICA.....	37
2	A COMUNICAÇÃO CIENTÍFICA E A CIÊNCIA ABERTA.....	40
2.1	PUBLICAÇÃO DE DADOS DE PESQUISA	47
2.2	<i>DATA JOURNALS</i> : PERIÓDICOS ESPECIALIZADOS EM PUBLICAR DADOS DE PESQUISA	53
2.3	<i>DATA PAPERS</i> : ANALISANDO SEU CONCEITO E FUNÇÕES	62
2.4	REPOSITÓRIOS DE DADOS DE PESQUISA: PARTE INTEGRANTE DO PROCESSO DE PUBLICAÇÃO DOS DADOS	70
2.5	A REPRESENTAÇÃO DOS <i>DATA PAPERS</i> E <i>DATA JOURNALS</i> NA LITERATURA.....	74
3	PROCEDIMENTOS METODOLÓGICOS	85
3.1	CLASSIFICAÇÃO DA PESQUISA	85
3.1.1	Pesquisa bibliográfica a partir do uso da busca sistematizada da literatura	90
3.2	UNIVERSO DA PESQUISA.....	94
3.3	COLETA DOS DADOS	95
3.4	TRATAMENTO E ANÁLISE DOS DADOS	102
3.4.1	Dados sobre as características temporais e temáticas dos <i>data papers</i>	103
3.4.2	Características das instituições de filiação dos autores e das agências de financiamento	104
3.4.3	Características dos periódicos.....	107
3.4.4	Características dos repositórios	110
3.5	LIMITAÇÕES DA PESQUISA.....	112
4	RESULTADOS E DISCUSSÕES	114
4.1	CARACTERÍSTICAS TEMPORAIS E TEMÁTICAS DOS <i>DATA PAPERS</i>	114

4.1.1	Considerações da seção	122
4.2	CARACTERÍSTICAS DA FILIAÇÃO DOS AUTORES DOS DATA PAPERS E DOS FINANCIAMENTOS.....	123
4.2.1	Características das instituições brasileiras	130
4.2.2	Características dos financiamentos	134
4.2.3	Considerações da seção	147
4.3	CARACTERÍSTICAS DOS PERIÓDICOS	150
4.3.1	Avaliação dos periódicos a partir do Journal Impact Factor (JIF) e do SCImago Journal Rank (SJR) de 2023.	162
4.3.1.1	<i>Avaliação dos periódicos a partir do Qualis Periódicos (2017-2020)</i>	170
4.3.2	O processo de revisão por pares	174
4.3.3	Considerações da seção	181
4.4	A DISPONIBILIDADE DOS DADOS DE PESQUISA E AS CARACTERÍSTICAS DOS REPOSITÓRIOS DE DADOS	183
4.4.1	Características dos repositórios utilizados pelos autores filiados às instituições brasileiras	188
4.4.2	Cobertura temática e tipologia dos repositórios	199
4.4.3	Depósito de dados versus orientações dos periódicos	206
4.4.4	Considerações da seção	213
5	CONSIDERAÇÕES FINAIS	215
5.1	CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS TEMPORAIS E TEMÁTICAS DOS <i>DATA PAPERS</i>	217
5.2	CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DAS FILIAÇÕES DOS AUTORES E DOS FINANCIADORES	218
5.3	CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DOS PERIÓDICOS	218
5.4	CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DOS REPOSITÓRIOS DE DADOS.....	219
5.5	RECOMENDAÇÕES E ESTUDOS FUTUROS.....	220
	REFERÊNCIAS	223
	APÊNDICE A – LISTA DE DOCUMENTOS CONSULTADOS NAS PÁGINAS DOS PERIÓDICOS E NOS <i>DATA PAPERS</i>	247
	APÊNDICE B – LISTA DOS DOCUMENTOS CONSULTADOS NAS PÁGINAS DOS REPOSITÓRIOS	255

APÊNDICE C – LISTA DE GRANDES ÁREAS, ÁREAS E TEMAS DOS <i>DATA PAPERS</i>	260
APÊNDICE D – LISTA DE PAÍSES E TIPOLOGIAS DE INSTITUIÇÕES DE FILIAÇÃO	265
APÊNDICE E – LISTA DE PAÍSES E TIPOLOGIAS DE INSTITUIÇÕES DE FINANCIAMENTO	270
APÊNDICE F – INDICADORES DOS PERIÓDICOS	272
APÊNDICE G – LISTA DOS MANTENEDORES DOS REPOSITÓRIOS IDENTIFICADOS	277
ANEXO A – ESQUEMA DE CLASSIFICAÇÃO DA <i>ALL SCIENCE</i> <i>JOURNAL CLASSIFICATION (ASJC)</i>	280

1 INTRODUÇÃO

A Ciência Aberta tem o movimento pelo Acesso Aberto como seu antecessor, cuja consolidação se deu com as declarações de Budapeste (2002), Bethesda (2003) e Berlim (2003). Esse movimento teve, e ainda tem, a iniciativa de promover discussões e propor ações que garantam o acesso aberto à informação científica, especialmente artigos de periódicos, nas diversas áreas de conhecimento (Pinto, 2022).

A Ciência Aberta surge como um movimento mais amplo, a partir da necessidade de tornar a prática científica mais acessível, transparente e reprodutível (Batthyány; Vommaro, 2023; Unesco, 2022). Desta forma, seus princípios e valores buscam desenvolver práticas alternativas que melhorem processos de pesquisa individuais e colaborativos, bem como sua comunicação e reprodutibilidade, a fim de acelerar a produção e utilização de novos conhecimentos pela sociedade (Batthyány; Vommaro, 2023).

A abertura da ciência vem sendo conduzida por atores públicos, tais como governos nacionais, instituições públicas e agentes financiadores que buscam criar um ecossistema de pesquisa baseado em novas habilidades, práticas, padrões e infraestruturas tecnológicas (Clinio, 2019; González-Teruel, 2022). Essas iniciativas institucionais têm influência direta sobre práticas de publicação e compartilhamento de dados de pesquisa, especialmente no contexto de periódicos, *data journals* e repositórios, objetos centrais desta investigação.

A Comissão Europeia apoia o programa do Horizon 2020, o qual tem como um dos objetivos exigir acesso aberto imediato a todas as publicações científicas e uma gestão responsável dos dados de pesquisa. Além disso, o CoAlition S criado por um grupo de organizações nacionais de financiamento da investigação, com o apoio também da Comissão Europeia e do Conselho Europeu de Investigação (ERC), anunciou o lançamento do Plano S, cujo objetivo é tornar aberto o acesso às publicações de pesquisa geradas por meio do financiamento público (European Commission, 2021; European Science Foundation, c2024).

Em 2021, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) publicou suas recomendações sobre a Ciência Aberta (UNESCO, 2022). Nos Estados Unidos, o Escritório de Ciência e Tecnologia da Casa Branca (OSTP) designou 2023 como o ano da Ciência Aberta. Por fim, mais recentemente,

na cúpula do Grupo dos Sete (G7)¹, em Hiroshima, o governo do Japão estabeleceu que até 2025 todas as publicações financiadas pelo governo serão de acesso aberto (Arita *et al.*, 2024).

Como parte da difusão de recomendações, diretrizes e regulamentos em vários níveis para promover e implementar práticas de Ciência Aberta, diferentes caminhos emergem para promover a colaboração científica (Batthyány; Vommaro, 2023). Dentre esses caminhos, emerge o compartilhamento² de dados de pesquisa como forma de enfrentar grandes desafios, como o da pandemia da COVID-19 e da crise climática, fatores que aceleraram a necessidade de ter maior abertura a esses dados e aos seus resultados (Rovelli; González, 2023).

Embora o compartilhamento dos dados de pesquisa tenha se acelerado nos últimos anos, ele não é um tópico novo nas discussões entre pesquisadores e formuladores de políticas de Ciência e Tecnologia (C&T). Teve seu início, em 2004, quando a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) passou a recomendar que os dados de pesquisas com financiamento público dos países membros fossem disponibilizados publicamente (OECD, 2007). Essa recomendação se materializou na publicação dos “Princípios e Diretrizes para acesso a dados de pesquisa de financiamento público” (em inglês *Principles and Guidelines for Access to Research Data from Public Funding*), em 2007 (OECD, 2007).

As práticas de colaboração e compartilhamento que norteiam a Ciência Aberta vêm sendo enfatizadas em diversas comunidades acadêmicas, tornando indissociáveis as etapas de produção e comunicação dos resultados da pesquisa (Dias, 2025; Seo; Kim, 2020). Com isso, desponta a publicação de dados como um novo modelo de compartilhamento que garante qualidade dos dados por meio de procedimentos básicos da comunicação científica, que incluem a submissão de dados, revisão por pares, armazenamento permanente, citação e avaliação de impacto dos dados, o que pode contribuir para o reúso (Kong *et al.*, 2019; Seo; Kim, 2020).

¹ O Grupo dos Sete (G7) é o grupo dos países mais industrializados do mundo, composto por: Alemanha, Canadá, Estados Unidos, França, Itália, Japão e Reino Unido.

² O compartilhamento ocorre quando o pesquisador intencionalmente disponibiliza seus próprios dados para outras pessoas para uso em pesquisas ou outros empreendimentos científicos relacionados, podendo variar desde a troca privada, através de e-mails, mediante solicitação, na disponibilização num repositório de dados abertos, até o fornecimento como um material suplementar em artigos de periódicos (Enwald *et al.*, 2022; Tenopir *et al.*, 2015).

Similar às vias para a promoção do acesso aberto, Vitón-Castillo, García-Espinosa e Arencibia-Paredes (2020) afirmam que existem duas vias para a publicação de dados de pesquisa: a via verde e a via ouro. A via verde consiste no depósito e publicação de dados em repositórios, como Zenodo e Dryad, enquanto a via ouro envolve armazenar os dados da pesquisa como material suplementar em artigos ou publicar *data papers*³ em *data journals* ou em periódicos tradicionais que possuem seções específicas para este tipo de publicação (Seo; Kim, 2020; Vitón-Castillo; García-Espinosa; Arencibia-Paredes, 2020). Dessas vias, esta pesquisa foca na publicação de *data papers*, seja por meio de *data journals* ou de periódicos tradicionais.

O *data paper* é um tipo de publicação cujo principal objetivo é descrever uma coleção ou conjunto de dados de pesquisa. Ao invés de analisar e discutir os dados, como ocorre num artigo convencional, dedica-se à descrição dos métodos para obtenção/coleta dos dados, permite informar a comunidade científica sobre a existência de coleções de dados para potencial reúso e valoriza o conjunto de dados, conferindo-lhe maior visibilidade, por meio de documentação e metadados estruturados e rigorosos (Curty; Aventureur, 2017; Santos; Nassi-Caló, 2020).

Para orientar esta pesquisa, os *data papers* serão considerados documentos publicados por periódicos científicos, cuja finalidade é descrever detalhadamente os conjuntos de dados de pesquisa (ou *datasets*⁴) e disponibilizá-los em formato bruto⁵ para os pesquisadores interessados, em repositórios de dados.

1.1 DELIMITAÇÃO DO PROBLEMA DE PESQUISA

O avanço das práticas de Ciência Aberta perpassa pela existência de políticas públicas e institucionais que assegurem os recursos, infraestruturas e incentivos, as quais devem abarcar estratégias e ações, na forma de leis,

³ Ao longo desta tese usaremos a expressão em inglês *data paper*, pois acreditamos que tradução literal em “artigos de dados” não representa o objetivo desse tipo de publicação cujo objetivo é descrever um conjunto de dados de pesquisa específico. Para isso, a tradução ideal seria “artigos sobre dados”, no entanto, a literatura brasileira ainda carece de mais reflexões sobre esse tipo de publicação e, por sua vez, sobre a sua tradução para o português.

⁴ Para Souza *et al.* (2022) o termo ‘*dataset*’ (no idioma inglês) se refere a coleções de dados estruturados, enquanto conjuntos de dados (a forma traduzida) abarca tanto dados estruturados quanto não estruturados. Nesta tese usaremos os dois termos como equivalentes, uma vez que a análise dos conjuntos de dados em si, não foi objeto de análise.

⁵ Dado bruto: é aquele dado disponibilizado na forma que foi coletado. Sem passar por um tratamento prévio que antecede a análise.

regulamentos, normativos, diretrizes e programas, que buscam apoiar, facilitar, reconhecer e promover a abertura, pois a implementação da Ciência Aberta envolve custos (Dias, 2025).

No cenário político, o fortalecimento do acesso aberto, dos dados abertos e da ciência é observável por meio de um conjunto de regulamentações nacionais, em diversos países da América Latina, tais como: Peru, Argentina, México, Chile e Colômbia. No Brasil, Equador, Panamá e El Salvador, existem políticas apenas em nível estadual e/ou institucional para o acesso aberto e Ciência Aberta. No entanto, no caso brasileiro não há um quadro regulatório nacional, mas existem diversas ações institucionais isoladas por parte de entidades federais, estaduais e civis. (Dias, 2025; Fortaleza; Bertin; Drucker, 2022; Mugnaini; Igami; Krzyzanowski, 2022; Rovelli; González, 2023).

Dias (2025) identificou 57 documentos que de alguma forma abordam o tema da Ciência Aberta e seus subtemas no Brasil. Os dados de pesquisa (n=27) e o acesso aberto a publicações (n=15) são os mais mencionados e as ações mais reportadas são relacionadas à implantação de repositórios (n=28) e ações de sensibilização e divulgação, de elaboração de diretrizes, princípios e políticas, de elaboração de estudo ou projeto de pesquisa e de articulação, coordenação e gestão (n=14 cada). Todos esses documentos estão vinculados às iniciativas institucionais dos entes federativos, sendo os que mais concentraram ações: Fundação Oswaldo Cruz (Fiocruz), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Controladoria Geral da União (CGU) e Ministério da Ciência, Tecnologia e Inovação (MCTI) (Dias, 2025).

As ações em relação à Ciência Aberta estão relacionadas com as iniciativas lançadas no âmbito da Parceria para Governo Aberto (*Open Government Partnership*), lançada em 2011, da qual o Brasil é um dos fundadores (Brasil, 2013). Trata-se de uma parceria com cerca de 70 países que tem como objetivo tornar os “[...] governos mais abertos, efetivos e responsáveis por meio da construção de compromissos concretos que promovam a transparência, a luta contra a corrupção, a participação social e o fomento ao desenvolvimento de novas tecnologias” (Brasil, 2013, p. 2).

O “Compromisso com a Ciência Aberta”, no âmbito nacional, começa a ser debatido em 2018, no 4º Plano de Ação Brasileiro para Governo Aberto (2018-2021). Posteriormente foram acrescentados novos compromissos e ações no 5º Plano

(2021-2023), publicado em 2021, e no 6º Plano de ação (2023-2027), publicado em 2023 (Brasil, 2023a; Clinio, 2019; Fortaleza; Bertin; Drucker, 2022; Ramírez; Samoilovich, 2021).

No 4º Plano de Ação ficou definido o 3º compromisso de “Estabelecer **mecanismo de governança de dados científicos** para o avanço da Ciência Aberta no Brasil”, culminando no objetivo de “Aprimorar instrumentos de governança da ciência para o avanço da Ciência Aberta” (Brasil, 2018a, p. 35, grifo nosso). O compromisso do 5º Plano de Ação representou a sequência do compromisso anterior, no sentido de contribuir para um “[...] esforço coletivo para que a Ciência Aberta seja valorizada e consolidada, com vistas ao desenvolvimento científico, tecnológico, econômico e social do Brasil”, culminando para o 8º compromisso de criar uma “[...] proposta de **modelo de avaliação que fomenta a Ciência Aberta**” (Brasil, 2021a, p. 45, grifo nosso). No 6º Plano de Ação, ficou definido o 3º compromisso de promover “**Práticas colaborativas** para a ciência e a tecnologia” (Brasil, 2023a, grifo nosso). Cada um desses compromissos se desdobra em marcos mensuráveis, os quais estão arrolados no Quadro 1.

Quadro 1 - Marcos mensuráveis dos planos brasileiros de ação do governo aberto

Nº	4º Plano de Ação	5º Plano de ação	6º Plano de Ação
1	Implantação de uma rede interinstitucional pela Ciência Aberta	Levantamento dos critérios nacionais e internacionais de avaliação de pesquisadores e instituições de pesquisa, com enfoque nas práticas de ciência aberta	Inclusão da pauta de colaboração, transparência e reprodutibilidade da ciência na Estratégia Nacional De Ciência e Tecnologia (ENCTI), por meio das conferências nacionais (CNCTI)
2	Realização de diagnóstico nacional e internacional da Ciência Aberta	Proposição de critérios de Qualificação de Repositórios de Dados e Repositórios de Publicações	Realização de debate sobre a temática do compromisso na comunidade científica para subsidiar a elaboração de diretrizes nacionais para a Ciência Aberta
3	Definição de diretrizes e princípios para políticas institucionais de apoio à Ciência Aberta	Proposição de diretrizes para concessão de fomento e desenvolvimento de projetos de pesquisa	Identificação de áreas, temas e tecnologias estratégicas e críticas para a autonomia tecnológica do país
4	Promoção de ações de sensibilização, participação e capacitação em Ciência Aberta	Proposição de Indicadores de Métricas Alternativas (<i>altmetrics</i>) para medição de Impacto de Pesquisas Científicas	Elaboração de proposta de como incorporar práticas abertas nos critérios de avaliação de pesquisadores e instituições, a partir do 5º Plano de Ação Nacional

Nº	4º Plano de Ação	5º Plano de ação	6º Plano de Ação
5	Articulação com agências de financiamento para a implantação de ações de apoio à Ciência Aberta	Proposição de Indicadores para a Ciência Cidadã	Identificar ações orçamentárias já existentes que possam vir a dar suporte à ciência aberta, propondo a criação de planos orçamentários
6	Articulação com editores científicos para a implantação de ações em apoio à Ciência Aberta	Proposta de qualificação das revistas científicas e critérios alternativos nas dimensões de Ciência Aberta para enriquecer a estratificação do Qualis	Implementação de ferramentas de monitoramento de práticas de Ciência Aberta
7	Implantação de uma infraestrutura federada piloto de repositórios de dados de pesquisa	Levantamento de percepções, e promoção da sensibilização sobre os produtos do compromisso	Elaboração da política de Ciência Aberta do país aderente à Estratégia Nacional de Ciência e Tecnologia
8	Proposição de padrões de interoperabilidade para repositórios de dados de pesquisa	Proposição de Indicadores para a avaliação que contemple a Ciência Aberta	Elaboração de plano de ações integradas para operacionalização de práticas científicas transparentes, responsáveis, colaborativas e reprodutíveis
9	Proposição de conjunto de indicadores para aferição da maturidade em Ciência Aberta	Implantação de observatório para o progresso de práticas de Ciência Aberta no Brasil	Desenvolvimento de ações de sensibilização para atores relevantes relacionados ao compromisso
10		Estudo comparativo do impacto esperado vs. impacto observado na produção dos programas de pós-graduação	Realização de estudos sobre infraestruturas de suporte à Ciência Aberta
11		Proposição de critérios de qualificação de produtos técnicos, tecnológicos e artísticos que favoreçam a Ciência Aberta	Identificação da presença da Ciência Aberta nas iniciativas já existentes para fixar e apoiar a permanência de pesquisadores brasileiros no país e para repatriar pesquisadores
12			Elaboração e divulgação de recursos educacionais abertos sobre práticas de pesquisa transparentes, colaborativas e reprodutíveis
13			Criação de mecanismos de incentivos a práticas de Ciência Aberta (prêmios, hackathon)

Fonte: adaptado de Brasil (2018a, 2021a, 2023a)

Pode-se observar que os compromissos foram sendo ampliados ao longo dos planos de ações, perpassando pela governança dos dados de pesquisa, pela avaliação da Ciência Aberta e pelo fomento das práticas colaborativas para a ciência. Junto com os compromissos, os marcos mensuráveis também foram sendo ampliados, tanto em termos quantitativos, de nove para treze, quanto em iniciativas que versam para a criação de infraestruturas como os repositórios, identificação de ações orçamentárias e criação de mecanismos de incentivos. Com isso, é possível ter uma noção da complexidade que envolve a abertura da ciência e os atores que precisam ser envolvidos.

Os planos de ações têm envolvido vários órgãos governamentais e da sociedade civil, como por exemplo: Ministério da Defesa, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), CNPq, Fiocruz, Comissão Nacional de Energia Nuclear (CNEN), Open Knowledge Brasil (OKBR), Universidade de Brasília (UnB), Rede Brasileira para Educação e Pesquisa (RNP), Scientific Electronic Library Online (SciELO), Associação Brasileira de Editores Científicos (ABEC), Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa (CONFAP), Rede Brasileira de Reprodutibilidade (RBR), Sociedade Brasileira para o Progresso da Ciência (SBPC), dentre outros (Brasil, 2018a, 2021a, 2023a).

Em relação ao alcance dos compromissos estabelecidos nos planos de ações anteriores, é possível destacar que:

- a) o 4º Plano de ação resultou na realização de um mapeamento do panorama da Ciência Aberta no Brasil e no mundo; na elaboração de uma metodologia a ser utilizada na elaboração de políticas institucionais; na realização de ações de sensibilização e capacitação tais como: cursos à distância, encontros e a disponibilização de vídeos curtos sobre o tema no YouTube; na realização de um consórcio de instituições para a aquisição de sequências de identificadores persistentes a serem atribuídos a conjuntos de dados de pesquisa; na formação do Consórcio Nacional de Ciência Aberta (CoNCiênciA)⁶, na assinatura de um acordo para o desenvolvimento e implementação do repositório de dados,

⁶ Esse consórcio tem o objetivo de integrar as instituições interessadas em promover a Ciência Aberta.

o Lattes Data; na realização de uma articulação com a comunidade nacional de editores científicos; na implementação de repositórios de dados em universidades e institutos de pesquisa federais; na elaboração de um conjunto mínimo de padrões de metadados para garantir a integração dos repositórios de dados de pesquisa; e, por fim, na criação de indicadores para aferir o grau de maturidade das instituições quanto à abertura dos dados de pesquisa (Bertin, 2021; Dias, 2025);

- b) no caso do 5º Plano de ação foi realizado o lançamento do Observatório de Ciência Aberta do Brasil (OCABr)⁷, com o objetivo de fornecer informações com indicadores, literatura científica e análises sobre o progresso da Ciência Aberta no Brasil e servir como fonte para subsidiar políticas públicas e institucionais sobre o tema (Brasil, [2021?]; Dias, 2025).

Nesse contexto, Dias (2025) destaca que a falta de uma política federal específica para a Ciência Aberta ou a baixa adesão dos entes federais ao movimento não significa que os resultados dos 4º e 5º planos de ação, tenham sido um vazio de políticas envolvendo a Ciência Aberta, pois a visibilidade trazidas pelos planos através de políticas, ações e atos normativos de entes ligados às pastas de Ciência e Tecnologia, Educação, Saúde, mostra que o tema vai pouco a pouco se disseminando pela Administração Federal.

Em consonância aos compromissos assumidos e alcançados nos planos de ações supracitados, foram identificadas algumas iniciativas de instituições de ensino e de pesquisa referentes ao compartilhamento de dados de pesquisa. Dentre elas destacam-se:

- a) a recomendação para a disponibilização dos dados de pesquisa da Scientific Electronic Library Online (Scielo) (Aventurier; Alencar, 2016);
- b) a publicação do Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã pelo IBICT, cujo objetivo foi demonstrar o valor estratégico e informacional dos dados de pesquisa, além de estimular e apoiar movimentos e iniciativas para Ciência Aberta no Brasil (USP, 2016);

⁷ Disponível em: <https://ocabr.org/pt-br/search/>

- c) a exigência do plano de gestão de dados em editais de fomento a pesquisa, como no caso da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Fapesp, [2023?]) e algumas chamadas do CNPq⁸, dentre outros;
- d) a adaptação das infraestruturas existentes nos repositórios institucionais para a criação de comunidades para disponibilização dos dados, como foi o caso da Universidade Federal do Paraná⁹ e da Universidade Federal do Rio Grande do Sul¹⁰;
- e) a criação de repositórios de dados de pesquisa, tais como o caso do Arca Dados¹¹ mantido pela (Fiocruz, [2023?]), do Repositório de Dados de Pesquisa da Embrapa (Redape)¹², das universidades Paulistas: Universidade de São Paulo (USP)¹³, da Universidade Estadual de Campinas (UNICAMP)¹⁴, da Universidade Estadual Paulista (UNESP)¹⁵, da Universidade Federal de São Carlos (UFSCAR)¹⁶, da Universidade Federal de São Paulo (UNIFESP)¹⁷; da Universidade Federal do ABC (UFABC)¹⁸, da Fundação Getúlio Vargas (FGV), do repositório mantido pelo Scielo, o ScieloData¹⁹, do LattesData²⁰, mantido pelo CNPq, do Aléia²¹ e do Deposita Dados²², ambos mantidos pelo IBICT;
- f) disponibilização pela FAPESP do Metabuscador²³ para possibilitar acesso aos dados de pesquisa, produzidos pelas universidades públicas paulistas, pelo Instituto Tecnológico de Aeronáutica (ITA) e pela Embrapa (Fapesp, [2023?]);

⁸ A exemplo a chamada Chamada Pública MCTI/CNPq/CISB/SAAB AB nº 72/2022.

⁹ Disponível: <https://bdc.c3sl.ufpr.br/>

¹⁰ Disponível em: <https://cedap.ufrgs.br/jspui/handle/2050011959/90>.

¹¹ Disponível em: <https://arcadados.fiocruz.br/>

¹² Disponível em: <https://www.redape.dados.embrapa.br/>

¹³ Disponível em: <https://repositorio.uspdigital.usp.br/?codmnu=9980>

¹⁴ Disponível em: <https://redu.unicamp.br/>

¹⁵ Disponível em: <https://repositorio.unesp.br/handle/11449/183294>

¹⁶ A UFSCar criou uma coleção destinada ao depósito de dados de pesquisa no seu repositório institucional. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/11758>

¹⁷ Disponível em: <https://repositoriodedados.unifesp.br/dataverse/unifesp>

¹⁸ Disponível em: <http://dataverse.ufabc.edu.br/>

¹⁹ Disponível em: <https://data.scielo.org/dataverse/scielodata>

²⁰ Disponível em: <https://lattesdata.cnpq.br/>

²¹ Disponível em: <https://aleia.ibict.br/>

²² Disponível em: <https://depositadados.ibict.br/>

²³ Disponível em: <https://metabuscador.uspdigital.usp.br/>

- g) ainda com o objetivo de estimular a disseminação dos repositórios de dados, o CNPq e a RNP, lançaram um edital para selecionar três instituições de ensino brasileiras para apoiar na criação de seus repositórios de dados de pesquisa (IBICT, 2021) e
- h) recentemente, a partir dos esforços da Fiocruz e do IBICT para tornar seus repositórios de dados em fontes confiáveis, os repositórios Arca e Aléia receberam a certificação CoreTrustSeal²⁴ (Fiocruz, 2024; IBICT, 2025). Essas iniciativas estão sintetizadas na Figura 1.

²⁴ A CoreTrustSeal é uma organização internacional, comunitária, não-governamental e sem fins lucrativos que promove infraestruturas de dados sustentáveis e confiáveis. A certificação visa reconhecer que o repositório de dados adota práticas que tornam os dados em arquivos úteis, significativos e confiáveis no futuro.

Figura 1- Síntese das iniciativas voltadas à institucionalização dos dados de pesquisa



Fonte: elaborado pela autora

Contudo, pode-se verificar que muitas dessas iniciativas foram lançadas e desenvolvidas por meio de decisões administrativas, em universidades e instituições de pesquisa sem o envolvimento direto em políticas ou diretrizes governamentais (Ribeiro; Oliveira; Santos, 2021), o que pode acarretar na sua descontinuidade e na ausência de participação dos pesquisadores e da comunidade interessada.

Partindo desse panorama nacional das iniciativas voltadas à consolidação da Ciência Aberta no Brasil, esta tese se insere nas discussões sobre os dados de pesquisa, mais especificamente naquela que envolve a publicação desses dados em periódicos a partir dos *data papers*.

A adesão aos *data papers* pelos periódicos brasileiros ainda é incipiente sendo identificada nos periódicos: *Encontros Bibli* (PGCIn/UFSC), da área da Ciência da Informação e áreas correlatas, que começou a publicá-los a partir de 2020 (Schiessl; Shintaku, 2022), BioSCIENCE (Associação Médica do Paraná), especializado na área da Saúde, que abriu uma seção para *data papers* em 2022 e o *Entomological Communications* (Sociedade Entomológica do Brasil) especializado na área da Entomologia, que aderiu em 2024, visando se adequar às práticas da Ciência Aberta (Fernandes; Pitta, 2024). Com isso, torna-se importante identificar as práticas dos autores filiados às instituições brasileiras na publicação dos *data papers* para contribuir para a elaboração de políticas que sustentarão tanto a promoção como o desenvolvimento de iniciativas e implementação da Ciência Aberta nas instituições brasileiras (Santos, 2024).

Os *data papers* se inserem no bojo das discussões da Ciência Aberta como uma forma de promover o reconhecimento pelo compartilhamento dos dados de pesquisa (Pampel; Dallmeier-Tiessen, 2014), uma vez que são publicados em periódicos tradicionais ou em *data journals* e tem como proposta fornecer informações sobre o quê, por que, como e quem coletou os dados de pesquisa (Callaghan *et al.*, 2012).

Os *data papers* demonstram propriedades essenciais da publicação de dados, a saber: disponibilidade, documentabilidade, citabilidade, verificabilidade e confiabilidade, uma vez que são revisados por pares. Com isso podem ser atraentes para os pesquisadores, uma vez que fornecem o reconhecimento científico (Chavan; Penev, 2011; Kratz; Strasser, 2014; Schöpfel *et al.*, 2019).

Os *data papers* vêm sendo considerados uma saída para a falta de incentivos e recompensas aos esforços dos pesquisadores para tornar os dados abertamente

disponíveis. Por exemplo, eles elevam o status dos dados de pesquisa a uma publicação científica legítima, revisada por pares e capaz de ser indexada por bases de dados, possibilitam a descrição minuciosa dos dados, facilitando a verificação, replicação e reprodutibilidade em pesquisa e permitem o acesso a diferentes produções associadas ao conjunto de dados, possibilitando o reconhecimento por meio das citações e oportunidades de colaboração entre pesquisadores com interesses comuns (Curty; Aventurier, 2017; Pampel; Dallmeier-Tiessen, 2014; Seo; Kim, 2020; Walters, 2020).

Essa falta de incentivos também foi mapeada pelos especialistas do governo e da sociedade civil que trabalharam nos compromissos do Governo Aberto, como um dos desafios para a implementação da governança dos dados de pesquisa e para a avaliação do avanço da Ciência Aberta (Brasil, 2018b, 2021b).

Portanto, considerando o contexto brasileiro e os benefícios inerentes à publicação dos dados de pesquisa, tem-se como problema de pesquisa: como se caracteriza a publicação de dados de pesquisa, por meio de *data papers* de autores filiados às instituições brasileiras?

1.2 OBJETIVO GERAL E ESPECÍFICOS

O objetivo geral desta pesquisa é caracterizar a publicação de dados de pesquisa por meio de *data papers* por parte dos autores filiados às instituições brasileiras. Sendo os objetivos específicos:

- a) identificar o conjunto de *data papers* publicados por autores filiados às instituições brasileiras;
- b) apresentar as características temporais e temáticas desses *data papers*;
- c) descrever a autoria por meio da filiação institucional, países e fontes de financiamento;
- d) caracterizar os periódicos nos quais os *data papers* foram publicados e
- e) analisar os repositórios de dados de pesquisa, nos quais os dados utilizados nos *data papers* foram publicados.

1.3 JUSTIFICATIVA

Os *data papers* surgem nesse contexto da abertura da ciência como uma saída para os pesquisadores atenderem às exigências das agências de financiamento e dos periódicos em compartilhar os dados de pesquisa, visto que os trazem para o cerne da comunicação científica por meio de práticas convencionais, como autoria, publicação e citação, além poderem ser indexados em bases de dados e, serem potencialmente reusáveis (Walters, 2020).

Este gênero documental se firma como um importante meio para disponibilizar dados de pesquisa em diversas disciplinas e tem ampliado as possibilidades de compreensão sobre como os dados são utilizados, citados e integrados ao processo de investigação (Jiao; Li; Fang, 2023).

Os *data papers*, enquanto componentes da publicação dos dados de pesquisa, apresentam uma interdependência entre os repositórios de dados, periódicos de dados (ou *data journals*) ou periódicos científicos (Schiessl; Shintaku, 2022). Com isso os *data papers* são capazes de envolver diversos atores, tais como: os pesquisadores, as agências de financiamento e os editores de periódicos.

Os pesquisadores são motivados a publicar os *data papers* para atender às regulamentações externas, como a obrigatoriedade para conseguir financiamentos, para ganhar influência pessoal, satisfazer interesses e fortalecer o relacionamento com a comunidade acadêmica, embora também sejam confrontados pelo *habitus* da ciência desenvolvida nas áreas do conhecimento e pelos critérios pelos quais são avaliados (Huang; Jeng, 2022; Ramírez; Samoilovich, 2021).

Desta forma, o pesquisador confronta-se, por um lado, com a exigência de publicar seus dados em repositórios de acesso aberto e, por outro lado, com a avaliação pelo número de publicações em periódicos de alto impacto e pelo número de citações que recebeu (Ramírez; Samoilovich, 2021). Além disso, os custos para a publicação dos *data papers* a partir das taxas de processamento do artigo, os custos associados para depositar os conjuntos de dados em um repositório que pode solicitar outra taxa e todo o trabalho envolvido na preparação dos dados de pesquisa e na redação do *data paper* podem ser barreiras que afetam todo o movimento de publicação de dados (Candela *et al.*, 2015).

As agências de financiamento, além de exigirem o compartilhamento dos dados de pesquisa, exercem um papel importante e crucial no processo de geração

e disseminação do conhecimento no contexto da Ciência Aberta, uma vez que podem estabelecer políticas e critérios como pré-requisitos a serem adotados pelos pesquisadores solicitantes de financiamento público (Mugnaini; Igami; Krzyzanowski, 2022).

Com isso, elas têm o papel de trabalhar em estreita associação com pesquisadores, instituições de investigação e sociedades científicas. Também precisam dar voz ao público e aos seus vários constituintes, trabalhar com editores que estejam dispostos a apoiar o desenvolvimento de um sistema de publicação e comunicação acadêmica que corrija as falhas atualmente observadas, e que esses tenham interesse em cooperar de maneira que seus modelos de negócio estejam entrelaçados a um quadro de avaliação onde o valor intelectual e econômico não se sobreponha aos valores científicos (Guédon *et al.*, 2019).

Os editores de periódicos vem tomando uma série de iniciativas, dentre as quais destacam-se a mudança de estratégias de negócios das assinaturas tradicionais para o acesso aberto, coincidindo com desenvolvimentos globais, como o Plano S e o aumento dos mandatos e políticas de acesso aberto (Butler *et al.*, 2023), criação de seções específicas para publicação dos *data papers* e de periódicos especializados na publicação de dados, os *data journals* (Candela *et al.*, 2015), criação de políticas de disponibilidade de dados (Kong *et al.*, 2019), além da criação de repositórios de dados, como o *Mendeley Data*, mantido pela Elsevier.

Com isso, os editores comerciais têm agido no ecossistema da Ciência Aberta, o que demanda atenção da comunidade acadêmica (Huang; Jeng, 2022). Além disso, especificamente em relação aos *data papers*, estes podem aumentar a bolha das publicações acadêmicas, uma vez que algumas revistas se apresentam abertamente como uma forma de aumentar o currículo do pesquisador com a publicação desses documentos (García-García; López-Borrull; Peset, 2015), além dos lucros já obtidos com a publicação dos artigos tradicionais pelos grandes editores comerciais.

Os *data papers* entram um cenário complexo e com disputas. Arita *et al.* (2024) destacam que a situação atual da publicação dos dados de pesquisa é um campo de batalha entre *publishers* (editores comerciais), autores e agências de financiamento que buscam a melhor solução para publicar artigos com custos de publicação ideais, mantendo a qualidade e a credibilidade das publicações com pouca ou nenhuma restrição de acesso a artigos e aos dados associados.

Para Schiessl e Shintaku (2022), este tema é abrangente e ainda requer estudos e aplicações, mas apresenta certa inovação na comunicação científica. Pinto (2023a) descreve que a comunicação científica via publicação dos dados produzidos em pesquisas ainda é jovem e tende a aumentar, contudo, tal comportamento e sua absorção por áreas e por atores que compõem o cenário da comunicação científica estará sempre em sintonia com o comportamento e características diversificadas de cada área do conhecimento, oportunizando estudos de domínio de acordo com seus padrões de informação e dados.

Esta tese visa contribuir para o entendimento de como os autores filiados às instituições brasileiras estão aderindo a essa prática, podendo ajudar a mapear caminhos, visto que este tipo de publicação tem aumentado entre as diferentes áreas do conhecimento (Ünal; Akbulut, 2025). Com isso, esta pesquisa justifica-se por buscar compreender como se dão as autorias, com ênfase nos aspectos geográficos e institucionais, os financiamentos, os padrões, as políticas editoriais e as práticas de comunicação científica vinculadas aos *data papers*.

Ao investigar esses aspectos, espera-se contribuir para formulação de políticas públicas e institucionais, uma vez que o panorama da Ciência Aberta, embora repleto de iniciativas promissoras, ainda enfrenta barreiras estruturais e culturais, como o *habitus* da ciência e a diversidade de interesses entre disciplinas e regiões, além dos custos envolvidos com as publicações.

Em relação ao *habitus* da ciência, em algumas áreas, como a saúde pública, apenas o artigo original pode demonstrar o valor da investigação e é a prioridade de publicação. Com isso, os *data papers* são considerados como um complemento ou produto secundário que acompanha um trabalho de pesquisa. Com isso, essa impressão pode estar relacionada à sua consideração na avaliação do desempenho acadêmico e dos costumes acadêmicos (Huang; Jeng, 2022).

Corroborando com a considerações dos *data papers* pelas áreas do conhecimento, Silveira, Dias e Neubert (2023a) identificaram que dentre as áreas avaliadas pela Capes, apenas a Odontologia e a Biodiversidade os mencionavam nos seus documentos de área.

Portanto, considerando as atualizações previstas para o Sistema de Avaliação da pós-graduação brasileira, que preveem considerar as mudanças impostas pelo movimento de Ciência Aberta na formação e produção intelectual dos docentes e discentes (Capes, 2025), espera-se que esta tese seja uma fonte de

informação sobre as práticas de publicação de dados de pesquisa para as áreas do conhecimento, principalmente para aquelas com maior adesão por parte dos pesquisadores.

Do ponto de vista prático, esta tese busca entender como os dados são publicados a partir dos *data papers*, o que envolve conhecer todo o ecossistema que permeia essa prática, e inclui identificar os periódicos e os repositórios de dados utilizados. Com isso, dentre as possíveis contribuições que a tese pode trazer, está a discussão sobre a adesão dos periódicos às práticas da Ciência Aberta, a partir das políticas editoriais voltadas à disponibilidade dos dados de pesquisa e a revisão dos conjuntos de dados e *data papers*. Em relação aos repositórios de dados de pesquisa, busca-se apresentar um rol de características, que levem os pesquisadores a refletirem sobre o processo de seleção desses repositórios e as instituições a investirem na capacitação desses pesquisadores no preparo dos dados de pesquisa e seleção de repositórios para compartilhamento desses dados.

Uma vez que a publicação dos dados de pesquisa impacta na comunicação científica (Pinto, 2023a; Schiessl; Shintaku, 2022), esta pesquisa possui aderência à linha de pesquisa “Informação, Comunicação Científica e Competência” do Programa de Pós-Graduação em Ciência da Informação (PGCIN) da Universidade Federal de Santa Catarina (UFSC), cujo objetivo é investigar “[...] aspectos teóricos, conceituais, metodológicos e ações relativas à produção da informação, à **comunicação científica** [...]. Desenvolve estudos relativos ao acesso, à disseminação, à **publicação** [...] e à **preservação** da informação (UFSC, 2023, p. 1, grifo nosso). Esta pesquisa vem contribuir para as discussões em torno das novas tendências da comunicação científica, pois a prática da publicação de dados de pesquisa é uma janela oportuna para a observação da ciência em transformação (Pinto, 2022).

Para exemplificar essa tendência, a temática vem sendo discutida no âmbito do Grupo de Trabalho 7 (GT-7) voltado para a discussões em torno da produção e comunicação da informação em Ciência, Tecnologia & Inovação da Associação de Pesquisa e Pós-graduação em Ciência da Informação (ANCIB), a partir dos trabalhos publicados de Curty e Aventurier (2017), Pinto e Pinheiro (2019), Araujo, Príncipe e Alencar (2021, 2022), Araujo e Príncipe (2023), Silva *et al.* (2024), Silveira, Dias e Neubert (2023b, 2024) apresentados no Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), o que mostra ser um tema em ascensão que impacta em várias áreas da comunicação científica.

Do ponto de vista científico e social, esta tese visa fomentar a produção científica qualificada no âmbito da Ciência da Informação, uma vez que as publicações sobre a temática no âmbito nacional são escassas, conforme seção 3.1.1. Além disso, pretende contribuir para o avanço nas discussões sobre o campo a partir das sugestões práticas e sistematização de evidências empíricas sobre a publicação de dados de pesquisa no contexto brasileiro, campo ainda pouco explorado na Ciência da Informação, especialmente no que se refere à articulação entre *data papers*, periódicos e repositórios. Por fim, acredita-se que a Ciência da Informação é campo profícuo para as discussões neste âmbito, visto que desde a sua gênese, a produção científica e por sua vez, a comunicação da ciência, se constitui como seu interesse de pesquisa.

1.4 ABORDAGEM EPISTEMOLÓGICA

Definir a epistemologia da pesquisa e os estudos na Ciência da Informação é uma tarefa árdua “[...] em razão de sua complexidade e de sua ainda incipiente consolidação científica fazendo com que a mesma ainda tenha uma atuação de forma bastante recursiva para lidar com seu polêmico objeto de estudo [a informação]” (Pando; Almeida, 2021, p. 694).

Portanto, considerando a complexidade epistemológica envolvida na Ciência da Informação, para essa tese o paradigma adotado foi o pragmatismo que se orienta na resolução de problemas práticos do mundo real. A construção do conhecimento é baseada na experiência do pesquisador, onde as percepções do mundo são influenciadas pelas experiências sociais, na qual o conhecimento de cada pessoa é único (Kaushik; Walsh, 2019). Nesta tese, a resolução do problema proposto, se baseia nas leituras e conhecimentos da autora da tese a partir da sua troca com outros pesquisadores ao longo do processo de construção do trabalho.

O pragmatismo foi proposto por Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) como o método que melhor coaduna com abordagem mista, considerada a forma mais adequada ao problema de pesquisa, sendo utilizada na coleta e análise dos dados de pesquisa, o qual baseia-se na proposição de que os pesquisadores devem usar uma pluralidade de métodos que melhor funcione para o problema de pesquisa que está sendo investigado (Kaushik; Walsh, 2019).

Na Ciência da Informação a problemática apresentada situa-se no campo da comunicação científica, que está intimamente ligado às suas origens na segunda metade da década de 1940 até a década de 1960, onde cresceu o interesse pela produção científica (Muller, 2007; Araújo, 2009). Esse cenário foi fortemente influenciado pela Guerra Fria, o conflito entre Estados Unidos e União das Repúblicas Socialistas Soviéticas, no qual, o desenvolvimento científico e tecnológico tornou-se estratégico. Nesse contexto, para o aumento da produtividade e da velocidade de produção de novos conhecimentos científicos, percebeu-se a importância da informação (Araújo, 2009).

Para Mueller (2007) a comunicação desempenha papel central na ciência, onde para ser considerado científico, um determinado conhecimento, num primeiro momento, deve ser validado por outros pesquisadores, o que ocorre antes da publicação, numa espécie de teste de qualidade a que são submetidos os manuscritos candidatos à publicação, processo geralmente denominado “avaliação por pares”. A segunda validação ocorre depois de publicação: se aprovado, o manuscrito contendo o novo conhecimento será publicado como artigo científico, ficando exposto à crítica de todos.

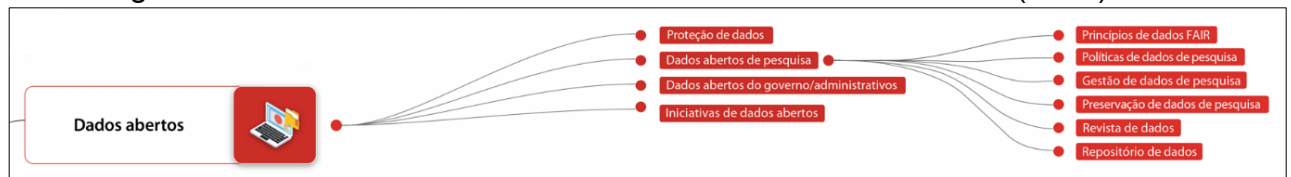
A comunicação científica formal ocorreu desta maneira, até o final do século XX e início do século XXI, onde novos fenômenos e desafios se colocaram para a Ciência da Informação. O advento dos computadores e da internet propiciou um acesso amplo a todo tipo de documentos e registros de conhecimento, do passado e do presente, de todos os lugares do mundo, em tempo real, a partir de diversos dispositivos, inclusive móveis (Araújo, 2018). Somadas a esses fenômenos, destacam-se as discussões que envolvem a abertura da ciência (Gama; Cianconi; González de Gómez, 2022). É nesse contexto que esta tese se situa, pois busca contribuir para as discussões da Ciência Aberta, definida e apoiada pela UNESCO

[...] 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 (UNESCO, 2022, p. 7).

Considerando que a Ciência Aberta envolve várias práticas, Silveira *et al.* (2023) propuseram uma taxonomia que visou representar o conhecimento em volta

do tema. A Figura 2 é um extrato dessa taxonomia que posiciona a tese na faceta “Dados Abertos”, especificamente no rótulo “dados de pesquisa abertos”.

Figura 2 - Extrato taxonomia da Ciência Aberta de Silveira *et al.* (2023)



Fonte: extraído de Silveira *et al.* (2023)

Nota: a figura da taxonomia completa pode ser consultada em: <https://zenodo.org/records/7836884>.

Cabe ressaltar que o tema que estrutura a argumentação da tese, o termo publicação dos dados, não está contemplado especificamente na taxonomia, uma vez que sua terminologia e os conceitos práticos, identificados por Pinto (2023a) ainda são incipientes na Ciência da Informação.

Nesta tese, a publicação dos dados de pesquisa será considerada um tema que unifica a prática da publicação formal dos dados de pesquisa dos autores filiados às instituições brasileiras a partir dos *data papers*, *data journals* e repositórios de dados de pesquisas, sendo que esses dois últimos são rótulos que aparecem na taxonomia.

2 A COMUNICAÇÃO CIENTÍFICA E A CIÊNCIA ABERTA

O conceito de comunicação científica remonta a John Bernal que designa o processo específico de produção, consumo e transferência de informação no campo científico, possuindo duas características principais: a quantidade de informação e a qualidade da informação comunicada (Cunha; Cavalcanti, 2008a). A comunicação científica situa-se no coração da ciência, sendo tão vital quanto a própria pesquisa, uma vez que a sua legitimidade é reivindicada quando analisada e aceita pelos pares (Meadows, 1999). Logo, o sistema de comunicação científica é a infraestrutura da comunidade científica (Mueller, 2006).

A comunidade científica é o grupo social formado por indivíduos que têm como profissão a pesquisa científica, tendo como uma das funções a comunicação, como forma de assegurar os trabalhos em andamento (Cunha; Cavalcanti, 2008b). A comunidade científica, é um dos muitos grupos sociais que compõem a sociedade estando, portanto, sujeita às forças presentes nesse meio. Tem sua autoridade legitimada pelos pares, a qual está ancorada no prestígio individual, conquistada por mérito reconhecido pelos demais, geralmente ao longo de uma carreira (Mueller, 2006).

Os pesquisadores, que são aqueles que formam a comunidade científica, tendem a buscar autoridade por meio do reconhecimento a partir da sua produtividade, a qual é conferida pelas citações recebidas. No entanto, alguns fatores influenciam nesse reconhecimento, como sexo, tempo de carreira e localização geográfica, onde cientistas de países mais proeminentes cientificamente tendem a ser mais citados (Meadows, 1999).

Os periódicos científicos são os meios formais e institucionalizados pelos quais os pesquisadores comunicam os resultados científicos. Seu surgimento remonta ao século XVII, com o *Journal des Savants*, em Paris, e a *Philosophical Transactions*, em Londres. Ambos são marcos da institucionalização das publicações científicas, sendo um passo importante no caminho da partilha e validação de resultados científicos (Kosmopoulos; Schöpfel, 2025).

O papel dos periódicos científicos na comunicação científica está atrelado à necessidade de uma comunicação mais eficiente entre uma clientela interessada em novas realizações e a necessidade de registrar a propriedade intelectual, oferecer audiências e criar visibilidade. Com isso, os periódicos passaram a formalizar a

comunicação científica, que aos poucos recebeu da comunidade científica o status de canal preferencial para a certificação do conhecimento científico e para a comunicação autorizada da ciência (Guédon, 2001; Meadows, 1999; Mueller, 2006).

A legitimidade e a confiabilidade dos artigos científicos publicados nos periódicos é conferida pela avaliação dos pares, que foi proposta, em 1831, por William Whewell à *Royal Society* de Londres, sugerindo que uma comissão de acadêmicos avaliasse os artigos publicados no periódico *Philosophical Transactions*, com o objetivo de aumentar a visibilidade da ciência e assim obter maior reconhecimento público, no Reino Unido. Logo nesse primeiro momento, a proposta não tinha intenção de selecionar artigos de qualidade ou criar um sistema para apoiar decisões editoriais (Santos; Nassi-Caló, 2020).

Apenas no século XIX é que os editores passaram a adotar sistematicamente a avaliação e revisão ao processo de publicação científica como forma de assegurar a integridade dos relatos de pesquisa, emergindo o conceito de revisão por pares como forma de garantir a qualidade dos artigos a serem publicados, bem como a confiabilidade da publicação científica (Kong *et al.*, 2019; Santos; Nassi-Caló, 2020).

A avaliação por pares é tida como um dos pilares da confiabilidade da comunicação científica e tem como objetivos o controle de qualidade das informações publicadas, garantir a correção do trabalho relatado, não permitindo que alegações comprovadamente falsas sejam credenciadas, certificar o trabalho dos autores como válido, atribuição de prioridade pela descoberta científica e fornecimento de crédito pelo trabalho (Richardson; Wolski, 2014).

Embora os estudiosos tenham identificado muitos vieses potenciais que podem afetar sua qualidade e sua neutralidade, a revisão por pares ainda é reconhecida na maioria das instituições de pesquisa como o melhor método para avaliar os méritos do trabalho científico (Mayernik *et al.*, 2015; Santos; Nassi-Caló, 2020). Com vantagens e desvantagens associadas, os tipos de revisão por pares mais conhecidos são a revisão por pares *simples cego*, onde o avaliador tem informação sobre a identidade e filiação do autor, mas o autor não o conhece, a *duplo cego*, onde as identidades do avaliador e do autor são mantidas anônimas, e a revisão aberta, que pode considerar a identidade dos autores e avaliadores e/ou no baseada na abertura do parecer (Santos; Nassi-Caló, 2020).

A qualidade dos artigos científicos foram fornecendo prestígio aos periódicos científicos, que emergiram como o mecanismo mais importante para regular o sistema

internacional de competição científica. Uma vez que os pesquisadores estão numa busca constante por reconhecimento, essas publicações incorporaram um poder enorme, constituindo a base para a gestão de carreiras científicas em todo o mundo (Guédon, 2011; Meadows, 1999).

A posição de prestígio dos cientistas e dos periódicos é mantida e sustentada por um sistema de avaliação baseado em vários indicadores, tais como quantidade de publicações, índices de citação e visibilidade internacional. Entre os indicadores mais utilizados, mas nem por isso isentos de muitas críticas e insatisfações, estão as citações e os diversos índices derivados de sua contagem, especialmente o fator de impacto, que é uma medida da penetração ou visibilidade (Mueller, 2006).

Com os periódicos ganhando centralidade na comunicação científica, a tradição da revisão qualitativa por pares tornou-se cada vez mais criticada por ser subjetiva, e, em meados da década de 1970, indicadores quantitativos baseados na contagem de citações começaram a ser usados como uma medida da produtividade do trabalho científico (Gingras, 2020).

Eugene Garfield criou um índice de citações científicas chamado *Science Citation Index* (SCI) que serviu para reconfigurar a estrutura de uso e avaliação das publicações científicas, trazendo critérios quantitativos. Com isso o número de citações (impacto) e o número médio de citações por artigo em um periódico ao longo de um período de dois anos (fator de impacto) tornaram-se formas aceitas de mensurar a qualidade dos periódicos (Guédon, 2011).

Publicado em 1975, o *Journal Impact Factor* (JIF), ou Fator de Impacto de Periódicos, tornou-se a medida do valor de periódicos como medida de visibilidade na área científica, contribuindo para os *publishers* divulgarem seus periódicos e atrair novos autores. Assim essa mudança deu aos periódicos uma centralidade adicional no sistema de avaliação, pois o acesso a eles passou a ser medido e valorizado e não mais o conteúdo do artigo em si, mas sim a sua visibilidade (medida pelas suas citações) (Gingras, 2020).

Outros indicadores destinados a medir o prestígio dos periódicos foram surgindo, como o *SCImago Journal Rank* (SJR), um indicador bibliométrico que mede o prestígio ou a influência de um periódico, a partir da importância de um periódico que o cita (González-Pereira; Guerrero-Bote; Moya-Anegón, 2010). Seu cálculo é realizado a partir de uma base de dados, sendo que nesse caso a rede de citações é realizada a partir da Scopus, e usa uma janela de citação de três anos que é larga o

suficiente para incluir a maioria das citações, e dinâmica o suficiente para medir a evolução dos periódicos científicos (González-Pereira; Guerrero-Bote; Moya-Anegón, 2010).

No século XX, os periódicos passaram a figurar como um meio de comunicação institucionalizado, dentro do sistema de comunicação formal da ciência. O crescimento de um modelo de negócios também pôde ser notado e a chegada de editoras comerciais especializadas na publicação científica, os transforma em produtos comerciais (Kosmopoulos; Schöpfel, 2025; Santos-D'Amorim, 2021). Essas editoras viram no prestígio adquirido pelos periódicos científicos centrais, ou seja, aqueles mais utilizados e citados pelos pesquisadores, e na dependência dos periódicos na corrida pelo reconhecimento científico dos pesquisadores e instituições de pesquisa, uma fonte de renda baseada no prestígio e por isso pouco afetada pelo preço (Guédon, 2001).

Na década de 1980, veio a crise dos periódicos cujo gatilho foi a impossibilidade das bibliotecas universitárias e de pesquisa americanas continuarem a manter suas coleções e a atenderem a uma crescente demanda de seus usuários, devido aos altos custos cobrados pelas editoras comerciais para manterem as assinaturas de periódicos. Embora essa crise já existisse no Brasil, novas alternativas só surgiram, quando a referida crise atingiu as bibliotecas americanas (Mueller, 2006).

No início da década de 1990, motivada pela crise dos periódicos e pelo surgimento da Web que foi vista como uma oportunidade para a reapropriação da publicação científica pela comunidade numa lógica não comercial, se iniciam as discussões em prol de um conhecimento livre (Kosmopoulos; Schöpfel, 2025).

Os periódicos eletrônicos de acesso livre começaram a aparecer no início da década de 1990, sendo muito semelhantes, em aparência, ao modelo tradicional de periódico impresso, mas inicialmente acessíveis sem pagamento (Mueller, 2006). Logo, as possibilidades e potencialidades das publicações eletrônicas não passaram despercebidas pelas editoras comerciais, que compreenderam o perigo e as oportunidades que elas representavam para sua sobrevivência e atuação no sistema de comunicação científica e se prepararam para isso. Programaram e executaram um plano de transição para o meio eletrônico que envolveu, na maioria dos casos, edição em meio eletrônico e impresso, ambos com acesso via assinatura paga (Mueller, 2006).

Esse plano envolveu a introdução de assinaturas de consórcios ou grandes acordos que, embora tenham expandido a acessibilidade da literatura científica, onde o acesso eletrônico evitou a necessidade de viajar fisicamente até o local onde o artigo estava disponível para lê-lo, na prática, não alteraram substancialmente duas características da comunicação científica na época: o aumento contínuo dos preços dos periódicos e as restrições de acesso derivadas da capacidade de pagar por assinaturas (Anglada; Abadal, 2023).

No início dos anos 2000, na Europa, inicia-se o movimento de acesso livre à informação científica, que tinha como objetivo promover discussões e propor ações que garantissem o acesso aberto à informação científica nas diversas áreas de conhecimento, dando origem aos movimentos e declarações associados ao acesso aberto que tem como marcos fundadores as seguintes declarações: Budapeste (2002), Bethesda (2003) e Berlim (2003) (Kosmopoulos; Schöpfel, 2025; Pinto; Pinheiro, 2019).

O Manifesto de Budapeste sobre Acesso Aberto descreve duas abordagens para o acesso aberto. A primeira abordagem, chamada de **via dourada**, defende a criação de periódicos de acesso aberto ou a transformação dos já existentes, onde, de maneira geral, custos de produção são transferidos para o autor ou para o financiamento público, em vez de serem repassados ao leitor. A segunda abordagem é **via verde**, defende o autoarquivamento por parte dos autores de artigos publicados em títulos por subscrição em repositórios institucionais (Guédon, 2011).

Os movimentos associados ao Acesso Aberto, assim como o que aconteceu com a transição para o periódico eletrônico, novamente levaram as editoras comerciais a flexibilizar seu controle sobre o acesso aos textos que publicam, as quais passam a tomar a frente do movimento como forma de manter ou dobrar os altos lucros com o produto fornecido pela comunidade científica, como o que se conhece por dupla taxação (do inglês, *double dipping*), ou seja, a cobrança simultânea pela taxa de *Article Publishing Charges* (APC) ou Taxa de processamento de artigos e pela assinatura do periódico (Mueller, 2006; Santos-D'Amorim, 2021).

O que o Acesso Aberto fez foi desconstruir o sistema de comunicação científica que se estabeleceu no ambiente impresso, a fim de criar um modelo alternativo, onde a centralidade do leitor, com o modelo "pague para ler" está mudando para a centralidade no autor, com o modelo "pague para publicar" (Anglada; Abadal, 2023).

O movimento do Acesso Aberto trouxe alguns desafios para a comunicação científica, os quais são destacados por Yang e Huang (2025):

- a) o primeiro deles está relacionado à polarização da reputação acadêmica da publicação aberta, onde foram identificados problemas de distorções nas citações de periódicos, a partir da manipulação do fator de impacto. Alguns *publishers* que publicam periódicos predatórios, escolhem aumentar os lucros com APC, sem considerar a qualidade dos artigos e a influência acadêmica do periódico;
- b) o segundo grande desafio está relacionado aos preços cobrados pelas editoras comerciais, que não são transparentes;
- c) o terceiro desafio é o caminho de desenvolvimento de acesso aberto multipolar. Diferentes países e regiões propuseram diferentes regimes de acesso aberto, o que dificulta a promoção do pleno acesso aberto em escala global. Alguns países não conseguem entrar na fase de acesso aberto devido à sua localização geográfica e dependência econômica, especialmente no Sul Global, pelo qual não é razoável pagar o mesmo montante de APC que os países do Norte Global e
- d) o quarto desafio é a falta de cooperação nas políticas de acesso aberto. A contribuição dos artigos científicos é incompatível com a distribuição geográfica dos periódicos científicos. A fim de construir uma estratégia de acesso totalmente aberto, é urgente estabelecer uma plataforma de dados científicos global-local para que se possa indexar e acessar revistas científicas de alta qualidade.

Esses desafios, aliados às mudanças no sistema de comunicação científica, têm levado a distorções no acesso aberto, com o aumento dos preços das APCs cobradas pelas editoras comerciais e a necessidade dos pesquisadores em publicarem em periódicos de maior prestígio (Anglada; Abadal, 2023). Os autores apontam três ações que podem corrigir as disfunções no funcionamento do acesso aberto:

- a) mudar o modelo de avaliação da ciência: o sistema predominante para avaliar publicações e autores baseia-se quase exclusivamente no fator de impacto da revista. Com isso, enquanto houver esse modelo de avaliação os preços continuarão a subir. Portanto, alguns documentos sugerem métricas alternativas, e até qualitativas, para a avaliação dos

- pesquisadores, tais como: a Declaração de São Francisco (DORA), Manifesto de Leiden e o Acordo sobre a reforma da avaliação da pesquisa;
- b) estabelecer medidas para regular os preços das APCs: promover a intervenção nos preços com o estabelecimento de preços máximos para a publicação de artigos, os quais seriam regulados por agências de financiamento da pesquisa público e
 - c) promover modelos alternativos de publicação: explorar e consolidar modelos alternativos aos periódicos, como repositórios com revisão por pares, por exemplo, o Open Research Europe (Anglada; Abadal, 2023).

Portanto, a comunicação científica não é um elo independente. Se a ciência se torna mais aberta, o conteúdo científico também deve ser aberto, e, assim, como os dados de pesquisa, deve garantir eficiência científica (em termos de custo), eficácia (em relação à função que desempenha) e equidade (em relação à sociedade que serve e que a financia) (Anglada; Abadal, 2023).

Nesse íterim, um novo tipo de publicação emerge no campo da ciência e amplia a maneira de comunicá-la: os dados de pesquisa, matéria prima da pesquisa que deram *corpus* à investigação, ganham protagonismo no campo da comunicação científica e compõem o elenco do movimento da Ciência Aberta (Pinto; Pinheiro, 2019).

A comunicação dos dados de pesquisa já é uma realidade, exige estrutura e impõe novos modelos de escrita e produção, o que culmina na criação de publicações específicas com formatos próprios e tipologia de informações diferentes dos periódicos tradicionais, possibilitando novas formas de transmitir e fazer a ciência no Brasil e no mundo (Kosmopoulos; Schöpfel, 2025; Pinto; Pinheiro, 2019).

A Ciência Aberta oportuniza a criação de novas ferramentas, métodos e infraestruturas que têm sido desenvolvidos para a disseminação, processamento, análise e preservação de dados de pesquisa. Com isso, a comunicação científica está mudando, e os *data papers* fazem parte dessa mudança (Candela *et al.*, 2015; Schöpfel *et al.*, 2019). O simples compartilhamento dos dados passa a dar lugar à sua publicação em canais formais da ciência, como periódicos.

2.1 PUBLICAÇÃO DE DADOS DE PESQUISA

Os dados de pesquisa fazem parte de qualquer disciplina e podem apresentar-se de muitas formas, podendo ser brutos sendo coletados diretamente por um instrumento ou um sensor, podem ser produtos de um modelo teórico, simulação ou visualização, ou dados de experimentos conduzidos na bancada de um laboratório, podem ser textos, bibliotecas de imagens digitais e modelos em 3D, tais como os usados para a reconstrução de sítios históricos e mitológicos (Sayão; Sales, 2020). Logo, o termo “dado de pesquisa” tem uma amplitude de significados que vai se transformando de acordo com domínios científicos específicos, objetos de pesquisas, metodologias de geração e coleta de dados e muitas outras variáveis (Sayão; Sales, 2020).

Sayão e Sales (2020, p. 8, grifo nosso) definem os dados de pesquisa como “[...] todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, **tratado e aceito como evidência** pela **comunidade científica** e necessário para analisar, validar e produzir resultados de pesquisa”.

Os dados de pesquisa não são estáticos e nem sempre são de interesse apenas para a comunidade onde são produzidos. Eles se movem entre situações e comunidades, existindo em diferentes contextos e sendo adaptados a diferentes propósitos. Portanto, os caminhos que os dados percorrem também dependem da sua criação e dos contextos de descoberta e reutilização (Gregory *et al.*, 2020).

Para Kong *et al.* (2019) os dados de pesquisa não são apenas resultado da investigação e da literatura científica, mas tornaram-se um importante conteúdo, ferramenta e infraestrutura da investigação científica e até do progresso social. Com isso, eles têm sido motivo de interesse de instituições públicas e privadas que vem criando mecanismos para seu compartilhamento, sendo, portanto, vistos como um bem público (Stuart, 2017).

O compartilhamento de dados é uma iniciativa para tornar a investigação científica e os dados acessíveis a todos. Além disso, visa incitar maior colaboração científica interdisciplinar, acessibilidade e maior reprodutibilidade, aumentando a transparência e a confiabilidade no trabalho científico, características essenciais da ciência (Mayernik *et al.*, 2015; Pinto; Pinheiro, 2019; Tenopir *et al.*, 2020).

Dentre alguns benefícios associados ao compartilhamento dos dados de pesquisa, destacam-se:

- a) a comprovação de resultados obtidos;
- b) a detecção de erros e a identificação de fraudes;
- c) a reprodução e validação dos dados de pesquisa;
- d) a colaboração entre pesquisadores, possibilitando que esses apresentem e respondam perguntas secundárias que estavam fora da investigação original;
- e) a redução de custos no processo de investigação e
- f) o retorno para a sociedade dos investimentos públicos realizados através de agências de financiamento, contribuindo para um melhor impacto socioeconômico das pesquisas públicas (Dias *et al.*, 2020; Vitón-Castillo; García-Espinosa; Arencibia-Paredes, 2020).

Os desafios voltados para o compartilhamento dos dados estão atrelados às resistências dos pesquisadores, principalmente relacionada a preocupações entre os autores dos dados sobre a sua perda de controle, o esforço necessário para compartilhar os dados de forma compreensível e padronizada, os incentivos insuficientes, as práticas de citação não padronizadas e os problemas de qualidade dos dados (Farina *et al.*, 2025; Jiao *et al.*, 2023; Thelwall, 2020).

Somado a esses desafios, no caso da área das Ciências Sociais e Humanas, alguns pesquisadores apontam que o risco da simplificação excessiva ao reduzir fenômenos complexos a conjuntos de dados estruturados, a perda de interpretações e a falta de conhecimento técnico por parte dos pesquisadores, podem prejudicar o compartilhamento e o reuso nessa área (Farina *et al.*, 2025; Jiao *et al.*, 2023; Thelwall, 2020).

Alguns desafios técnicos também criam barreiras ao compartilhamento, tais como: a dificuldade de garantir a interoperabilidade entre diferentes conjuntos de dados, complica os esforços para conduzir análises em larga escala e criar recursos reutilizáveis, a ausência de padrões amplamente compartilhados para codificação de dados e a ausência de uma descrição minuciosa dos dados, por meio de sua gestão, impacta diretamente na usabilidade do reuso de dados em novas pesquisas. Portanto, publicar o conjunto de dados nem sempre garante seu entendimento e a possibilidade de reuso em novas pesquisas (Farina *et al.*, 2025; Pinto, 2022).

Superar esses obstáculos exigirá um esforço conjunto de pesquisadores, instituições e órgãos de financiamento para promover práticas de dados abertos e garantir que os conjuntos de dados gerados sejam tratados como contribuições acadêmicas valiosas e independentes (Farina *et al.*, 2025).

A emergência dos dados de pesquisa, vem demandando que as publicações expandam as formas de socialização dos resultados de pesquisa. Assim, novos fluxos de trabalho envolvendo os dados de pesquisa foram sendo implementados, permitindo que novas modalidades de comunicação na ciência e disseminação da informação fossem exploradas pela comunidade científica, de maneira que os dados possam apoiar a replicação de experimentos, testes de veracidade e reutilização em outros contextos de pesquisa (Pinto, 2022; Santos; Nassi-Caló, 2020).

Dessa maneira, a ideia de compartilhamento de dados vem sendo considerada antiga e a linguagem que está ganhando evidência é a publicação de dados. Essa mudança decorre do papel que os conjuntos de dados vêm ganhando no âmbito da Ciência Aberta, sendo espelhado no modelo de publicação científica, passando a ingressar no registro acadêmico e assim, recebendo o mesmo status que os produtos de pesquisa tradicionais, como os artigos de periódico (Candela *et al.*, 2015; Kratz; Strasser, 2014).

A publicação de dados serve como um veículo poderoso para a introdução dos dados em estruturas científicas existentes, como resposta ao reposicionamento dos dados de pesquisa no sistema de comunicação científica (Li; Greenberg; Dunic, 2020). No entanto, para que a publicação de dados de pesquisa tenha êxito é necessário que ela seja um requisito obrigatório em propostas de projetos de pesquisa e na avaliação de desempenho individual, tenha envolvimento e conscientização e capacitação da comunidade de pesquisa no processo. Portanto, o comprometimento institucional e as declarações obrigatórias por agências de financiamento e periódicos acadêmicos são essenciais (Chavan; Penev, 2011; Dias *et al.*, 2020).

O protagonismo dos dados de pesquisa tem levado a algumas tendências no âmbito da comunicação científica: a criação de periódico de dados (os *data journals*), a criação dos *data papers* e a solicitação pelos periódicos do depósito de dados para acompanhar a submissão de artigos. Em alguns casos, para vincular o artigo aos dados subjacentes, os periódicos vem exigindo a declaração de disponibilidade dos dados de pesquisa, uma declaração na qual os autores explicam como os leitores

podem acessar os dados utilizados em determinado artigo (Federer *et al.*, 2018; Jiao; Li; Fang, 2023; Richardson; Wolski, 2014).

Em relação à disponibilidade dos dados de pesquisa, as primeiras tentativas dos editores de apoiar a sua publicação foram: os dados serem parte integrante do artigo ou os dados serem arquivos suplementares, anexados ao artigo (Candela *et al.*, 2015). Ainda de acordo com esse autor, a primeira abordagem é afetada principalmente pela dificuldade de separação entre os dados de pesquisa e o restante do material e na segunda abordagem, o inconveniente está relacionado com a restrição do tamanho dos arquivos e questões autorais, além de não permitir que leitores encontrem o *link* dos dados independentemente do *link* da publicação (Candela *et al.*, 2015).

Em seguida, a publicação dos dados foi sendo alterada para uma desconexão do artigo original, sendo que os três métodos que passaram a ganhar adesão por parte dos periódicos e autores foram: a) o envio de dados como materiais suplementares para periódicos tradicionais; b) o envio de dados para um repositório de dados e c) publicação da descrição dos dados como um *data paper*, publicados em periódicos tradicionais ou periódicos especializados, os *data journals* (Penev, 2017).

Assim, os veículos de comunicação científica estão também se transformando, trazendo conceitos inéditos, cujos dados vêm sendo evidenciados e incorporados aos artigos de periódicos e de conferência e a outros produtos de pesquisa, transformando os documentos acadêmicos em objetos digitais complexos (Sayão; Sales, 2019).

Pinto (2022, p. 62) define a publicação de dados de pesquisa como “[...] o ato de comunicar e dar publicidade aos dados produzidos por uma pesquisa, quer seja por meio de periódico científico, quer seja por meio de repositório de dados [...]”. Para Dias *et al.* (2020) o processo de publicar dados de pesquisa consiste em um conjunto de ações que agregam valor a um conjunto de dados, possibilitando que sejam acreditados por uma comunidade específica e que tenham a possibilidade de serem universalmente acessíveis através de ambientes de redes.

Para que esse conceito se amplie é preciso ir além do que foi proposto por Pinto (2022), que busca relacionar a publicação dos dados de pesquisa ora aos repositórios de dados ora aos *data journals*. É preciso, trazer elementos práticos ao que foi proposto por Dias *et al.* (2020), portanto, para esta tese, essas abordagens

não são consideradas excludentes, mas sim complementares, ou seja, a publicação de um *data paper*, seja num *data journal* ou num periódico convencional, requer a publicação dos dados de pesquisa num repositório de dados confiável para garantir sua preservação e reuso, além de garantir o reconhecimento da comunidade científica (Candela *et al*, 2015).

Sendo assim, a associação entre *data papers*, periódicos convencionais ou *data journals* e repositórios formam o ecossistema da publicação dos dados de pesquisa (Candela *et al.*, 2015; Lee; Kim, 2021), conforme Figura 3.

Figura 3 - Representação do ecossistema da publicação de dados de pesquisa



Fonte: elaborado pela autora

O ecossistema tem a função de unir todos os elementos associados, onde os *data journals* são produtos de pesquisa de primeira classe totalmente integrados ao sistema de comunicação acadêmica por meio de processos como autoria, validação pelos pares, disseminação, preservação e citação e garantem a confiabilidade a partir da revisão por pares e a recompensa aos autores a partir da citação e indexação nas bases de dados (Lee; Kim, 2021; Walters, 2020).

Os *data papers* são um novo tipo de comunicação científica que fazem parte da documentação produzida pelo processo de publicação de dados, fornecem os detalhes necessários para explicar a coleta e análise dos dados, devendo estar sempre atrelados a um conjunto de dados depositados num repositório (Candela *et*

al., 2015; Curty; Aventurier, 2017; Roa-Martínez; Vidotti; Santana, 2017), onde esses garantem o compartilhamento e o gerenciamento de dados no contexto de repositórios que são realizados de acordo com padrões e procedimentos para sua preservação e utilização, sendo componentes essenciais da infraestrutura para compartilhamento e preservação de dados relacionados aos *data journals* (Lee; Kim, 2021).

Os *data papers* e os *data journals* são fundamentais na publicação de toda a história da pesquisa. Facilitam a interação entre criadores e usuários dos dados, ajudando na visualização e interpretação dos conjuntos de dados, mitigando a sua dispersão e fragmentação, ao mesmo tempo em que complementam a documentação e os metadados dos conjuntos de dados, a partir da sua disponibilização em repositórios. Essas ações promovidas pelos *data papers* e *data journals* viabilizam a transparência dos processos e métodos científicos, e, consequentemente, potencializam e fomentam o avanço da ciência (Jiao *et al.*, 2023; Pinto, 2023b).

Para Soliman, Tripier-Mondancin e Bourg (2024), a redação de um *data paper* é uma etapa do ciclo da publicação de dados a fim de torná-los FAIR e, assim, legíveis por humanos e máquinas, uma vez que esse tipo de artigo visa disponibilizar os dados, indicando as circunstâncias e os métodos de sua coleta, demonstrando sua originalidade, confiabilidade e potencial de reutilização.

Os princípios FAIR foram criados por um grupo de trabalho que publicou, em 2016, na Nature, uma série de princípios, onde *Findable* (**Encontrado**) refere-se a dados e metadados que podem ser descobertos pela comunidade científica, o que requer que esses sejam identificados de forma única e persistente, bem como indexados e descritos em detalhes, *Accessible* (**Acessíveis**). Após serem encontrados, é necessário que sejam publicados usando protocolos padronizados, sendo que *Interoperable* (**interoperáveis**) refere-se ao uso de padrões de compartilhamento e *Reusable* (**reutilizáveis**) garante que a procedência e as condições de reutilização dos dados de pesquisa sejam preservadas (Kosmopoulos; Schöpfel, 2025; Pérez-Luque; Ros-Candeira, 2019; Wilkinson *et al.*, 2016). Em síntese, os princípios FAIR fornecem passos ao longo de um caminho em direção à capacidade de ação mecânica, ou seja, a capacidade das máquinas encontrarem os dados (Wilkinson *et al.*, 2016).

Para que máquinas ou humanos consigam encontrar e reutilizar dados, os metadados também devem fornecer uma descrição rica e detalhada desses dados,

com isso, a documentação escrita nos *data papers* pode ser usada para garantir esses requisitos e contribuir para o círculo virtuoso da Ciência Aberta, reprodutível e transparente, tanto para produtores/autores quanto para reutilizadores de dados (Kosmopoulos; Schöpfel, 2025).

Por fim, concorda-se com Pinto (2022) que afirma que, na comunicação científica, a natureza de cada área define seu comportamento quanto à forma de produzir, disseminar e comunicar a ciência, não existindo um padrão geral. Logo, a comunicação científica via publicação dos dados produzidos em pesquisas, devido à juventude da sua prática, se apresentará de diversas maneiras, podendo ser nomeada de diversas formas. Contudo, estará sempre em sintonia com o comportamento e características diversificadas de cada área do conhecimento (Pinto, 2022).

Para Stuart (2017) uma das áreas mais recentes e menos estudadas de publicação de dados são os *data journals*. No entanto, como eles estão posicionados para desempenhar um papel fundamental no ecossistema de dados, é necessário ter um entendimento básico do papel que eles desempenham. Na próxima seção serão explorados os conceitos referentes aos *data journals*.

2.2 DATA JOURNALS: PERIÓDICOS ESPECIALIZADOS EM PUBLICAR DADOS DE PESQUISA

Os periódicos que compartilham processos e resultados de análises de pesquisa estão no centro da comunicação científica. Porém, com a recente ênfase na importância do compartilhamento e reutilização de dados, os *data journals* surgem como um novo canal para esse fim (Lee; Kim, 2021). Portanto, eles configuram-se nesse cenário como veículos de comunicação, legitimados pela comunidade científica e ampliam o diálogo entre as áreas de conhecimento, abrindo portas para possíveis parcerias interdisciplinares (Pinto; Pinheiro, 2019).

Os periódicos acadêmicos dedicados a dados existem há muitas décadas em áreas especializadas, muito antes da web, como o *Journal of Chemical and Engineering Data*²⁵ publicado pela primeira vez em 1956, *Journal of Physical and*

²⁵ ISSN 1520-5134. Disponível em: <https://pubs.acs.org/toc/jceaax/70/6>

*Chemical Reference Data*²⁶, publicado em 1972, e o *Nuclear Data Sheets*²⁷, uma revista científica, que possui um conselho editorial, revisão por pares e publica artigos que relatam dados, desde 1975, o que pode ser visto como um periódico tradicional com ênfase em dados ou *data journal* com ênfase na avaliação de dados (Stuart, 2017; Thelwall, 2020; Walters, 2020).

No entanto, os *data journals* começaram a ganhar evidência, a partir de 2010, com a criação de alguns deles, tais como: *Journal of Open Archaeology Data*²⁸ (2011), *Geoscience Data Journal*²⁹ (2012) e o *Scientific Data*³⁰ (2013). Seu crescimento tem se dado forma rápida e exponencial nos últimos anos, considerando o protagonismo que os dados de pesquisa vêm ganhando, além de passaram a ser valorizados como ativos de pesquisa autônomos, de alto valor intrínseco e, logo, passíveis de publicação (Aventurier; Alencar, 2016; Santos; Nassi-Caló, 2020; Walters, 2020).

A definição de *data journal* ainda é imprecisa, pairando em torno da tipologia de documentos que ele publica, bem como a frequência e conteúdos dessas publicações e a sua ligação com os repositórios, conforme Quadro 2.

Quadro 2 - Definições de *data journal*

	Autor e ano	<i>Data journal</i>
Tipo de publicação	Candela et al. (2015)	Precursor na categorização dos periódicos que publicam <i>data papers</i> . Designou periódicos “puros” aqueles que publicavam apenas <i>data papers</i> e periódicos “mistos” aqueles publicavam qualquer tipo de artigo, incluindo <i>data papers</i> .
	Stuart (2017)	Todos os <i>data journals</i> devem necessariamente estar dispostos a publicar data papers , mas a proporção de um periódico que é composto por eles pode diferir significativamente, desde aqueles que ainda se concentram principalmente em artigos de pesquisa, mas publicam artigos de dados, até aqueles que se concentram principalmente em artigos de dados com editoriais ocasionais.
	Huang (2018)	Os <i>data journals</i> descrevem conjuntos de dados científicos valiosos para promover o compartilhamento e a reutilização de dados científicos. Eles tratam os conjuntos de dados científicos como os melhores resultados de pesquisa e tornam os dados facilmente acessíveis.
	Thelwall (2020)	Definiu o <i>data journal</i> como um periódico que se anunciava como publicando principalmente artigos sobre dados de alguma forma, independentemente de esses dados estarem ou não associados a outros artigos ou com um conjunto de dados separados.

²⁶ ISSN 1529-7845. Disponível em: <https://pubs.aip.org/aip/jpr>

²⁷ ISSN 1095-9904. Disponível em: <https://www.sciencedirect.com/journal/nuclear-data-sheets>

²⁸ ISSN 2049-1565. Disponível em: <https://openarchaeologydata.metajnl.com/>

²⁹ ISSN 2049-6060. Disponível em: <https://rmets.onlinelibrary.wiley.com/journal/20496060>

³⁰ ISSN 2052-4463. Disponível em: <https://www.nature.com/sdata/>

	Autor e ano	<i>Data journal</i>
	McGillivray <i>et al.</i> (2022)	Os <i>data journals</i> são periódicos acadêmicos revisados por pares dedicados à publicação de <i>data papers</i> , ou seja, artigos de pesquisa que descrevem um ou mais conjuntos de dados com curadoria proficiente
	Santos e Nassi-Caló (2020, p. 29, grifo nosso)	“[...] os periódicos de dados são veículos que se destinam exclusivamente à publicação de artigos de dados [...], fornecem modelos que orientam os pesquisadores sobre como descrever adequadamente os dados e onde depositá-los, [...] podem utilizar guias próprios para a apresentação e descrição dos dados, e [...] podem também adotar critérios específicos de uma área do conhecimento ou disciplina.”
Frequência	Walters (2020, p. 4)	Os <i>data journals</i> são aqueles que publicam <i>data papers</i> regularmente, exclusivamente ou como um tipo de artigo principal , mas outros tipos de documentos também podem ser publicados, a exemplo, os artigos de pesquisa convencionais
Relação entre <i>data journal</i> e repositórios de dados	García-García, López-Borrull e Peset (2015)	São publicações diferentes dos periódicos tradicionais, pois ampliam a forma clássica de conceber artigos de pesquisa e defendem uma aproximação, entre os repositórios de dados e as revistas científicas.
	Pinto e Pinheiro (2019)	Tais periódicos possuem características próprias, além de normas específicas para que sejam validados pela comunidade científica, publicarão <i>data papers</i> que contenham, detalhadamente, a descrição do conjunto de dados da pesquisa, apontando para o repositório onde esses estão depositados .

Fonte: elaborado pela autora a partir de Candela *et al.* (2015), García-García, López-Borrull e Peset (2015), Huang (2018), Pinto e Pinheiro (2019), Santos e Nassi-Caló (2020), Stuart (2017) e Walters (2020)

A partir dessas definições pode-se sintetizar que os *data journals* são canais de comunicação tem como objetivo publicar *data papers* e conjuntos de dados (Huang, 2018; Santos; Nassi-Caló, 2020; Stuart, 2017) de modo exclusivo ou não, com ênfase na qualidade (Huang, 2018) e na publicação regular (Walters, 2020), cujos dados devem estar depositados em repositórios (García-García; López-Borrull; Peset, 2015; Pinto; Pinheiro, 2019).

Os *data journals* se constituem como canais de comunicação, reconhecidos pela comunidade, embora isso não seja unânime em todas as áreas, tendo como objetivo publicar dados de pesquisa e não necessariamente *data papers*, de maneira periódica, permitindo dar visibilidade aos dados oriundos das pesquisas, reconhecimento que não ocorre em trabalhos de pesquisa tradicionais (McGillivray *et al.*, 2022; Pinto; Pinheiro, 2019).

Dentre as suas características, destacam-se:

a) publicação de conjuntos de dados com alto potencial de reutilização;

- b) acesso totalmente aberto para que os conjuntos de dados possam ser citados e reutilizados;
- c) os conjuntos de dados devem poder ser copiados, distribuídos, transmitidos e reutilizados, desde que sejam dados créditos aos autores;
- d) processo de revisão por pares;
- e) exigem a preservação em repositórios públicos de acesso aberto e com longa duração;
- f) identificação segura por meio de um identificador de objeto digital, como o *Digital Object Identifier* (DOI) ou Identificador de Objeto Digital e
- g) frequência rápida de publicação (García-García; López-Borrull; Peset, 2015).

Dentre as vantagens identificadas para os **leitores**, os *data journals* se concentram em dados de pesquisa abertos, o que incentiva a reutilização, uma vez que descrevem a metodologia para obter e analisar os dados, com profundidade semelhante à exigida em patentes, pois em muitos artigos convencionais, a descrição da coleta de dados é reduzida para dar destaque aos resultados e sua discussão (García-García; López-Borrull; Peset, 2015).

Para os **autores**, a vantagem mais importante talvez, seja o fornecimento de crédito de autoria e citação pela publicação dos dados de pesquisa, uma vez que a revisão por pares garante a qualidade e a integridade dos dados e da documentação, facilitando assim a replicação e a reutilização. Ainda como vantagens, tem-se: reconhecimento acadêmico; a indexação dos *data journals* aumenta sua capacidade de descoberta, incentivando assim o uso e a citação de dados, além de promover oportunidades de colaboração entre criadores de dados e outros estudiosos; a hospedagem de dados na plataforma on-line do editor ajuda a garantir o acesso de longo prazo, reduzindo a dependência dos usuários de dados em várias organizações e vários links (Lee; Kim, 2021; Walters, 2020).

Na Coreia do Sul descobriu-se que a principal motivação para os pesquisadores coreanos submeterem em *data journals* com alto fator de impacto era publicar rapidamente os *data papers* para serem reconhecidos como realizações de pesquisa, uma vez que esses passam por avaliações anuais de desempenho com base nos fatores de impacto dos periódicos (Lee; Kim, 2021).

As **instituições dos autores** têm como vantagem a oportunidade de se destacar em rankings que avaliam a produtividade editorial ou o impacto de citações

de pesquisas científicas (Walters, 2020). No entanto, o mesmo autor ressalta que o crédito de autoria é um incentivo confiável apenas se a comunidade reconhecer o valor dos *data journals* para a área do conhecimento.

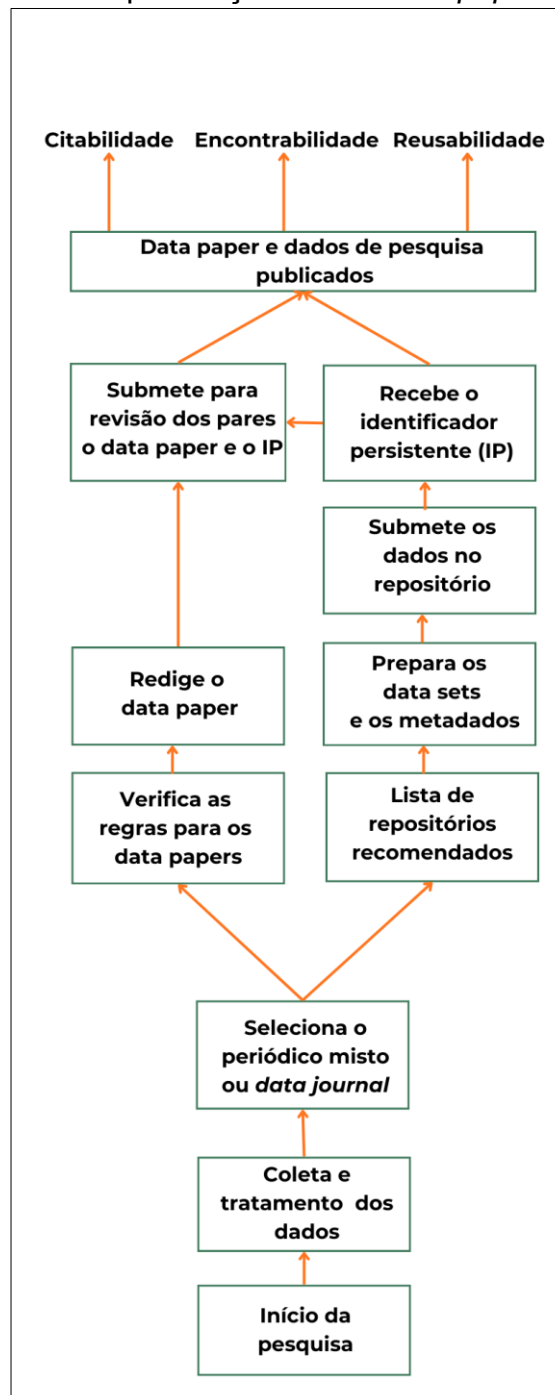
Dentre os desafios encontrados pelos *data journals*, estão aqueles atrelados a atrair os autores para a publicação dos dados e aqueles atrelados à sustentabilidade desses canais, os quais destacam-se:

- a) a resistência dos pesquisadores em publicar: em áreas como a Ecologia, onde projetos de pesquisa de longo prazo financiados externamente (até várias décadas de duração) são a norma, uma vez que é improvável que o crédito associado a publicação dos dados compense os benefícios de manter o acesso exclusivo a esses materiais por um período prolongado (Walters, 2020);
- b) o efeito excludente das APCs, que embora seja o mesmo problema para os periódicos tradicionais pode afastar os autores (Walters, 2020) ou ainda fortalecer a dependência de financiamento para as pesquisas (Huang, 2018);
- c) a publicação em *data journals* pode ser demorada ou excessivamente complexa, pois os periódicos exigem que os autores organizem, descrevam e compartilhem seus dados brutos, aumentando sua carga de trabalho (Huang, 2018; Lee; Kim, 2025);
- d) os pesquisadores frequentemente se preocupam com o fato de o sistema de avaliação acadêmica não valorizar *data journals* (Huang, 2018);
- e) construção de um sistema de avaliação adequado para os *data journals* é um aspecto a ser considerado na sua sustentabilidade dentro do sistema de comunicação científica (Huang, 2018);
- f) a necessidade de práticas sustentáveis de gerenciamento de dados, sendo que, dos *data journals* analisados por Walters (2020), nenhum possui políticas formais e acessíveis ao público que descrevem como garantirão a preservação e o acesso de dados a longo prazo e
- g) os editores de periódicos, tendem a ter dificuldade para atrair e identificar *data papers* de alto nível e, além disso, promovê-los para o público-alvo e manter toda a plataforma online exigem custos operacionais substanciais, o que a longo prazo pode acarretar o aumento dos custos das APCs (Huang, 2018).

Os *data journals* possuem um papel crucial nesse contexto de emergência dos dados de pesquisa, uma vez que a realização de experimentos ou pesquisas para gerar dados científicos valiosos envolve custos substanciais e procedimentos complexos. Portanto, avaliar a qualidade dos dados de pesquisa e sua potencial aplicabilidade é fundamental para determinar a utilidade prospectiva dos resultados da pesquisa, permitindo um sistema eficiente e eficaz de validação e disseminação de pesquisas baseadas em dados (Wang; Xu, 2025).

O processo de publicação de *data journals* é semelhante ao dos periódicos científicos convencionais, sendo que o que difere é que na revisão por pares, os conjuntos de dados depositados em repositórios de dados são incluídos no processo (Seo; Kim, 2020). A partir da perspectiva do autor, na Figura 4 estão elencadas as etapas envolvidas no processo de publicação de um *data paper* num *data journal*.

Figura 4 - Processo de publicação de um *data paper* num *data journal*



Fonte: adaptado de Silveira, Neubert, Dias (2025) e elaborado a partir de Callaghan *et al.* (2012), Lee e Kim (2021) e Santos e Nassi-Caló (2020)

Após a coleta e tratamento dos dados de pesquisa, os autores selecionam um periódico adequado à pesquisa e verificam quais são os repositórios aceitos ou recomendados, bem como demais regras voltadas para a publicação. Em seguida, redigem o texto do *data paper* de acordo com as instruções, modelos e ferramentas recomendados pelo periódico e, em paralelo, os autores preparam os arquivos com

os dados brutos e os metadados para serem submetidos ao repositório. (Callaghan *et al.*, 2012; Lee; Kim, 2021).

É importante ressaltar que cada periódico possui políticas diferentes para repositórios. Uma vez que os dados se tornam disponíveis abertamente no momento da publicação, os repositórios são componentes essenciais da infraestrutura para o compartilhamento e a preservação desses dados de pesquisa (Lee; Kim, 2021).

Na etapa seguinte, os autores submetem o conjunto de dados ao repositório e recebem um identificador e os metadados do conjunto, mas não necessariamente disponibilizam os dados abertamente, podendo deixá-los abertos somente ao editor do periódico, para fins de avaliação pelos pares (Callaghan *et al.*, 2012; Lee; Kim, 2021). Candela *et al.* (2015) alertam que os metadados e os *data papers* não devem ser confundidos, pois os primeiros visam detalhar aspectos dos conjuntos de dados, enquanto os segundos são a metadescrição do próprio conjunto de dados, ou seja, um texto sobre ele.

Em relação à submissão do conjunto de dados é importante salientar que os autores devem realizar a preparação dos dados, definindo por exemplo, a quantidade de arquivos, a sua nomenclatura, a anonimização dos dados e a elaboração de um arquivo “Readme” com o objetivo de complementar as informações descritas nos metadados do repositório. Além disso, o autor precisa escolher uma licença de distribuição de dados, uma vez que ela define as condições de utilização dos dados, sendo que as mais comuns são as Creative Commons (Soliman; Tripiier-Mondancin; Bourg, 2024).

Após receberem o identificador do repositório de dados, os autores submetem o *data paper* ao periódico, adicionando o identificador e os metadados fornecidos pelo repositório durante o arquivamento. Na etapa seguinte, o artigo é submetido ao processo de avaliação por pares, e caso seja aceito, os dados devem ser disponibilizados de forma aberta e sem restrições de acesso para os demais interessados em reutilizarem (Callaghan *et al.*, 2012; Lee; Kim, 2021).

Em relação à confiabilidade, validade, precisão e reprodutibilidade dos *data papers* e dos dados de pesquisa, os *data journals* os garantem por meio do processo de revisão (Lee; Kim, 2021). Assim, a revisão por pares dos dados é um procedimento complexo que requer a acessibilidade dos dados e conhecimentos das estruturas de dados e padrões de metadados, que precisa ser aplicado ao avaliar conjuntos de dados. Sendo assim, os *data journals*, precisam garantir editores de assunto e

estabelecer conselhos consultivos de assunto. Além disso, para aprimorar a qualidade técnica dos dados e sua curadoria, a atribuição de cargos relevantes é cada vez mais necessária (Seo; Kim, 2020).

Numa pesquisa realizada nos *data journals*, Seo e Kim (2020) identificaram cientistas de dados, engenheiros de softwares, analistas de sistemas, programadores e especialistas em banco de dados e em mídias sociais. Além disso, a colaboração entre os responsáveis pela qualidade e os autores é essencial para o aprimoramento dos *data papers* (Wang; Xu, 2025).

Ao contrário da revisão por pares de artigos de pesquisa, a revisão por pares de dados carece de acordo sobre critérios ou padrões consistentes, ou seja, não existem definições claras sobre como os processos de revisão por pares tradicionais podem ser aplicados aos dados, ou o que deve ser garantido por meio da revisão por pares (Kong *et al.*, 2019; Seo; Kim, 2020). Com isso, a revisão por pares aberta tende a ser altamente indicada, sendo que a palavra aberta refere-se à visibilidade das identidades de revisores e autores e transparente significa que o relatório de revisão é publicado juntamente com o artigo (Wang; Xu, 2025).

O que se sabe é que a revisão por pares nos *data journals* precisa considerar os *data papers* e os conjuntos de dados. Com isso ela possui algumas peculiaridades que são destacadas por Kong *et al.* (2019) e apresentadas no Quadro 3.

Quadro 3 - Comparação entre a revisão por pares em periódicos tradicionais e *data journals*

	Periódico tradicional	<i>Data journal</i>
Foco	Nos argumentos e conclusões da pesquisa	Na qualidade e reutilização dos dados, bem como na consistência do <i>data paper</i>
Seleção de revisores	Especialistas no assunto	Especialistas no assunto e especialistas em dados
Modo de revisão	Modo de revisão tradicional	Demanda por revisões abertas

Fonte: adaptado de Kong *et al.* (2019)

Portanto, a revisão por pares nos *data journals* pode ser considerada complexa, pois envolve a qualidade dos dados e dos *data papers*, bem como seu potencial de reutilização, especialistas no assunto ao qual os dados se referem e nos dados propriamente, além disso, as revisões abertas são recomendadas.

Por fim, em relação ao futuro dos *data journals*, é necessário uma maior padronização para descrever, representar e disponibilizar o conjunto de dados, além

do estabelecimento de critérios mais significativos para a avaliação dos dados de pesquisa. A sensibilização dos pesquisadores e demais atores envolvidos na pesquisa sobre a importância de valorizar os dados, desde a sua concepção até a sua disponibilização é outro aspecto que merece atenção para garantir a adesão e a qualidade dos dados publicados (Lee; Kim, 2021; Pinto; Pinheiro, 2019). Os *data papers* surgem nesse contexto para minimizar os desafios elencados.

2.3 DATA PAPERS: ANALISANDO SEU CONCEITO E FUNÇÕES

Outro desenvolvimento recente significativo sobre dados de pesquisa é o gênero acadêmico *data papers*, que gradualmente tomou forma a partir do início dos anos 2010. A nomenclatura dos *data papers* varia conforme os periódicos, sendo também são conhecidos como “*data article*” no *Data in Brief* (Elsevier), “*data description paper*” no *Earth System Science Data* (Copernicus Publications) e “*data descriptor*” no *Scientific Data* (Springer) (Jiao; Li; Fang, 2023). Jiao et al. (2023) identificaram que a falta de padrão na nomenclatura dos *data papers*, ocorre até mesmo entre diferentes periódicos de um mesmo *publisher*.

Embora a disseminação dos *data papers* venha ocorrendo desde 2010, Chavan e Penev (2011) identificaram que a *Ecological Society of America* vem publicando dados desde o ano 2000 na *Ecological Archives* e o *BMC Research Notes*, desde 2008, embora esses periódicos publicassem esporadicamente os dados.

No entanto, as diferentes nomenclaturas dificultam uma definição exata de quando surge a modalidade de artigo voltada para divulgar dados de pesquisa. Os conceitos dos *data papers* estão disponíveis no Quadro 4.

Quadro 4 - Definição de *data paper*

Autor e ano	Definição de <i>Data paper</i>
Chavan e Penev (2011, p. 1, tradução nossa)	“[...] uma publicação acadêmica de um documento de metadados pesquisável que descreve um conjunto de dados acessível on-line específico, ou um grupo de conjuntos de dados, publicado de acordo com as práticas acadêmicas padrão” ³¹ .
Pérez-Luque e Ros-Candeira (2019, p. 154, tradução nossa)	“O <i>data paper</i> é uma publicação científica que documenta em detalhes um ou vários conjuntos de dados acessíveis, descrevendo o contexto em que foram obtidos.” ³²
Schöpfel <i>et al.</i> (2019)	São artigos de autoria, revisados por pares e citáveis em periódicos acadêmicos ou acadêmicos, cujo conteúdo principal é uma descrição de conjuntos de dados de pesquisa publicados, juntamente com informações contextuais sobre a produção e aquisição dos conjuntos de dados, com o objetivo de facilitar a localização, disponibilidade e reutilização de dados de pesquisa. Eles fazem parte do gerenciamento de dados de pesquisa e estão vinculados a repositórios de dados.
Li, Greenberg e Dunic (2020, p. 172, tradução nossa)	“O <i>data paper</i> , um gênero acadêmico emergente, descreve conjuntos de dados de pesquisa e visa preencher a lacuna entre a publicação de dados de pesquisa e artigos científicos.” ³³
Walters (2020, p. 4, tradução nossa)	“[...] descreve os dados que fundamentam um artigo empírico ou um projeto de pesquisa mais amplo. [...] geralmente incluem mais detalhes metodológicos do que normalmente seriam encontrados em um trabalho de pesquisa – informações sobre os procedimentos usados para gerar ou compilar os dados, a população de interesse, os métodos de amostragem, os nomes das variáveis e os códigos de resposta, às dificuldades encontradas, as decisões tomadas, notas do usuário e sugestões para uso posterior dos dados. [...] eles normalmente não incluem tabulações cruzadas, análises multivariadas ou outras tentativas de descrever as relações entre as variáveis.” ³⁴

³¹ Original: “We define a data paper as a scholarly publication of a searchable metadata document describing a particular online accessible dataset, or a group of datasets, published in accordance to the standard academic practices” (Chavan; Penev, 2011, p. 3).

³² Original: “El **data paper** es una publicación científica que documenta detalladamente uno o varios conjuntos de datos accesibles, describiendo el contexto en el que fueron generados y su contenido. (Pérez-Luque; Ros-Candeira, 2019, p. 154)

³³ Original: The data paper, an emerging scholarly genre, describes research data sets and is intended to bridge the gap between the publication of research data and scientific articles” (Li; Greenberg; Dunic, 2020, p. 172).

³⁴ Original “Each data report describes the data that underlie an empirical paper or a broader research project. Data reports often include greater methodological detail than would normally be found in a research paper – information on the procedures used to generate or compile the data, the population of interest, the sampling methods, the variable names and response codes, difficulties encountered, decisions made, user notes and suggestions for further use of the data. Nearly all data reports are peer-reviewed. Although some present simple descriptive statistics or frequency tables, they do not normally include cross-tabulations, multivariate analyses, or other attempts to describe the relationships among the variables. For any empirical study that draws on the data, three elements – the study itself, the data report and the data file – should provide all the information needed to replicate the analysis” (Walters, 2020, p.4)

Autor e ano	Definição de <i>Data paper</i>
Kosmopoulos e Schöpfel (2025)	Os <i>data papers</i> são, portanto, parte de um processo emergente, interdisciplinar e instável, com formatos e práticas que variam de acordo com as disciplinas.

Fonte: elaborado pela autora a partir de: Chavan e Penev (2011), Kosmopoulos e Schöpfel (2025), Li, Greenberg e Dunic (2020), Pérez-Luque e Ros-Candeira (2019), Walters (2020)

Embora haja uma imprecisão quanto ao surgimento e nomenclatura dos *data papers*, no quadro acima é possível verificar que há um consenso sobre seus objetivos: descrever dados ou um conjunto de dados. Schiessl e Shintaku (2023) já haviam identificado uma certa consolidação no termo em pesquisa anterior.

Os artigos científicos estão no cerne da comunicação científica, onde apresentam resultados a partir das evidências geradas pelos dados de pesquisa. Com a sua emergência na comunicação científica, os *data papers* têm como principal função informar e vincular dados de pesquisa em repositórios de dados, contribuindo para sua encontrabilidade e reutilização (Schöpfel *et al.*, 2019).

Dito de outra forma, os *data papers* são uma publicação cujos objetivos são conter fatos sobre dados e descrevê-los (Chavan; Penev, 2011). Assim, enquanto os artigos científicos incluem literatura que fundamenta as discussões com base nos achados da pesquisa, os *data papers* destinam-se a relatar as etapas metodológicas de obtenção dos dados científicos e a detalhar os metadados de forma a permitir maior contextualização para ampliar a reusabilidade dos dados (Santos; Nassi-Caló, 2020).

Para Chavan e Penev (2011) os objetivos dos *data papers* são triplos: fornecer uma publicação de periódico citável que traga crédito acadêmico aos editores de dados; descrever os dados em um formato estruturado e legível por humanos e, trazer a existência dos dados à atenção da comunidade acadêmica, uma vez que no caso da Biodiversidade há uma dificuldade para descobrir e acessar de maneira atualizada e confiável os dados primários de biodiversidade, o que leva a uma duplicação desnecessária de esforços, dificulta a verificação dos resultados e o investimento em pesquisa.

De acordo com Lee e Kim (2021) a função dos *data papers* é fornecer sentido aos dados a partir de descrições detalhadas de como eles foram coletados e tratados, uma vez que os dados disponíveis em repositórios gerais, como Github, *National Center for Biotechnology Information* (NCBI) e sites de instituições, nem sempre conseguem fornecer.

Os *data papers* são catalisadores para o compartilhamento de dados e promovem a transparência. Com isso, eles podem ter um impacto transformador na comunidade acadêmica e na pesquisa em geral, visto que os conjuntos de dados podem ser reutilizados por várias disciplinas para abordar diferentes questões. (Wigdorowitz *et al.*, 2024).

A publicação de *data papers* pode trazer benefícios tanto para o pesquisador individual quanto para a comunidade científica como um todo. Esses benefícios são apresentados no Quadro 5.

Quadro 5 - Benefícios associados a publicação de dados de pesquisa

Pesquisador individual	Comunidade científica
Cumprir os requisitos de promoção na carreira	Ajudar outras pessoas do mesmo campo a descobrir e reutilizar o conjunto de dados para o avanço científico
Aumentar o número de publicações	Combater a crise de reprodutibilidade da pesquisa
Marcar território de pesquisa	Aumentar a credibilidade, a qualidade e o valor dos dados associados
Aumentar a visibilidade dos resultados da investigação, a partir da sua indexação em bases de dados	Qualificador dos <i>data papers</i> como independentes registros acadêmicos formais e/ou para facilitar a subjetividade dos dados em publicações acadêmicas
	Facilitar a vinculação de dados por meio de pesquisas por palavras-chave, bem como, permitir atribuições e reutilização de dados por meio de citações formais, sem a necessidade de obter permissão do responsável pelos dados
Registro de prioridade e autoria em uma publicação acadêmica convencional em qualquer periódico adequado	Oferecer recursos valiosos para futuros pesquisadores, profissionais e pesquisadores interdisciplinares
	Apresentar os dados de forma mais detalhada e completa, aumentando a eficácia da verificação e, contribuindo para a integridade da pesquisa
Processo desafiador de produção de <i>data papers</i>	Possibilitar o rastreamento de uso e citações de dados publicados
	Ofertar uma rica documentação dos dados, que é essencial para a sua reutilização.
Ajudam a disseminar os resultados da pesquisa de acordo com sua citabilidade	Cobrir a falta de dados em uma área de pesquisa específica.
Obter <i>feedback</i> da comunidade científica	Preservar o conhecimento.
	Fomentar a colaboração e o engajamento interdisciplinar e novas interpretações.

Fonte: elaborado a partir de Chavan, Penev (2011); Dimitrova *et al.* (2021); Farina *et al.* (2025); Huang, Jeng (2022, 2023); Lee, Kim (2021); Kim, 2020; Kotti *et al.* (2020), Le Fournier; Schöpfel (2025)

Apesar dos benefícios associados aos *data papers*, foram identificados na literatura alguns pontos negativos, sendo eles:

- a) medo de constrangimentos relacionados à exposição de erros de dados ou práticas inadequadas de manuseio, bem como medos relacionados à perda da vantagem competitiva associada ao compartilhamento de seus dados (Gorgolewski; Margulies; Milham, 2013);
- b) alguns pesquisadores não os consideram como publicações acadêmicas prioritárias, mas sim como um complemento ou produto secundário que acompanha um trabalho de pesquisa e que não tem como objetivo substituir os artigos de pesquisa (Huang; Jeng, 2022; Lee; Kim, 2021);
- c) existem dúvidas se os procedimentos de revisão por pares conseguem verificar completamente os dados associados aos *data papers* (Huang; Jeng, 2023);
- d) alguns pesquisadores os consideram como publicações secundárias em áreas como a Biotecnologia e Genética, uma vez que os repositórios de dados já são predominantes, como por exemplo, o NCBI (Huang; Jeng, 2023);
- e) o escopo limitado dos conjuntos de dados associados aos *data papers* pode ser considerado restrito quanto à sua utilidade prática em comparação com os *datasets* mais extensos depositados repositórios de dados (Huang; Jeng, 2023);
- f) a curadoria dos dados implica num investimento considerável de tempo e custo para os pesquisadores, o que pode comprometer a velocidade de disseminação dos dados (Curty; Aventurier, 2017; Huang; Jeng, 2023) e
- g) algumas instituições não classificam os *data papers* nos seus indicadores de avaliação de desempenho, o que significa que publicá-los pode não trazer benefícios práticos para seus autores (Lee; Kim, 2021; Huang; Jeng, 2023).

É importante ressaltar que esses pontos negativos podem estar influenciados pela visão limitada de algumas áreas e/ou pelo desconhecimento dos pesquisadores sobre os *data papers*, bem como dos editores de periódicos (Yi *et al.*, 2023). Para sanar essas limitações e desconfianças por parte dos pesquisadores e editores de periódicos, Yi *et al.* (2023) propõem programas de treinamento sobre compartilhamento de dados e a publicação de *data papers* para editores, juntamente

com o estabelecimento de políticas obrigatórias para a aceitação pela comunidade e para a adesão dos editores a adoção de políticas nos periódicos.

Em relação à estrutura, os *data papers* apresentam características diferentes dos artigos convencionais, descrevendo metodicamente conjuntos de dados compartilhados em repositórios, desde sua estrutura e formato aos métodos utilizados em sua coleta, organização e processamento, com o objetivo de apoiar a reprodutibilidade e o reúso dos dados (Araujo; Príncipe, 2023).

Le Fournier e Schöpfel (2025) afirmam que de modo geral, os *data papers* costumam ser mais curtos do que um artigo científico, mas não existe uma unanimidade quanto ao seu tamanho, alguns periódicos exigem um máximo de 3.000 palavras enquanto em outros, o *data paper* pode se estender a dezenas de páginas. Embora não haja uma estrutura comum, a literatura é unânime em afirmar que seu foco é a descrição dos dados (Chavan; Penev, 2011; Le Fournier; Schöpfel, 2025). No Quadro 6 estão descritos os elementos comuns que podem ser encontrados nos *data papers*.

Quadro 6 - Elementos comuns dos *data papers*

Tipo	Estrutura	Descrição do conteúdo
Elementos associados aos metadados do <i>data paper</i>	Título	Nome do <i>data papers</i> , devendo preferencialmente estar associado ao conjunto de dados
	Autores	Nome dos autores, podendo ter informações sobre Open Researcher and Contributor ID (ORCID), contribuições e afiliações institucionais
	Resumo	Breve descrição do conteúdo do artigo e dos dados.
	Palavras-chave	Termos controlados ou não que representam o tema
Elementos associados ao conteúdo do <i>data paper</i>	Introdução	Apresenta o contexto geral e específico da pesquisa que levou à coleta de dados, trabalhos anteriores sobre o tema relacionado aos dados
	Métodos	Descreve os métodos e condições (tempo, espaço etc.) de coleta e reprodução de dados, metodologia de amostragem, estratégia de recrutamento dos participantes, metodologia de consentimento informado, recursos para proteger a privacidade dos participantes, materiais utilizados, quem coletou os dados, detalhes de qual <i>software</i> foi usado para criar os dados ou poderia ser usado para visualizá-los e guia para reutilização de dados

Tipo	Estrutura	Descrição do conteúdo
	Descrição do conjunto de dados	Abarca as características dos dados tais como: formatos, estrutura de dados, linguagem, domínios ou faixa de valores que os dados podem assumir, explicações de dados com valores especiais, identificador de conjunto de dados, nome do repositório de dados, data do depósito e licenças adotadas para a reutilização dos dados
	Valor dos dados	Descreve o valor e/ou potencial dos dados para outros pesquisadores ou outros atores (sociedade civil, tomadores de decisão, organizações não governamentais, dentre outros)
Elementos complementares do <i>data paper</i>	Agradecimentos	Agradecimentos a pessoas que contribuíram com a pesquisa
	Citações	Ligação com os outros trabalhos que utilizam os conjuntos de dados
	Referências	Listas dos trabalhos citados ao longo do <i>data paper</i>
	Financiamento	Informações sobre os financiamentos

Fonte: adaptado de Silveira, Neubert, Dias (2025) e elaborado a partir de Chavan e Penev (2011), Chen (2017), Gorgolewski, Margulies e Milham (2013), Kembellec e Le Deuff (2022), Kim (2020), Le Fournier e Schöpfel (2025), Li e Jiao (2021), Liu (2022), Roa-Martínez, Vidotti e Santana (2017), Schöpel *et al.* (2019)

Considerando a estrutura apresentada e o tempo para preparar os conjuntos de dados para o compartilhamento, Kotti *et al.* (2020) identificaram junto aos autores da área de Engenharia de Software que o tempo necessário para produzir um *data paper* é de três a seis meses.

Com isso, a jornada de um *data paper*, da concepção à publicação, passa por uma etapa crucial conhecida como revisão por pares, que exige uma avaliação abrangente não apenas da metodologia da pesquisa e dos resultados científicos, mas também da acessibilidade, qualidade e reutilização dos dados da pesquisa, bem como a integridade das descrições fornecidas, o que implica no controle de qualidade e seleção e, por isso, contribuem para a confiabilidade dos dados de pesquisa (Gorgolewski; Margulies; Milham, 2013; Schöpfel *et al.*, 2019; Wang; Xu, 2025).

Terui, Minamiyama e Yamaji (2019), identificaram os envolvidos e as responsabilidades no processo de revisão dos dados de pesquisa, sendo:

- a) o **autor**: é o proprietário tanto do manuscrito submetido na revista quanto dos dados registrados no repositório de dados, devendo descrever como os dados foram coletados e como usá-los, registrar num repositório de dados, não realizar alteração de dados durante o processo de revisão, podendo fazê-las apenas após a aceitação do trabalho, quando solicitado;

- b) o **revisor**: tem a responsabilidade de verificar o conteúdo do manuscrito e a qualidade dos dados;
- c) o **editor**: desempenha o papel de maior responsabilidade entre as partes interessadas da revista, pois são os responsáveis por nomear o revisor adequado, divulgar os resultados da revisão por pares e devem ficar atentos à manipulação dos dados durante as revisões por pares e revisões pós-publicação;
- d) o **repositório de dados**: é um serviço de informação para registro e publicação de dados. Para que sejam confiáveis, precisam possuir uma página inicial de acesso livre, uma política de licença aberta e uma função de publicação de identificador persistente, além de permitir alguma estratégia de controle de versão dos dados depositados e
- e) a **secretaria**: é responsável pelos procedimentos administrativos do processo de revisão. Na maioria das vezes, o editor também acumula as funções da secretaria.

Portanto, identificar as partes interessadas nesse processo pode contribuir para manter a transparência no processo de publicação dos resultados da pesquisa em periódicos. Por fim, em relação ao processo de revisão por pares, deve-se considerar as necessidades específicas de cada disciplina. Dada a diversidade na prática e nos objetivos da revisão por pares, não existe um princípio uniforme que governe a revisão de *data papers* (Richardson; Wolski, 2014; Wang; Xu, 2025).

A adoção do *data paper* pela comunidade científica, significa a aceitação de que a pesquisa não é (mais) de um pesquisador que submete um artigo a um periódico, mas sim de um conjunto de atores dentro de uma estrutura extensa construída em torno dos dados de pesquisa de um projeto, desde sua produção, curadoria até o arquivamento (Le Fournier; Schöpfel, 2025). Por fim, os *data papers* cumprem seu papel apenas na medida em que se referem aos dados depositados e acessíveis em uma infraestrutura, como repositórios. Sem esse vínculo o *data paper* perde seu valor intrínseco (Le Fournier; Schöpfel, 2025). Portanto, estudar os repositórios de dados de pesquisa enquanto parte do ecossistema de publicação dos dados é imprescindível.

2.4 REPOSITÓRIOS DE DADOS DE PESQUISA: PARTE INTEGRANTE DO PROCESSO DE PUBLICAÇÃO DOS DADOS

Ao longo dos últimos anos, diversas agências internacionais de fomento à pesquisa têm passado a requerer, como critério para a concessão de recursos, o comprometimento por parte dos pesquisadores de que os dados derivados das pesquisas sejam gerenciados e arquivados em repositórios temáticos ou institucionais, a fim de garantir preservação em longo prazo e garantir maior facilidade de compartilhamento (Aventurier; Alencar, 2016; Caballero-Rivero; Sánchez-Tarragó; Santos, 2019). As recomendações perpassam pelo depósito dos dados em repositórios de dados qualificados, reconhecidos internacional ou institucionalmente, que garantam a disponibilidade a longo prazo e o acesso permanente aos conjuntos de dados (Caballero-Rivero; Sánchez-Tarragó; Santos, 2019).

A exigência de depósito dos dados de pesquisa levou os pesquisadores a buscarem recompensas pelo esforço despendido para compartilhar esses dados, levando ao surgimento de novos modelos de publicação, os *data papers* e *data journals*, que evoluíram para dar suporte a dados de pesquisa. Frequentemente, esses modelos se sustentam na estrutura de um repositório, que se encarrega da captura e gestão da descrição dos dados, sendo essenciais para permitir o acesso confiável aos dados subjacentes à pesquisa em diversas disciplinas, sendo que seu uso está arraigado nas práticas de algumas comunidades de pesquisa e nas políticas editoriais (Hrynaskiewicz *et al.*, 2016; Richardson; Wolski, 2014).

Além das agências de financiamento, os editores também passaram a ingressar no ecossistema da publicação dos dados, exigindo que os dados de pesquisa sejam disponibilizados em repositórios apropriados, ou seja, repositórios confiáveis e aceitos pela comunidade científica (Vázquez *et al.*, 2021). Os repositórios de dados estabelecem-se, cada vez mais, como locais para o armazenamento e publicação de dados de pesquisa, uma vez que eles implementam práticas de gestão de dados, para promover a coleta, curadoria, preservação, disponibilidade, disseminação e acesso de conjuntos de dados de pesquisa (Assante *et al.*, 2016; Pinto, 2022).

Um repositório de dados de pesquisa é um serviço online que gerencia o armazenamento de longo prazo e a preservação de dados, fornecendo um arcabouço para a descoberta e acesso dos dados de pesquisa. Além dessas funcionalidades

básicas, alguns repositórios de dados de pesquisa também incluem metainformações sobre dados, licença e identificador único persistente (Vázquez *et al.*, 2021).

Os repositórios podem ser disciplinares ou temáticos ou generalistas ou multidisciplinares. No caso dos primeiros, as plataformas de gestão proporcionam uma curadoria com alto grau de integração e conexão com as demandas das comunidades disciplinares-alvo, onde o design das funcionalidades requer uma compreensão de cada campo de domínio, problema científico, estrutura de informação, bem como questões organizacionais e o contexto do repositório, as necessidades de metadados (Sayão; Sales, 2022; Vázquez *et al.*, 2021). Com isso, esses repositórios tornam-se sistemas de referência para seus respectivos campos de estudos, tais como o *GenBank*³⁵, na área de genômica, e o *Protein Data Bank*, na área das biociências (Sayão; Sales, 2022).

O segundo tipo, os repositórios de dados generalistas ou multidisciplinares, são repositórios que não fazem suposições nem arranjos especiais para aspectos específicos da comunidade ou do tipo de dado, estando, portanto, abertos para publicar qualquer tipo de conjunto de dados de pesquisa que variam em termos de formatos, volume, modelos e estruturas. Além disso, eles não tentam integrar ou harmonizar os dados depositados e interpõem poucas restrições aos metadados assinalados na publicação dos dados (Assante *et al.*, 2016; Sayão; Sales, 2022). Assim, um repositório de dados generalista ou multidisciplinar não está voltado para uma comunidade científica, mas sim para aquelas áreas que não tem repositório de dados de referência para o depósito do tipo de conjuntos de dados que está produzindo (Assante *et al.*, 2016; Sayão; Sales, 2022).

Esses tipos de repositórios têm crescido por causa das exigências de governos, agências de financiamento e periódicos e são desenvolvidos para apoiar a publicação de conjuntos de dados produzidos nos contextos científicos de “cauda longa”, ou seja, no contexto de domínios científicos em que a atividade é realizada em um grande número de laboratórios relativamente pequenos e por pesquisadores individuais que produzem coletivamente a maioria dos resultados (Assante *et al.*,

³⁵ Repositório voltado para o compartilhamento de dados sobre sequências de DNA. Disponível em: <https://www.ncbi.nlm.nih.gov/genbank/>.

2016). Alguns exemplos desse tipo de repositório, são: *Figshare*³⁶, *Dryad*³⁷, *Mendeley Data*, *Zenodo*³⁸, DataHub, entre outros (Sayão; Sales, 2022).

Os dados de pesquisa depositados em repositórios, quando bem geridos, podem trazer benefícios relacionados a economia de recursos, maior transparência para as pesquisas, visibilidade institucional, agilidade no ciclo científico. Os pesquisadores podem reutilizar dados compartilhados para reproduzir pesquisas, validar resultados de pesquisas e propor novas pesquisas em relação a dados relevantes existentes (Bonetti; Arakaki, 2022; Caregnato *et al.*, 2019; Cavalcanti; Sales, 2017; He; Nahar, 2016).

Apesar das exigências das agências de financiamento e das políticas dos periódicos e dos benefícios associados ao depósito dos dados, alguns desafios estão associados ao comportamento dos pesquisadores que embora reconheçam a necessidade, possuem resistências em compartilhar os dados. Também existem desafios associados aos dados, pois alguns são confidenciais dos pacientes ou estratégias que não devem ser publicadas imediatamente e, por isso precisam de restrições de acesso ou algum período de embargo. Além disso, existem desafios associados à falta de incentivo de políticas sobre dados das instituições e à falta de profissionais especializados em curadoria e disponibilização de dados de pesquisas (Aventurier; Alencar, 2016).

Sayão e Sales (2022) ressaltam que para dar suporte à curadoria e à preservação dos dados é necessária a criação de uma infraestrutura adequada, que possa apoiar o desenvolvimento da pesquisa como um todo e não apenas o depósito dos dados no final da pesquisa, o que vai exigir a criação de novos serviços.

Em relação à gestão dos dados, é preciso inicialmente obter-se informações a respeito dos pesquisadores, aqueles que na prática coletam, criam e utilizam estes dados, sendo essencial que os repositórios busquem aderir a princípios e boas práticas, como é o caso dos princípios FAIR (Bonetti; Arakaki, 2022; Caregnato *et al.*, 2019).

A descrição dos conjuntos de dados, também chamados de *datasets*, consiste em agregar informações adicionais, por meio de metadados associados, de modo a ampliar as possibilidades de interpretações sintáticas e semânticas destes

³⁶ Repositório multidisciplinar. Disponível em: <https://figshare.com/>.

³⁷ Repositório multidisciplinar. Disponível em: <https://datadryad.org/stash>.

³⁸ Repositório multidisciplinar. Disponível em: <https://zenodo.org/>.

objetos, tanto por seres humanos, quanto por sistemas automatizados. Este conjunto de informações representa a documentação dos conjuntos de dados, promovendo o intercâmbio e reutilização dos dados que servem de base para as novas descobertas em acesso aberto, aumentando a eficiência e sustentabilidade da pesquisa (Dias *et al.*, 2020; Santos; Nassi-Caló, 2020).

Para isso, os repositórios devem possuir uma série de requisitos:

- a) fornecer identificadores únicos, como o DOI, para os dados;
- b) permitir acesso público aos dados sem necessidade de restrições, exceto nos casos em que sejam necessários para garantir a proteção da privacidade dos sujeitos da pesquisa;
- c) implementar Acordos de Uso de Dados (DUAs) para os conjuntos de dados não públicos³⁹;
- d) possuir licença para acesso aos dados e seu reuso. Salvo em situações em que é necessário sigilo, os dados devem ser disponibilizados por meio de licença menos restritiva possível;
- e) permitir o acesso aos dados em tempo hábil e incluir uma revisão proporcional da fundamentação científica, sem introduzir barreiras desnecessárias;
- f) definir formatos de arquivos aceitos e limites de tamanho dos arquivos de dados;
- g) possuir metadados de descrição;
- h) ser amplamente apoiado e reconhecido dentro de sua comunidade científica;
- i) garantir a persistência e preservação de longo prazo dos conjuntos de dados em sua forma publicada e
- j) implementar requisitos de relatórios relevantes e aprovados pela comunidade (Aventurier; Alencar, 2016; Hrynaskiewicz *et al.*, 2016; Richardson; Wolski, 2014; Santos; Nassi-Caló, 2020).

Nos repositórios as estruturas de dados devem ser gerenciadas de acordo com a sua área de cobertura e com metadados úteis e utilizáveis, o conteúdo e a

³⁹ São conjuntos de dados que foram gerados por meio de pesquisa clínica experimental, como ensaios clínicos, e que não são de acesso aberto, mas estão disponíveis mediante solicitação (Hrynaskiewicz *et al.*, 2016).

qualidade dos dados devem ser gerenciados rotineiramente para evitar que os repositórios se tornem simplesmente contêineres de dados (Lee; Kim, 2021).

Por fim, de acordo com Sayão e Sales (2022), a gestão de dados de pesquisa em repositórios de dados, não deve se aplicar somente aos dados, no sentido mais estrito, mas também aos algoritmos, ferramentas, códigos e workflows que levam aos dados, posto que todos os componentes dos processos de pesquisa devem estar disponíveis para assegurar a transparência, a reprodutibilidade e a reusabilidade da pesquisa como um todo.

2.5 A REPRESENTAÇÃO DOS *DATA PAPERS* E *DATA JOURNALS* NA LITERATURA

Esta seção detalha os trabalhos relacionados aos *data papers* e *data journals*, os quais foram identificados e selecionados a partir de uma revisão sistematizada da literatura, detalhada na seção 3.1.1, que recuperou 77 documentos dentre artigos científicos, editoriais, capítulos de livros e anais de eventos.

Os *data papers* surgem com o objetivo de oferecer crédito acadêmico aos responsáveis pela criação e curadoria dos dados de pesquisa (Chavan; Penev, 2011). Inicialmente, eles surgem no contexto da área da Biodiversidade (Chavan; Penev, 2011; Silva; Stueber; Witt, 2022), mas aos poucos outras áreas viram seu potencial de reconhecimento e valorização acadêmica, como por exemplo na área da saúde, na Neuroimagem (Gorgolewski; Margulies; Milham, 2013) e Enfermagem (Pinto, 2023a), na Ecologia (Pérez-Luque; Ros-Candeira, 2019; Shin *et al.*, 2020) e na área das Ciências Sociais e Humanidades, especificamente nos estudos sobre o Mundo Antigo (Farina *et al.*, 2025).

A partir dos *data journals* *Journal of Open Humanities Data* e *Research Data Journal for the Humanities and Social Sciences*, McGillivray *et al.* (2022) verificaram que os *data papers* melhoram a visibilidade dos conjuntos de dados que descrevem, apoiam artigos de pesquisa que acompanham e aumentam a transparência e a agenda de pesquisa aberta nas Ciências Humanas e Sociais.

Em relação à autoria, os trabalhos estão relacionados às motivações e as percepções dos autores sobre a publicação e o papel dos *data journals* e *data papers* na comunicação científica (Huang; Jeng, 2022, 2023; Kotti *et al.*, 2020; Lee; Kim, 2021), a colaboração científica dos autores que publicam *data papers* (Chen; Li; Jiao, 2022), a contribuição dos autores brasileiros para a biodiversidade global por meio da

publicação de *data papers* e *data sets* (Silva *et al.*, 2024) e a divisão dos trabalhos dos autores que redigiram os *data papers* (Heng; Yonglin; Fenghong, 2025).

O estudo de Kotti *et al.* (2020) investigou as motivações dos autores e usuários dos *data papers* e conjuntos de dados e verificaram, que para os autores, as motivações estão relacionadas à contribuição de certos conjuntos de dados à comunidade, a possibilidade de cobrir a falta de dados em uma área de pesquisa específica, possuindo a expectativa de obter um alto número de citações. Enquanto que para os usuários de conjuntos de dados, o estudo identificou que estes apreciam a alta qualidade dos dados e são desencorajados a reutilizá-los devido à falta de replicabilidade na construção.

Lee e Kim (2021) buscaram compreender as motivações, as considerações e as perspectivas dos pesquisadores filiados a institutos de pesquisa sul-coreanos em relação à submissão em *data journals*. Com isso eles verificaram que os *data journals* fornecem aos pesquisadores incentivos para pesquisas orientadas por dados, uma vez que esses são validados por meio de um processo de revisão por pares podendo ser usados universalmente e preservados em um repositório. Sendo assim, os *data journals* podem promover o conhecimento do assunto relacionado aos dados e contribuir na aceleração dos avanços científicos e tecnológicos e na melhoria da comunicação acadêmica.

Chen, Li e Jiao (2022) investigaram como a geografia influencia os padrões de colaboração científica em *data papers*, especialmente considerando o índice de capacidade científica e tecnológica dos países.

Huang e Jeng (2022) também investigaram as motivações, as considerações e as perspectivas dos pesquisadores em relação à submissão de *data papers* como forma de publicação científica, a partir de entrevistas realizadas com professores do corpo docente de uma universidade de Taiwan que tenham publicado *data paper*. Os autores identificaram que as motivações estão relacionadas a cumprir regulamentos externos, ganhar influência pessoal e concretização de interesses e estreitamento da relação com a comunidade acadêmica; as considerações sobre a publicação dos *data papers* envolvem as características das revistas, como reputação e adequação, bem como o processo de submissão. No entanto, o valor dos *data papers* ainda é questionado na academia, sendo considerados documentos secundários.

Huang e Jeng (2023) exploraram como autores da área de Ciências Naturais percebem as funções dos *data papers* dentro do ecossistema da comunicação

científica e identificaram cinco funções: legitimação, disseminação, acesso/preservação/curadoria, recompensa e verificação (função inédita). Em suma, os artigos de dados não devem ser concebidos como independentes, mas sim como parte de um ecossistema mais amplo de comunicação acadêmica.

Silva *et al.* (2024) investigaram como pesquisadores e universidades brasileiras têm contribuído para a biodiversidade global por meio da publicação de *data papers* e *datasets* na plataforma internacional *Global Biodiversity Information Facility* (GBIF) e verificaram que as universidades brasileiras desempenham papel relevante na produção e atualização de dados sobre biodiversidade.

Heng, Yonglin e Fenghong (2025) a partir do *Contributor Roles Taxonomy* (Credit) ou Taxonomia de Contribuições de Autoria, analisaram a divisão de trabalho em atividades científicas intensivas em dados, com foco específico nos *data papers* publicados na revista *Data in Brief* (Elsevier). Com isso identificaram a necessidade de aprimorar os mecanismos de reconhecimento e atribuição de autoria numa ciência mais aberta e colaborativa.

O **impacto** dos *data papers* e dos dados de pesquisa associados a eles na comunidade foram estudados pelos autores a partir das citações. Stuart (2017) discute os desafios de medir o impacto de dados de pesquisa antes que existam normas consolidadas de citação e compartilhamento desses dados, uma vez que a diversidade dos dados de pesquisa e a cultura acadêmica, a dependência excessiva das citações são fatores que podem influenciar no seu impacto na comunidade.

Jiao e Darch (2020) apresentam os resultados preliminares da análise de citações baseada em conteúdo de *data papers* publicados em dois *data journals*, das áreas das Ciências da Terra e outro das Ciências Físicas e Químicas, na qual os dados indicaram uma subutilização desses, a partir das citações. Jiao *et al.* (2023) também exploraram as práticas de citação na Biomedicina e verificaram que, embora a sua produção venha aumentando, a citação formal dos conjuntos de dados ainda é rara. Li, Huang e Jeng (2025) também identificaram resultados semelhantes para a citação dos *data papers*, o que leva os autores a questionarem sua eficácia quando se fala em reuso de dados. Para Guifeng *et al.* (2024), a partir das citações recebidas por um único *data paper*, identificou-se que podem existir diferenças temáticas na utilização dos dados de pesquisa.

As **mudanças na comunicação científica** devido à criação dos *data papers* e dos *data journals*, foram estudadas a partir do aumento da produção desses

documentos, da criação de novos modelos de periódicos, das mudanças no fluxo e nas políticas editoriais e da ausência de padrões na estrutura dos *data papers* e na sua redação e da criação de maneiras de visualizar esses dados a partir dos periódicos.

O aumento da publicação dos *data papers* pelas várias áreas, vem impactando na comunicação científica e culminando para o aumento da produção desses documentos e a adesão de periódicos voltados para sua publicação (Silva; Stueber; Witt, 2022). Nesse sentido, Pinto e Pinheiro (2019) exploraram os conceitos e componentes dos *data journals*, Dias *et al.* (2020) discutiram o processo de publicação de conjuntos de dados de pesquisa a partir da apresentação dos elementos envolvidos nas fases relativas ao depósito, descrição, atribuição de identificador e revisão e Schiessl e Shintaku (2023), ao realizar um levantamento da literatura sobre *data paper* e suas relações com o fluxo editorial, descrevem informações voltadas aos editores que desejam atualizar as suas políticas editoriais e adicionar esse novo tipo de artigo.

Os *data papers* na comunicação científica ainda carecem de padrões quanto à sua estrutura de apresentação, os trabalhos de Roa-Martínez, Vidotti e Santana (2017) e Soliman, Tripier-Mondancin e Bourg (2024) visam descrever os elementos necessários à redação de um *data paper*. Dimitrova *et al.* (2021) visando automatizar a redação de um *data paper*, propuseram a criação de um protótipo para a importação simplificada de metadados de genômica do *European Nucleotide Archive (ENA)*. Guifeng, Chen, Han (2024) constróem um modelo de associação entre dados de pesquisa, *data papers* e artigos científicos, baseado nos princípios FAIR, para promover o desenvolvimento do compartilhamento aberto de dados científicos.

Em relação ao conteúdo dos *data papers*, Li e Cheng (2018) identificaram como as citações são distribuídas ao longo das seções, sendo verificado que a seção *Background* e *Summary*, por exemplo, funcionam como uma introdução, concentrando a maior parte das citações. Já as seções *Methods* e *Data Record* destacam mais referências a dados e ferramentas, refletindo seu caráter técnico. Kim (2020) também examinou os tipos de informações contextuais que os *data journals* solicitam para serem descritos nos *data papers*.

Li, Greenberg e Dunic (2020) realizaram uma análise de conteúdo em *data papers* publicados e verificaram que não há um padrão para a descrição entre os periódicos, o que também se confirmou nos estudos de Li e Jiao (2021) que afirmam

que dada a curta história dos *data papers*, ainda estamos longe de uma compreensão clara de sua presença no sistema de comunicação acadêmica. Ainda, Li e Jiao (2022) analisaram os resumos dos *data journals* *Scientific Data* (Springer) e *Data in Brief* (Elsevier) para verificar quais movimentos retóricos são usados no gênero e identificaram que ele é tanto uma publicação científica que segue normas de escrita tradicionais quanto um descritor de objetos de dados.

Breure, Doorn e Voorbij (2022) relatam uma experiência de “vitrine de dados” que consiste em tornar os *data papers* publicados no periódico *Research Data Journal for the Humanities and Social Sciences*, em hiperdocumentos contendo uma rede com multimídia incorporada e *links* para conteúdo aprofundado, claramente sinalizados por indicadores contextuais locais na forma de botões, capturas de tela e mapas clicáveis, tornando os conjuntos de dados mais acessíveis aos usuários, especialmente aos não especialistas.

Além da análise dos artigos científicos e de trabalhos apresentados em eventos para a elaboração dessa seção, também foram selecionados os **editoriais de periódicos**, tendo em vista identificar como os editores se manifestam quanto aos *data papers* e *data journals*, e, assim, identificar sua adesão a partir das áreas.

Em relação aos *data papers*, no editorial do periódico *Environmental Chemistry Letters*, Lichtfouse (2018) destaca que são publicações a serem realizadas pelos pesquisadores como forma de obter maior visibilidade, maior divulgação, mais facilidade para encontrar empregos e sucesso de bolsas. Machuca-Martínez (2020) destaca a importância dos *data papers* e afirma que eles são tão importantes quanto os artigos tradicionais.

O primeiro editorial identificado foi escrito por Li-Scholz (2001), editora do *data journal Atomic Data and Nuclear Data Tables*, destinado a publicação de resultados experimentais e teóricos em Física Atômica, Física Nuclear e áreas afins. A editora destaca que, com o crescimento da publicação eletrônica e com a complexidade dos dados computacionais, torna-se importante a disponibilização dos dados brutos como material suplementar dos artigos. Com isso, destaca-se a importância da disponibilidade dos dados para consulta dos leitores, muito antes das práticas associadas à Ciência Aberta. Posteriormente, no mesmo *data journal*, Pritychenko (2014), ao assumir como editor, anunciou que, devido ao alto padrão do periódico e das demandas que envolvem um rigoroso critério de triagem, o *data journal* passará por revisão do editor e da equipe editorial.

Newman e Corke (2009) relatam a criação de uma seção destinada a publicar *data papers* no periódico *International Journal of Robotics Research*, especializado em robótica, cujo objetivo é facilitar e incentivar a disponibilização de conjuntos de dados de alta qualidade e revisados por pares para a comunidade.

O editorial da *Nature* de 10 de outubro de 2013 anunciou lançamento do *data journal* multidisciplinar *Scientific Data*, com o objetivo de fornecer créditos aos pesquisadores que criam e compartilham dados, além de fornecer detalhes de como os dados foram produzidos e quais experimentos de controle de qualidade foram realizados, para que eles possam ser reutilizados (Announcement..., 2013).

Ainda nesse mesmo ano, Smit *et al.* relatam a criação do *data journal*, especializado na área da Biodiversidade, o *Biodiversity Data Journal* (Pensoft Publishers), cujo objetivo é integrar dados e texto em uma única publicação, convertendo vários tipos de dados de biodiversidade (por exemplo, ocorrências de espécies, listas de verificação ou tabelas de dados) em texto para uso legível por humanos, ao mesmo tempo em que torna as unidades de dados do mesmo artigo colhíveis e disponíveis para *download*.

Rushby (2015) anuncia a publicação de um número especial, em 2015, do periódico *British Journal of Educational Technology*, especializado em pesquisa educacional, cujo objetivo é publicar *data papers* para provar que a publicação de dados qualitativos é possível na área.

Le Deuff (2018) anunciou a criação de uma seção destinada à publicação de *data papers* no periódico *Revue Française des Sciences de l'information et de la Communication*, da área da Ciência da Informação e Comunicação, com o objetivo de alinhar-se à lógica do acesso aberto, publicando conjuntos de dados que permitam a outros replicar, verificar, comparar os resultados ou reutilizá-los em trabalhos derivados. Posteriormente, Kembellec e Le Deuff (2022), destacam que a publicação dos dados de pesquisa é necessária para atender às exigências dos órgãos de financiamento franceses. Os autores também destacam que, devido ao caráter qualitativo dos dados das pesquisas na área, exige que o formato dos *data papers* seja diferente daqueles produzidos pelas Ciências Exatas, sendo necessário uma contextualização mais expressiva de como os dados foram coletados.

El-Tawil e Agrawal (2019), no editorial da revista *Journal of Structural Engineering*, especializada em engenharia industrial, anunciaram o lançamento de uma seção para publicação de *data papers*, com o objetivo de promover o mais alto

nível de esforços de preservação de dados. Os editores destacam, ainda, que a iniciativa para criar uma seção para publicação de *data papers* e disponibilização dos dados de pesquisa, veio das pressões dos financiadores e para seguir as práticas de outros periódicos.

Puerta-Piñero, Pérez-Luque e Rodríguez-Echeverría (2020) descrevem a adesão à publicação de *data papers* no periódico *Ecosistemas*, da área da Ecologia, a partir do ano de 2021, com o objetivo de cumprir com uma lacuna de publicações em espanhol, além disso, visa publicar esse tipo de documento de maneira gratuita.

Ainda em 2020, Van Oorth e Noordegraaf (2020), no editorial da revista *Research Data Journal for the Humanities and Social Sciences*, anunciam a publicação de *data papers*, sendo que esse tipo de documento pode contribuir para melhorar a responsabilidade das pesquisas sobre a história das artes cênicas.

Raub *et al.* (2023) anunciaram a criação de uma seção para publicação de *data papers* no periódico *Arachnologische Mitteilungen*, especializado em aranhas, tendo em vista atender às exigências das agências alemãs para publicação de dados.

Nesse mesmo ano, Schneider e Hagener (2023) também informaram a criação de uma seção destinada a publicar *data papers* no periódico NECSUS, especializado em estudos de cinema e mídia, com o objetivo de divulgar os dados de pesquisa dessa área.

O editorial escrito por Farina, Marongiu e Rodda (2024) apresenta uma coleção especial de *data papers* voltada para promover a visibilidade e o compartilhamento aberto de dados de pesquisa na área de estudos do Mundo Antigo, publicados no *data journal Journal of Open Humanities Data*.

Por fim, Fernandes e Pita (2024) apresentam a criação de uma seção destinada à publicação de *data papers* no periódico *Entomological Communications*, da área da Entomologia, cujo objetivo é facilitar a disseminação e a reutilização de dados, se adaptando assim às práticas da Ciência Aberta. Cabe ressaltar que esse foi o único editorial recuperado na revisão de um periódico brasileiro.

Portanto, a partir dos editoriais é possível verificar que aos poucos as áreas estão se adaptando à publicação dos dados de pesquisa, seja a partir da criação de seções para publicar *data papers*, seja pela publicação de números especiais para esse fim ou seja com a criação de periódicos especializados, os *data journals*.

Os trabalhos que abordam as **diretrizes editoriais** tratam da adoção de modelos para conectar os dados clínicos não públicos aos repositórios de dados

(Hrynaskiewicz *et al.*, 2016), da identificação de diretrizes para revistas das Ciências da Saúde que publicam *data papers* (Araujo; Principe, 2023) e da identificação de atitudes dos editores de periódicos quanto à adoção de políticas de compartilhamento de dados e publicação de *data papers* (Yi *et al.*, 2023).

O trabalho de Hrynaskiewicz *et al.* (2016) traz a proposta de desenvolver um modelo de publicação que conecta artigos científicos revisados por pares a dados clínicos não públicos, por meio de repositórios confiáveis e metadados persistentes, garantindo transparência, reprodutibilidade e crédito científico.

O trabalho de Araujo e Principe (2023) teve como objetivo analisar as diretrizes editoriais de revistas científicas de Ciências da Saúde que aceitam *data papers*, tendo em vista identificar diretrizes voltadas à sua elaboração, visando apoiar editores de revistas brasileiras.

Yi *et al.* (2023) investigaram as atitudes de editores e publicadores de periódicos acadêmicos coreanos em relação às políticas de compartilhamento de dados de pesquisa e à publicação de *data papers*. A pesquisa revelou que apenas alguns periódicos coreanos implementaram políticas de compartilhamento de dados e os editores possuem uma experiência limitada no que condiz à publicação dos *data papers*.

Wigdorowitz *et al.* (2024) apresentam uma reflexão sobre a experiência de compor parte da equipe editorial do *Journal of Open Humanities Data*.

Os trabalhos sobre a **revisão por pares** exploram o cenário em evolução da publicação de dados, os processos de revisão por pares e a garantia de qualidade em periódicos de dados acadêmicos.

O trabalho de Richardson e Wolski (2014) examina os desafios da revisão por pares e propõe papéis expandidos para repositórios institucionais no suporte a novos modelos de revisão, fornecendo a base para entender como os repositórios digitais podem suportar processos modernos de publicação e revisão de dados.

Os autores Kong *et al.* (2019) propõem um índice hierárquico de avaliação da qualidade dos dados com base nos princípios FAIR, tendo em vista promover o compartilhamento de dados de qualidade e garantindo a acessibilidade e a reutilização de dados.

Terui, Minamiyama e Yamaji (2019) desenvolveram e testaram um sistema de revisão por pares para detectar e prevenir a manipulação inadequada de dados.

Com isso os autores demonstraram na prática medidas para garantir a integridade de dados em *data journals*.

Seo e Kim (2020) realizaram uma revisão sistemática para identificar sistemas de revisão por pares, critérios e estruturas editoriais em nove *data journals* indexados pela Web of Science (WoS) e, com isso, apresentaram evidências empíricas de práticas atuais na revisão por pares de *data journals*, mostrando convergência para critérios padronizados.

Por fim, Wang e Xu (2025) investigaram as diretrizes de periódicos que publicam *data papers* para identificar as preocupações tanto do ponto de vista dos periódicos quanto dos revisores em relação à avaliação desses documentos. Foram identificados trabalhos que tratam da indexação tanto de *data journals* quanto dos *data papers* em bases de dados.

Li, Lu e Jiao (2021) verificaram a indexação de 25 *data journals* nas bases de dados WoS, Scopus, Crossref e Dimensions, e identificaram que tanto o Crossref quanto o Dimensions possuem ampla cobertura de periódicos exclusivamente de dados, em comparação com a WoS e a Scopus, duas das bases de dados científicas mais populares, e que abrangem apenas cerca de metade dos periódicos de dados arroladas na lista elaborada pelos autores.

Jiao, Li e Fang (2023) exploraram a indexação de 18 *data journals* nas bases de dados Web of Science, Scopus, Dimensions e OpenAlex para verificar até que ponto essas bases de dados cobrem os *data papers* publicados por eles. Os resultados mostram inconsistências significativas na indexação e rotulagem de *data journals* e *data papers* pelas bases de dados estudadas.

Ambos os trabalhos citados afirmam que estudar a indexação dos *data journals* e *data papers* poderá contribuir para estudos sobre a utilização dos dados pelos pesquisadores, as citações dos dados, os papéis desempenhados pelos dados de pesquisa e pelos artigos científicos no sistema de comunicação acadêmica. Como exemplo, sobre o uso das bases de dados como fontes de pesquisa para esse tipo de documento, foi identificado o trabalho de Silveira, Neubert e Dias (2023) que buscaram *data papers* sobre a COVID para entender as características dos conjuntos dados, e assim, identificar ações que permitam a reutilização desses dados a longo prazo.

Por fim, foram identificados ainda trabalhos que tratam **das características gerais** dos *data journals*. O primeiro estudo que visou mapeá-las foi o de Candela et

al. (2015) que buscou entender as abordagens inerentes à implementação do conceito de *data papers*, incluindo como descrever adequadamente um conjunto de dados, como promover a sua disponibilidade de conjuntos de dados, como citá-los e como garantir a sua qualidade e o acesso aberto aos conjuntos de dados.

García-García, López-Borrull e Peset (2015) apresentaram uma compilação e análise de diversos periódicos especializados em dados, descrevendo suas origens, evolução atual e características. Schöpfel *et al.* (2019) identificaram as disciplinas, editores e modelos de negócios, bem como sua estrutura, tamanho, formatos, metadados e licenças utilizadas pelos periódicos.

Thelwall (2020) analisou o impacto do *data journal* que mais concentra a publicação de *data papers*, o *Data in Brief (Elsevier)*, e identificou uma contribuição positiva para a ciência, permitindo o acesso a vários tipos de dados, embora seus artigos raramente levem à reutilização de dados. Walters (2020) buscou identificar as características de 13 *data journals* das áreas da Biologia, Ciências Ambientais, Química, Medicina e Ciências da Saúde e Giner-Miguelez, Gómez e Cabot (2025) analisaram o conteúdo dos *data papers* de diferentes domínios, avaliando sua cobertura e tendências nas dimensões solicitadas e comparando-os com aqueles de um local focado.

Também foram identificado um grupo de trabalhos que objetivaram analisar as características relacionadas à publicação de *data journals* e *data papers* por autores filiados às instituições brasileiras. Os trabalhos de Araujo, Principe e Alencar (2021) buscaram identificar os periódicos utilizados por pesquisadores do Brasil para a publicação de seus dados de pesquisa. As mesmas autoras, em 2022, buscaram na literatura publicada entre 2016 a 2021, a existência de estudos que tratam sobre a publicação de *data papers* em periódicos científicos (Araujo; Principe; Alencar, 2022). Os trabalhos de Silveira, Dias e Neubert (2023b) analisaram como os *data papers* da área da Biodiversidade escritos por pesquisadores filiados às instituições brasileiras, estão sendo publicados. No ano de 2024, um outro trabalho desses autores buscou caracterizar a publicação dos dados de pesquisa dos autores afiliados à Universidade Federal de Santa Catarina (Silveira; Dias; Neubert, 2024).

Por fim, com base nesses estudos pode-se verificar que a publicação de *data papers* envolve diversos aspectos que variam desde os incentivos aos autores, os impactos nas áreas do conhecimento, os impactos nas práticas editoriais dos periódicos tradicionais e *data journals*, a identificação do alinhamento entre as

políticas editoriais e o efetivo depósito em repositórios de dados, e até o impacto no mercado editorial e financiamentos. Apesar das contribuições apresentadas, observa-se que houve menor atenção às práticas efetivas de publicação de dados, especialmente em contextos nacionais específicos, lacuna que esta tese busca explorar. Portanto, considerando os estudos apresentados, esse trabalho visa identificar alguns dos aspectos mencionados na publicação dos *data papers* pelos autores filiados às instituições brasileiras.

3 PROCEDIMENTOS METODOLÓGICOS

Para Pando e Almeida (2021, p. 686) o “[...] conhecimento científico possui um valor justamente porque se difere dos demais e se produz de outra forma. Sem tais características perderia a sua essência e, com isso, a utilidade à sociedade”. Ao se analisar a citação desses autores, pode-se observar que o valor do conhecimento científico está associado à sua capacidade de ser verificável, reproduzido e certificado pelos pares.

Sendo assim, os procedimentos que norteiam a elaboração de uma pesquisa devem ser precisos e, com isso entende-se que a “[...] pesquisa é um conjunto de processos sistemáticos, críticos e empíricos aplicados no estudo de um fenômeno” (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013, p.30).

Dessa forma, nesta seção são apresentados os procedimentos adotados para caracterizar a publicação de dados de pesquisa por meio dos *data papers* por parte dos autores filiados às instituições brasileiras.

3.1 CLASSIFICAÇÃO DA PESQUISA

Esta tese teve como característica propor um objetivo **exploratório** e **descritivo**. A característica exploratória teve relação direta com a transversalidade da pesquisa, onde se almeja analisar o que está acontecendo agora sobre a publicação de dados de pesquisa. O caráter descritivo buscou realizar um desenho de um quadro por meio da descrição das características relacionadas nos objetivos específicos, das relações entre variáveis e da compreensão através da literatura, fornecendo um quadro global do objeto de estudo (Gray, 2012; Lerma González, 2009; Prodanov; Freitas, 2013).

Nesse sentido, caracterizar significa reconhecer os elementos que darão personalidade às variáveis do objeto que se deseja conhecer (Volpato, 2019). Para esta pesquisa, as variáveis foram definidas a partir das leituras de Candela *et al.* (2015), Chen, Li e Jiao (2022), García-García, López-Borrull e Peset (2015), Huang, Hawkins e Qiao (2013), Schöpfel *et al.* (2019), Seo e Kim (2020) e Walters (2020), conforme Quadro 7.

Quadro 7 - Variáveis relacionadas aos objetivos específicos

Objetivos específicos	Variáveis	Significado
Apresentar as características temporais e temáticas dos <i>data papers</i>	- Ano de publicação dos <i>data papers</i>	Ano de publicação dos <i>data papers</i> nos periódicos
	- Áreas do conhecimento e tópicos	Classificação de tópicos estabelecida pelo OpenAlex que se divide em domínio (<i>domain</i>), campo (<i>field</i>), subcampo (<i>subfield</i>) e tópico (<i>topic</i>). Nesta pesquisa convencionou-se descrever, respectivamente, grande área, áreas, subáreas e assuntos dos <i>data papers</i>
Descrever a autoria por meio da filiação institucional e países, fontes de financiamento	- Instituições de filiação dos autores	Nome(s) da(s) instituições às quais os autores dos <i>data papers</i> estão filiados
	- Nome do(s) financiador(es)	Nome(s) da(s) instituição(ões) que financiou(aram) a pesquisa conforme metadados das bases de dados
	- País e continentes da instituição de filiação e da agência de financiamento	Países e continentes onde as instituições de filiação dos autores e financiamento estão localizadas
	- Estado e região brasileiros de filiação dos autores e da agência de financiamento	Refere-se à divisão do território brasileiro
Caracterizar os periódicos nos quais os <i>data papers</i> foram publicados	- Nomes dos periódicos	Nomes dos periódicos nos quais os <i>data papers</i> foram publicados
	- Total de <i>data papers</i> indexados	Quantidade total de <i>data papers</i> indexados nas bases Scopus e WoS
	- País	País de localização do periódico
	- Nome do <i>publisher</i>	Nome do editor do periódico
	- Ano de fundação do periódico	Ano de publicação do primeiro número do periódico
	- Ano de início de publicação de <i>data paper</i>	Ano de publicação do primeiro <i>data paper</i> ou documento com nomenclatura similar
	- Indexação no Directory of Open Access Journals (DOAJ)	Verificação da indexação do periódico no DOAJ
	- Valor da APC	Valor da APC conforme fornecido pelo site do periódico
	- Orientações para o depósito dos dados	Orientações voltadas para a disponibilidades dos dados de pesquisa e repositórios a serem utilizados pelos autores

Objetivos específicos	Variáveis	Significado
	- Adesão ao acesso aberto	Indicação do tipo de via de acesso aberto.
	- SCImago Journal Rank (SJR) (2023)	Número médio de citações recebidas num ano em relação ao número de documentos publicados na revista selecionada nos três anos anteriores. Os dados são obtidos a partir da Scopus
	- Journal Impact Factor (JIF), via JCR (2023)	São todas as citações recebidas pelo periódico no ano atual para itens publicados nos dois anos anteriores
	- Quartil	No SJR, os quartis (Q1, Q2, Q3 e Q4) indicam a posição de um periódico em relação aos demais na sua área de estudo, baseado no número de citações ponderadas recebidas por artigo. O Q1 representa os 25% melhores periódicos, o Q2 os 25% seguintes, o Q3 os 25% seguintes e o Q4 os 25% inferiores, segundo o SJR
	- Estrato qualis	Consiste na qualificação da produção intelectual a partir da análise da qualidade dos periódicos
	- Área do conhecimento Capes	Áreas do conhecimento representadas no qualis 2017-2020.
	- Área mãe	É a área do conhecimento que mais concentrou publicações no período de avaliação.
Analisar os repositórios de dados de pesquisa, nos quais os dados utilizados nos <i>data papers</i> foram publicados.	- Tipo de publicação dos conjuntos de dados	Indicação da publicação dos conjuntos de dados nos <i>data papers</i> recuperados
	- Nome do repositório	Nome do repositório onde os dados foram depositados
	- Quantidade de conjunto de dados (<i>datasets</i>)	Quantidade de <i>datasets</i> informadas nos <i>data papers</i>
	- Mantenedor(es)	Nome da instituição mantenedora do repositório, conforme informado no R3data
	- País do repositório	País onde está hospedado o repositório conforme site institucional
	- Acesso	Refere-se à limitação imposta pelos repositórios para acesso ou depósito dos dados
	- Licença	Refere-se às licenças que limitam ou não o manuseio e reuso dos dados
	- Cobrança	Refere-se às taxas cobradas para depósito dos dados, ampliação de tamanho ou adesão
	- Revisão dos dados	Indicação da existência de algum tipo de revisão dos dados por parte dos repositórios

Objetivos específicos	Variáveis	Significado
	- Tipo de repositório	Cobertura do repositório em relação às áreas dos dados de pesquisa a serem depositadas, se disciplinar ou multidisciplinar
	- Áreas do repositório	Áreas de cobertura do repositório, conforme R3data

Fonte: elaborado pela autora a partir de Candela *et al.* (2015), Capes (2023), Chen, Li e Jiao (2022), Clarivate (c2021), Clarivate (c2023); García-García, López-Borrull e Peset (2015), Huang, Hawkins e Qiao (2013), Santos e Nassi-Caló (2020); Schöpfel *et al.* (2019), Scimago (c2017-2025), Seo e Kim (2020) e Walters (2020).

Os **procedimentos** são os passos necessário para levantamento dos dados para a elaboração da pesquisa. Eles permitem traçar um modelo conceitual e operativo, denominado de design, envolvendo diagramação, previsão de análise e interpretação da coleta de dados, considerando o ambiente em que são coletados e as formas de controle das variáveis envolvidas (Prodanov; Freitas, 2013). Considerando a natureza da pesquisa e a classificação do objetivo, os procedimentos utilizados para a coleta dos dados foram a **pesquisa bibliográfica** e a **documental**.

A **pesquisa bibliográfica** tem sido utilizada com frequência em estudos exploratórios ou descritivos, casos em que o objeto de estudo proposto é pouco estudado, tornando difícil a formulação de hipóteses precisas e operacionalizáveis (Lima; Mito, 2007).

A pesquisa bibliográfica, para Stumpf (2012), é um conjunto de procedimentos que visa identificar informações bibliográficas e selecionar documentos pertinentes de forma a evidenciar o entendimento dos autores sobre o tema estudado. Ainda de acordo com a autora a pesquisa bibliográfica subsidia a revisão da literatura que é uma atividade contínua e constante em toda a pesquisa, que perpassa pela definição do problema e/ou objetivo, pela fundamentação teórica e conceitual, pela escolha da metodologia até a análise dos dados para a formulação dos resultados (Stumpf, 2012).

Para a coleta dos dados, a pesquisa bibliográfica se concentrou no uso de bases de dados, que são definidas como suportes informacionais compostos de artigos e trabalhos científicos, elaborados por organizações especializadas, nas diversas áreas do conhecimento. Em geral possuem credibilidade científica, autenticidade e confiabilidade, bem como mecanismos de busca que possibilitam a localização do material bibliográfico e filtros que possibilitam o refinamento dos resultados (Cunha, 2010; Ferreira; Ferro, 1983; Pizzani *et al.*, 2012).

Nas bases de dados foram identificados os *data papers* e seus metadados, fontes principais de dados para o alcance dos objetivos propostos. Além disso, nas bases de dados também foi identificada a literatura que serviu para fundamentar o referencial teórico e conferir significado à análise dos dados, contribuindo para a pesquisadora identificar as variáveis a serem medidas, e como elas se relacionam, a partir de estudos anteriores (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013).

A escolha pelos *data papers* se justifica pois esta foi a forma mais objetiva e exequível para identificar os dados compartilhados pelos brasileiros em repositórios de dados, tendo em vista buscar abarcar o máximo de áreas do conhecimento possíveis para preencher o objetivo da pesquisa. Em testes anteriores, realizados em repositórios de dados abertos (Zenodo, Dryad, R3data), não foi possível filtrar a busca por autores brasileiros. Portanto, as bases de dados em questão oportunizaram a consecução desta etapa de coleta de dados, uma vez que indexam esse tipo de material.

A **pesquisa documental** visa fornecer as fontes documentais necessárias para o levantamento das características do objeto de estudo. A motivação para uso desse método reside na flexibilidade e na possibilidade de ter inúmeros caminhos da pesquisa e poder levar à formulação de interpretações novas ou à modificação de pressupostos iniciais (Cellard, 2012). A pesquisa documental exige um esforço quanto ao reconhecimento das fontes potenciais de informação, e isto, não apenas em função do objeto de pesquisa, mas também em função do questionamento. Por isso, a utilização da pesquisa documental é destacada quando se pode organizar informações que se encontram dispersas, conferindo-lhe uma nova importância como fonte de consulta (Cellard, 2012; Prodanov; Freitas, 2013).

Nesta tese, a **pesquisa documental** teve como objetivo complementar as informações referentes aos periódicos e repositórios identificados, e que devido às suas características não estavam disponíveis nas bases de dados.

Portanto, considerando os procedimentos propostos, as informações identificadas nas pesquisas bibliográfica e documental, foram utilizadas para fornecer maior completude à análise dos objetos relacionados aos *data papers*.

A **análise dos dados** se deu com o uso de **métodos mistos**, o que implica no uso de dados quantitativos e qualitativos integrados para a elaboração de uma discussão conjunta, a fim de realizar inferências como produto de toda a informação

(Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013). Para Gray (2012) o uso dos métodos mistos complementa a análise, isto é, busca-se o esclarecimento dos resultados usando distintos métodos de investigação e análise. A análise dos dados quantitativos foi realizada utilizando a estatística descritiva, que visou resumir e descrever os dados coletados e sintetizá-los a partir de tabelas, gráficos e figuras (Connaway; Powell, 2010). Já a análise qualitativa buscou entender o significado das variáveis previamente definidas, a partir de categorias de análise (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013). Na seção 3.4 e suas subseções serão fornecidos os detalhes sobre a análise dos dados.

3.1.1 Pesquisa bibliográfica a partir do uso da busca sistematizada da literatura

Buscando identificar a produção na literatura mundial sobre o tema “*data paper*” e “*data journal*” foram realizadas buscas sobre a temática nas bases de dados *Library, Information Science & Technology Abstracts with Full Text* (LISTA), Scopus e *Web of Science* (WoS). A primeira foi selecionada por ser uma base da área da Ciência da Informação e as demais foram selecionadas pela cobertura multidisciplinar, idiomática e temporal, além do prestígio dos periódicos ali indexados. A busca foi realizada no dia 23 de junho de 2025, totalizando 284 registros. As estratégias utilizadas em cada base estão descritas no Quadro 8.

Quadro 8 - Estratégias de busca

Base	Estratégia
LISTA	TI (("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos")) OR SU (("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos"))
Scopus	(TITLE (("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos")) OR KEY (("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos"))) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ed") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "cr") OR LIMIT-TO (DOCTYPE , "ch"))
Web of Science	("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos") (Title) or ("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revistas de Datos") (Author Keywords) and Article or Editorial Material or Proceeding Paper or Book Chapters or Review Article (Document Types)

Fonte: elaborado pela autora

As buscas foram realizadas simultaneamente nos campos de títulos e palavras-chaves (ou assunto). Essa opção foi adotada pois o termo "*data paper*" é um termo designado para diferenciar o tipo de artigo adotado nas pesquisas, em contraposição ao uso do termo artigo científico. Por isso seu uso nos resumos tem sido adotado, o que acarreta uma recuperação exacerbada de registros que não tratam do termo como um assunto em si, o qual é o objetivo desta revisão.

Esta busca focou apenas em artigos científicos, anais de eventos, capítulos de livros e editoriais que tratassem do tema, motivo pelo qual os filtros foram utilizados nas bases Scopus e WoS. Na base LISTA, a recuperação se deu apenas por artigo científico, pois é o tipo documental disponível na base de dados. A escolha abrangente pelos tipos de materiais levou em consideração, o caráter pouco explorado da temática.

Adicionalmente, buscando abarcar a literatura não indexada nas bases citadas acima, foram realizadas buscas nas bases: Bielefeld Academic Search Engine (BASE)⁴⁰, Base de Dados em Ciência da Informação (BRAPCI)⁴¹ e Directory of Open Access Journals (DOAJ)⁴². A busca foi realizada no dia 26 de junho de 2025 e seguiu as especificidades e limitações de cada base de dados, conforme Quadro 9.

⁴⁰ Disponível em: <https://www.base-search.net/>

⁴¹ Disponível em: <https://www.brapci.inf.br/home>

⁴² Disponível em: <https://doaj.org/>

Quadro 9 - Estratégias de buscas em bases adicionais.

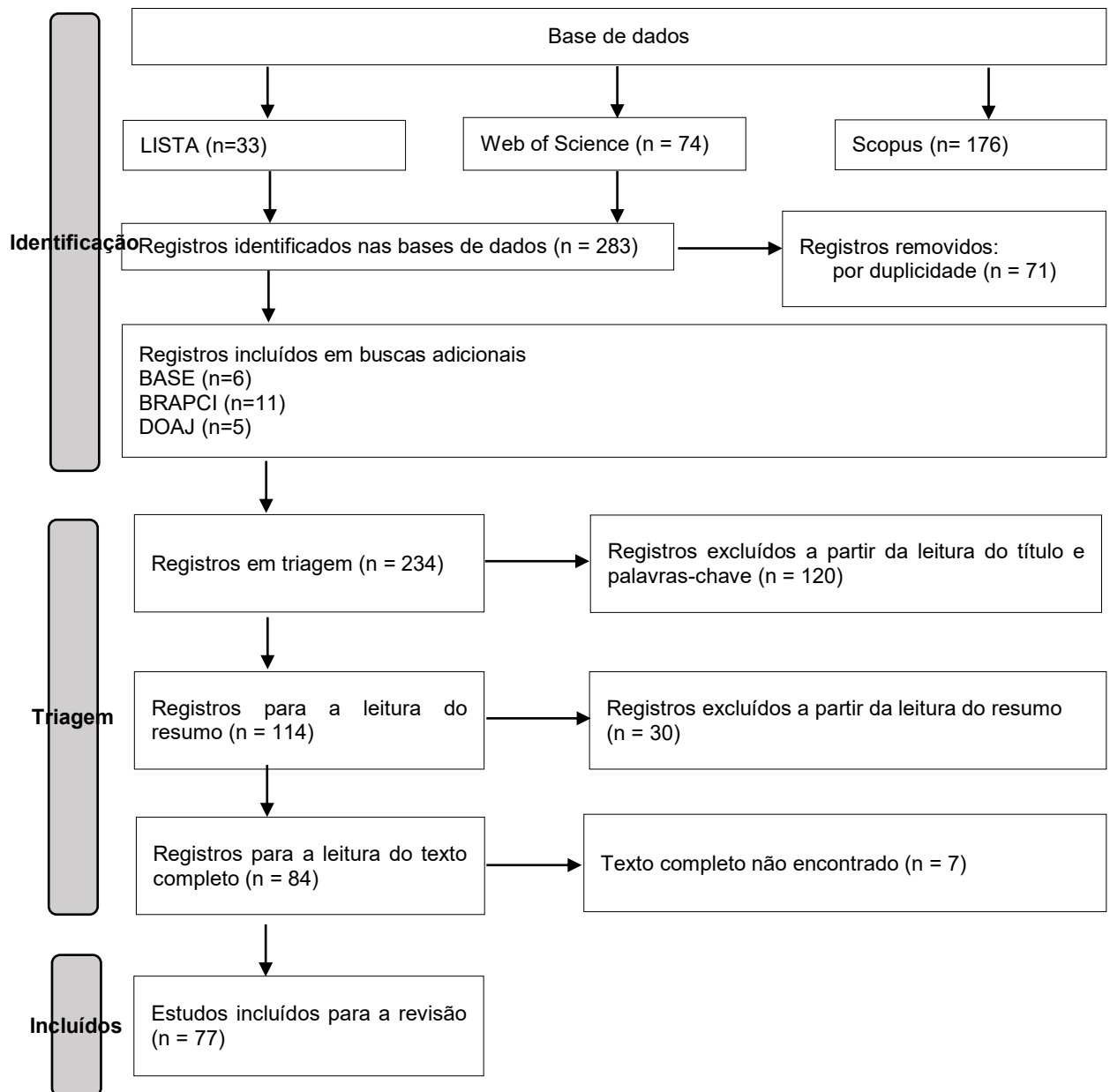
Base	Estratégia	Comentários
BASE	Foram realizadas quatro buscas: tit:("data paper" OR "artigo de dados" OR "artículo de datos") doctype:121 Apenas no título: 172 resultados subj:("data paper" OR "artigo de dados" OR "artículo de datos") doctype:121 Apenas nos assuntos: 197 resultados tit:("data journal" OR "jornal de dados" OR "Revistas de Datos" OR "periódico de datos") doctype:121 Apenas no título: 115 resultados subj:("data journal" OR "jornal de dados" OR "Revistas de Datos" OR "periódico de datos") doctype:121 Apenas nos assuntos: 11 resultados	Essa base recupera muitos documentos duplicados, por isso foi necessário separar as buscas Foram incluídos apenas artigos (<i>"Article contribution"</i>)
BRAPCI	A busca foi realizada separadamente por cada termo da estratégia: "data papers" - 10 resultados "artigo de dados" – 10 resultados "artículo de datos" – 2 resultados "data journal" – 6 resultados "jornal de dados" – 0 resultado "Revistas de Datos" – 1 resultado "periódico de datos" – 2 resultados	Foram incluídos artigos e anais de eventos A maioria dos resultados se repetiu nas buscas
DOAJ	Foram realizadas duas buscas: ("data paper" OR "artigo de dados" OR "artículo de datos" OR "data journal" OR "jornal de dados" OR "Revista de dados") Apenas no título: 26 resultados Apenas nas palavras-chave: 37 resultados	A base não oferece a busca simultânea nos campos título e palavras-chave. Portanto, a busca precisou ser realizada de maneira separada.

Fonte: elaborado pela autora

Embora essas bases possuam limitações quanto à forma de busca, foram encontrados documentos em acesso aberto que não estavam indexados nas bases anteriores. Foram incluídos apenas artigos científicos, anais de eventos e editoriais e excluídos outros documentos tais como: apresentações de slides, conjuntos de dados, textos didáticos ou documentos que não possuíam o texto completo.

A fim de ilustrar o fluxo utilizado para selecionar os documentos foi elaborada a Figura 5.

Figura 5 - Fluxo de identificação de estudos nas bases LISTA, Web of Science e Scopus



Fonte: adaptado de Page *et al.* (2021)

A triagem dos documentos foi realizada a partir da leitura dos títulos e das palavras-chaves para verificar se elas aderiam ao critério principal de inclusão que era selecionar documentos representativos sobre a temática *data paper* e *data journal* e que pudessem contribuir para o referencial teórico, análise e discussão da presente tese. Devido à escassez de documentos sobre a temática o critério utilizado foi amplo, o que culminou para a quantidade de textos incluídos nesta revisão.

Os registros foram extraídos das bases de dados no formato *comma-separated values* (.csv) e a partir do DOI dos registros, um campo em comum entre todas as bases de dados. Foi executada uma função que fez a integração desses registros, realizando a limpeza dos registros duplicados. Após a leitura do título, resumo e palavras-chave, foram selecionados 77 documentos para compor a justificativa, referencial teórico e análises desta pesquisa. Por fim, outros trabalhos que foram publicados após a realização das buscas foram adicionados às análises.

3.2 UNIVERSO DA PESQUISA

O universo da pesquisa abarcou o uso de três bases de dados como fontes de informação, sendo que essa decisão se deu pela complementaridade entre elas (Codina, 2024). Devido às características documentais do objeto analisado, os *data papers*, somente as bases de Scopus e WoS possibilitaram o uso do filtro para esse tipo de documento. Já a seleção da base de dados OpenAlex⁴³ ocorreu pelo fato dela oferecer uma estrutura de tópicos⁴⁴ que permitiram a análise comum das temáticas abordadas nos *data papers* identificados nas duas bases anteriores. Além disso, essa base possui o uso do *Research Organization Registry* (ROR), um registro global de identificadores persistentes abertos para organizações de pesquisa (ROR, [2025?]), que permitiu a descrição e a análise das instituições às quais os autores estavam filiados e das agências de financiamento.

A Scopus e a WoS estão entre as maiores bases de dados referenciais do mundo, possuem ampla cobertura multidisciplinar, temporal, idiomática e uma diversidade de ferramentas para realizar a filtragem e a exportação dos resultados. Além disso, a escolha por essas bases leva em consideração sua representação no sistema de avaliação brasileiro, que, em diferentes áreas, valoriza a publicação em periódicos indexados nelas (CAPES, 2023). São bases cujo acesso é pago, no entanto, nesta pesquisa o acesso foi viabilizado pela UFSC via rede *Virtual Private Network* (VPN) ou Rede Privada Virtual.

⁴³ Disponível em: <https://openalex.org/>

⁴⁴ Os tópicos no OpenAlex são definidos a partir do uso de um sistema automatizado que considera as informações disponíveis sobre o trabalho, incluindo título, resumo, nome de origem (journal) e citações. Os tópicos se referem aos assuntos do documento e em ordem decrescente se subdividem em subcampo, campo e domínio (Priem; Piwowar; Orr, 2022)

O OpenAlex também é uma base referencial, com uma cobertura documental maior do que as anteriores, sendo uma fonte totalmente aberta (com 100% dos dados abertos, API aberta e código-fonte aberto) (Codina, 2024; Priem; Piwowar; Orr, 2022). Seu nome é uma alusão a antiga Biblioteca de Alexandria e sua organização é realizada pela entidade sem fins lucrativos, *OurResearch*. Por isso, seus idealizadores consideram que o OpenAlex tem potencial para melhorar a transparência da avaliação, navegação, representação e descoberta de pesquisas (Borrego, 2024; Priem; Piwowar; Orr, 2022).

Para realizar esta pesquisa foi verificado que o OpenAlex possui uma interface de usuário mais limitada, oferecendo menos recursos do que a Scopus ou a WoS, ou seja, essa base não possui filtro com o tipo documental *data paper*, pois nesta base, são indexados como artigos científicos tradicionais. Assim, nesta pesquisa o OpenAlex teve o papel de fornecer os relatórios com as variáveis explicitadas no Quadro 7 (Codina, 2024).

Em geral, a busca por *data papers* como fontes de dados ocorre a partir de listas de *data journals* compiladas por outros pesquisadores (Jiao; Li; Fang, 2023). No entanto, essa compilação se torna limitada, uma vez que os critérios para a seleção dessas fontes são diversos. Devido a sua contemporaneidade no sistema acadêmico de publicação, os *data journals* nem sempre possuem os critérios de qualidade para estarem indexados nessas bases de dados (Jiao; Li; Fang, 2023).

No caso desta tese, o diferencial, em relação aos outros trabalhos, foi integrar as bases Scopus, OpenAlex e Web of Science, pois estão entre as fontes de dados bibliográficas de grande escala mais comumente usadas em estudos quantitativos de ciência (Jiao; Li; Fang, 2023). Embora o OpenAlex seja mais recente do que as duas anteriores, os números de artigos cobertos pelas bases de dados são semelhantes entre si (Jiao; Li; Fang, 2023), o que tornou possível a coleta dos dados para esta pesquisa.

3.3 COLETA DOS DADOS

A coleta dos dados ocorreu em etapas, tendo como objetivo complementar as informações sobre o objeto estudado. A pesquisa bibliográfica foi iniciada no dia 26 de abril de 2024, com o objetivo de identificar todos os *data papers* publicados por autores filiados às instituições brasileiras e indexados na Scopus e na Web of

Science. No dia 02 de maio de 2024 foi realizada a busca no OpenAlex com o objetivo de extrair os relatórios com as informações sobre *data papers* identificados na primeira coleta. Esse intervalo de tempo foi necessário para que fosse possível extrair os dados de todos os *data papers* identificados na busca anterior, uma vez que a indexação nas bases de dados não ocorre de maneira simultânea.

A busca pelos *data papers* obedeceu às especificidades da Scopus e da WoS. Na tela de busca avançada, foi selecionado o tipo de documento “data paper”, usando os códigos DOCTYPE (dp), na Scopus e “Document type”, na WoS conforme Figuras 6 e 7.

Figura 6 - Página da busca realizada na Scopus

The screenshot shows the Scopus Advanced search page. The 'DOCTYPE(dp)' filter is highlighted with a red box and a red arrow pointing to it. The page shows the search interface with various filters and a search button.

Fonte: Elsevier (c2024a)

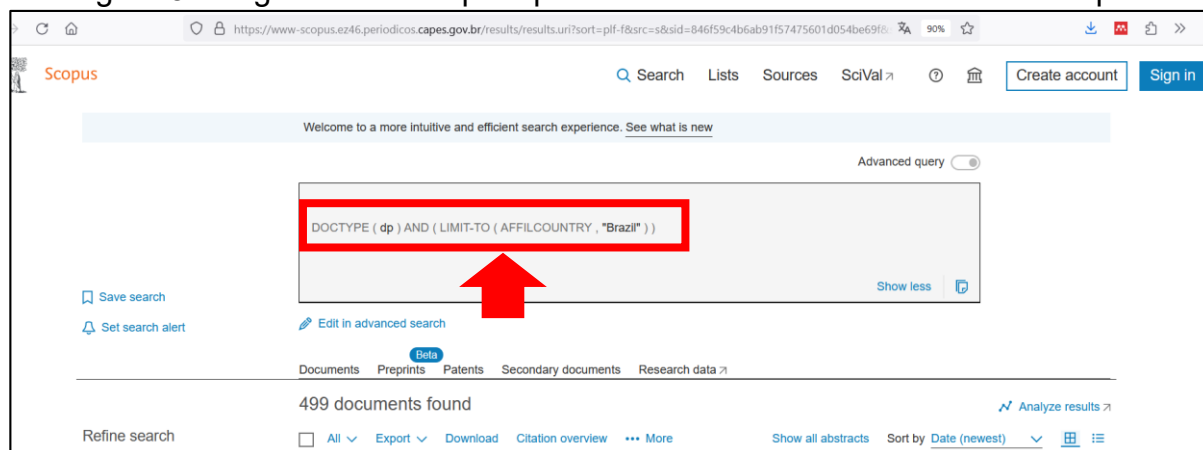
Figura 7 - Página da busca realizada na Web of Science

The screenshot shows the Web of Science Advanced search page. The 'DT=(Data Paper)' filter is highlighted with a red box and a red arrow pointing to it. The page shows the search interface with various filters and a search button.

Fonte: Clarivate (c2024)

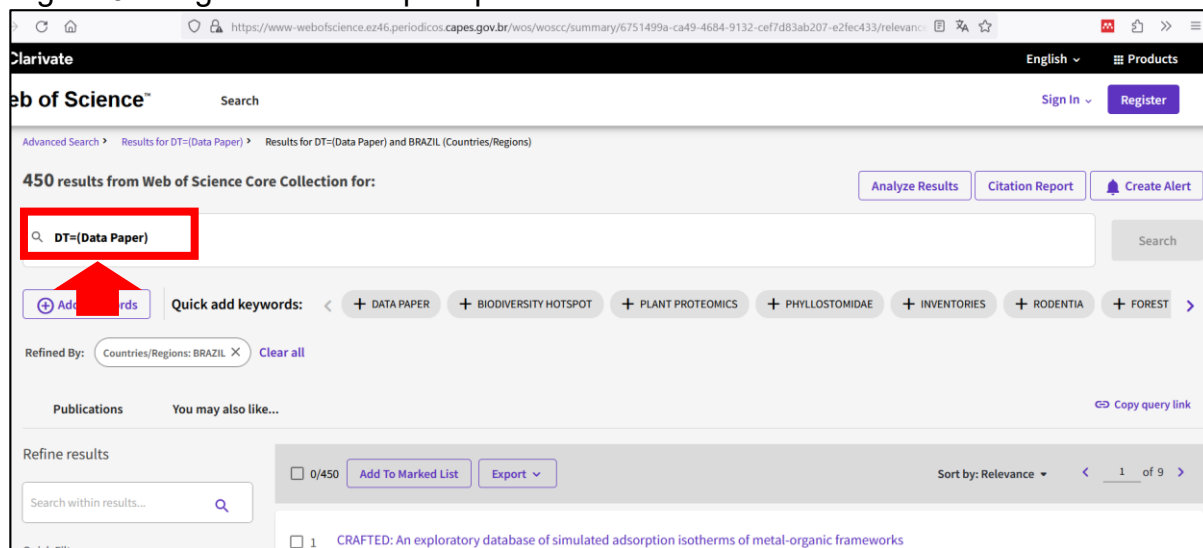
Após a identificação dos *data papers*, os resultados de ambas as bases foram filtrados utilizando a opção “Country/territory” onde foi selecionado “Brazil”, determinando que pelo menos um dos autores dos *data papers* fosse filiado à uma instituição brasileira, conforme Figuras 8 e 9.

Figura 8 - Página da busca por tipo de documento e filtro “Brazil” na Scopus



Fonte: Elsevier (c2024a)

Figura 9 - Página da busca por tipo de documento e filtro “Brazil” na Web of Science



Fonte: Clarivate (c2024)

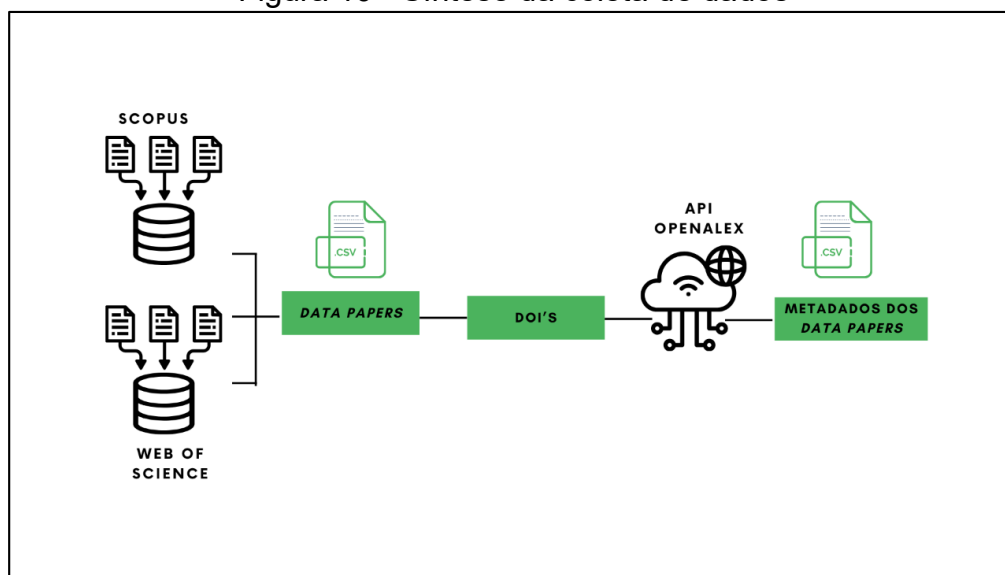
Portanto, para fins de reprodução, as pesquisas realizadas nas bases de dados podem ser sintetizadas pelas seguintes estratégias: DOCTYPE (dp) AND (LIMIT-TO(AFFILCOUNTRY,"Brazil")), na Scopus e DT=(Data Paper) AND CU=(Brazil) na WoS.

Após as buscas nas bases de dados, foi realizada a exportação dos metadados no formato *comma-separated values* (.csv) para o software *Microsoft Office Excel®* 2016, onde foi realizada a limpeza dos registros duplicados e a identificação dos DOIs, dos 634 *data papers* publicados com pelo menos um autor filiado à uma instituição brasileira.

Na etapa seguinte, com os DOIs dos *data papers* já identificados na etapa anterior, foi criado um *script* em linguagem de programação Python para a extração⁴⁵ dos respectivos metadados dos *data papers* na *Application Programming Interface* (API) do OpenAlex⁴⁶. Por fim, obteve-se um relatório em .csv com metadados dos *data papers* que subsidiaram a identificação das variáveis elencadas no Quadro 7.

Todas as etapas realizadas para a coleta dos *data papers*, DOIs e dos metadados estão sintetizadas na Figura 10.

Figura 10 - Síntese da coleta de dados



Fonte: elaborado pela autora

Esse relatório foi complementado pela pesquisa documental, na qual foram identificados como documentos úteis: as diretrizes para os autores e/ou para submissões, páginas dos periódicos, dos repositórios de dados, das instituições financiadoras e de filiação dos autores. No Quadro 10 estão sintetizadas as fontes originárias dos dados de cada variável, bem como a data de coleta de cada uma.

⁴⁵ A elaboração do *script* e a extração dos dados via API foi realizada pelo bibliotecário Fábio Lorensi do Canto, o qual teve contribuição fundamental na coleta dos dados. Muito obrigada!

⁴⁶ Disponível em: <https://docs.openalex.org/how-to-use-the-api/api-overview>

Quadro 10 - Fontes de coleta de cada variável

Objetivos específicos	Variáveis	Fonte “Campo”*/Origem dos dados	Data da coleta dos dados
Apresentar as características temporais e temáticas dos <i>data papers</i>	- Ano de publicação dos <i>data papers</i>	“publication_year”	02/05/2024
	- Áreas do conhecimento e tópicos	“topics”	02/05/2024
Descrever a autoria por meio da filiação institucional e países, fontes de financiamento	- Instituições de filiação dos autores	“authorships”	02/05/2024
	- Nome do(s) financiador(es)	“Grants” no OpenAlex “Funding Orgs” na Web of Science	02/05/2024
	- País e continentes da instituição de filiação dos autores e da agência de financiamento	Informação retirada do campo “Location” do ROR** para cada instituição	23/06/2024
	- Estado e região brasileiros de filiação dos autores e da agência de financiamento	Página da instituição a partir de busca realizada no buscador Google.	23/06/2024
Caracterizar os periódicos nos quais os <i>data papers</i> foram publicados	- nome do periódico	“Locations”	02/05/2024
	- Total de <i>data papers</i> indexados	Busca pelos periódicos na Scopus e Web of Science	26/04/2024
	- País	Informação retirada do campo “country” no site do Scimago*** para cada periódico	08/09/2024
	- Nome do <i>publisher</i>	“Locations”	02/05/2024
	- Ano de fundação do periódico	Lista de todos os periódicos, vide apêndice A*	15/08/2024
	- Ano de início de publicação de <i>data paper</i>	Lista de todos os periódicos, vide apêndice A*	15/08/2024
	- Indexação no DOAJ	Busca em https://doaj.org/	02/09/2024
	- Valor da APC	Taxas de processamento de artigos, Taxas e financiamento, Lista de preços de APCs, Opções de publicação de artigos, Taxas e licenciamento de acesso aberto, vide apêndice A*	14/03/2025

Objetivos específicos	Variáveis	Fonte “Campo”*/Origem dos dados	Data da coleta dos dados
	- Orientações para o depósito dos dados	Política de compartilhamento de dados, política de disponibilidade de dados, política de dados, etc., vide apêndice A*	14/03/2025
	- Adesão ao acesso aberto	“open_access”	02/05/2024
	- SCImago Journal Rank (SJR) (2023)	Informação retirada do campo “SJR” no site do Scimago*** para cada periódico	08/09/2024
	- Quartil	Informação retirada do campo “Quartiles” no site do Scimago*** para cada periódico	08/09/2024
	- Assunto e Categoria		08/09/2024
	- Journal Impact Factor (JIF)	Informação retirada do campo JIF**** para cada periódico	08/09/2024
	- Quartil e categoria	Informação retirada do campo JIF e quartile**** para cada periódico	08/09/2024
	- Estrato qualis	Informação retirada da coluna classificação da plataforma Sucupira*****	25/09/2024
	- Área do conhecimento Capes	Informação retirada da coluna área do conhecimento da plataforma Sucupira*****	25/09/2024
	- Área mãe	Informação retirada da coluna área mãe da plataforma Sucupira*****	25/09/2024
	- Tipo de revisão por pares	Política de revisão por pares, tutorial do revisor, diretrizes para revisores, processos para revisão, diretrizes do editor, Políticas editoriais e padrões dos relatórios, Políticas editoriais e considerações éticas, vide Apêndice A	27/05/2025
Analisar os repositórios de dados de pesquisa, nos quais os dados utilizados nos <i>data papers</i> foram publicados.	- Tipo de publicação dos conjuntos de dados	Próprio <i>data paper</i> , para verificar as seções de onde foram retiradas essas informações vide Apêndice A	18/10/2024
	- Nome do repositório	Próprio <i>data paper</i> , para verificar as seções de onde foram retiradas essas informações vide Apêndice A	18/10/2024
	- Quantidade de conjunto de dados (<i>datasets</i>)	Próprio <i>data paper</i> , para verificar as seções de onde foram retiradas essas informações vide Apêndice A	18/10/2024
	- Mantenedor(es)	Informação retirada do campo “Institutions” do site R3data*****	18/11/2024
	- País do repositório	Informação retirada do campo “Country” do site R3data*****	18/11/2024
	- Tipo de instituição mantenedora	Informação retirada do campo “Type of institution” do site R3data*****	18/11/2024
	- Acesso	Informação retirada do campo “Data access” do site R3data*****	16/05/2025

Objetivos específicos	Variáveis	Fonte “Campo”*/Origem dos dados	Data da coleta dos dados
	- Licença	Informação retirada do campo “Data licences” do site R3data*****	16/05/2025
	- Cobrança	Busca realizada em cada repositório utilizado, vide Apêndice B.	18/11/2024
	- Revisão dos dados	Busca realizada em cada repositório utilizado, vide Apêndice B.	18/11/2024
	- Tipo de repositório	Informação retirada do campo “Repository type” do site R3data*****	16/05/2025
	- Áreas do repositório	Informação retirada do campo “Subjects” do site R3data*****	16/05/2025

Fonte: elaborado pela autora

Notas: *No Apêndice A estão especificados todos os documentos utilizados de acordo com cada periódico identificado na pesquisa.

** Acesso ao site ROR: <https://ror.org/>

*** Acesso ao site Scimago: <https://www.scimagojr.com/>

**** Busca realizada por cada periódico no Journal Citation Reports, acesso via VPN na base Web of Science.

***** Busca realizada na plataforma Sucupira, por cada periódico, no quadriênio 2017-2020.

***** Acesso ao site R3Data: <https://www.re3data.org/>

O intervalo no qual ocorreu a coleta dos dados da pesquisa foi de abril de 2024 a maio de 2025. Essa variação se deve à análise de cada objetivo, onde a partir do tratamento e análise de cada variável as informações foram sendo complementadas pela pesquisa documental.

Para testar a viabilidade da pesquisa foi realizado um pré-teste com 35 *data papers* publicados por autores vinculados à UFSC, o qual foi apresentado e aprovado na banca de qualificação⁴⁷, realizada no dia 05 de março de 2024.

A síntese dos resultados apresentados na banca de qualificação originou um trabalho completo, o qual foi submetido e aceito no 24º Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação⁴⁸, realizado em Vitória, Espírito Santo, em 2024.

Por fim, os dados brutos, ou seja, os relatórios dos metadados utilizados para subsidiar a presente tese estão disponíveis no repositório Zenodo, e podem ser acessados a partir do link <https://zenodo.org/records/19192846>.

⁴⁷ A banca da qualificação foi composta pelos membros externos, os doutores Elizabete Cristina de Souza de Aguiar Monteiro e Washington Luís Ribeiro de Carvalho Segundo e pelo membro interno professora doutora Rosângela Schwarz Rodrigues.

⁴⁸ O acesso ao trabalho completo está disponível neste link: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2410>.

3.4 TRATAMENTO E ANÁLISE DOS DADOS

O tratamento visa organizar e limpar os dados buscando padronizá-los para a realização da análise quantitativa e qualitativa. Para tanto, todos metadados referentes aos *data papers* foram sintetizados em planilhas do *Microsoft Office Excel*® 2016.

Em seguida, as planilhas contendo os metadados foram exportados para o *Open Refine*⁴⁹, cujo objetivo foi realizar a limpeza de dados por meio de uma interface web (Open Refine's, [2023]). Neste aplicativo foi realizada a uniformização de letras maiúsculas e minúsculas e a padronização dos nomes das instituições para a forma completa, caso estivessem abreviados. No caso de instituições brasileiras foram realizadas as traduções para o idioma português quando necessário. Além disso, foi realizada a clusterização das instituições, com o objetivo unir as instituições cujas grafias eram diferentes embora fossem as mesmas e reunir as unidades subordinadas à uma unidade maior.

Na análise se busca dar sentido aos dados por meio das variáveis descritas no Quadro 7, almejando atender aos objetivos específicos propostos e assim responder ao problema de pesquisa.

Conforme mencionado na subseção 3.1, a análise dos dados se deu com uso do método misto, que pressupõe o uso dos enfoques quantitativos e qualitativos. No primeiro enfoque, os dados foram analisados com o uso da estatística descritiva que visou caracterizar os dados para cada uma de suas variáveis e depois estabelecer relações entre elas (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013). De maneira geral, as variáveis foram medidas usando a **distribuição de frequência** que corresponde à organização dos dados de acordo com a quantidade de ocorrências dos diferentes resultados, a **média**, que é a soma dos valores dividida pelo número de valores observados e a **porcentagem** que se refere à divisão do número total de ocorrências pela frequência de cada categoria (Barbetta, 2015). No enfoque qualitativo, os dados foram descritos em categorias, atreladas às variáveis e analisados conforme a literatura.

Ainda em relação à análise, foi utilizada a ferramenta Microsoft Copilot⁵⁰, uma ferramenta baseada em inteligência artificial, cujo objetivo foi ajudar na identificação

⁴⁹ Disponível em: <https://openrefine.org/>

⁵⁰ Disponível em: <https://copilot.microsoft.com/>

de padrões e similaridades, a partir do *upload* das tabelas e quadros já elaborados pela autora, além de ajudar na melhoria da redação de alguns parágrafos da tese.

Para a apresentação dos dados foram utilizados elementos visuais, tais como: quadros e tabelas, recursos visuais condizentes com o caráter descritivo do objetivo, conforme Lerma González (2009), elaborados a partir do próprio Excel ou do aplicativo Flourish⁵¹, cujo objetivo foi gerar as visualizações e compartilhar os dados.

Nas subseções seguintes, a partir das categorias relacionadas aos objetivos específicos, são detalhadas as análises de maneira específica.

3.4.1 Dados sobre as características temporais e temáticas dos *data papers*

O objetivo dessa categoria foi descrever a análise das características gerais dos *data papers* como a distribuição dos anos de publicação, temáticas abordadas e grandes áreas e áreas do conhecimento, conforme o modelo de Tópicos do OpenAlex, que é um modelo baseado no conjunto de dados do *Centre for Science and Technology Studies*, da Leiden University, que atribui tópicos às publicações na medida que são inseridas na base de dados.

Os tópicos são um sistema de rótulos que pode ser usado para descrever de forma precisa e sucinta o **assunto de um artigo**, garantindo uma relação entre os tópicos e os níveis superiores de campos e uma precisão na descrição dos artigos (Portenoy, 2024). Eles são gerados usando um sistema automatizado que considera as informações disponíveis do título, resumo, nome do periódico e citações. A esses campos o modelo gera pontuações que culminam na criação do assunto (Portenoy, 2024).

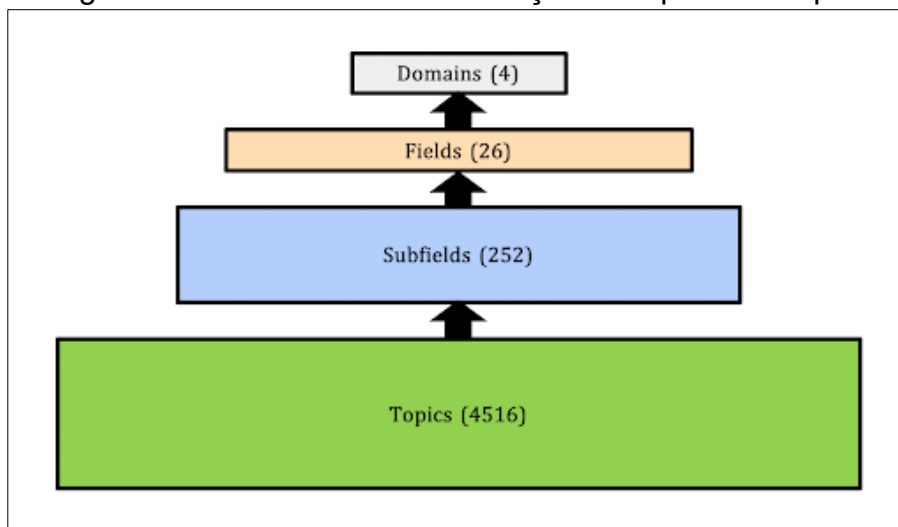
Esses tópicos são a base da estrutura de classificação no OpenAlex, que é formada ainda pelos subcampos (*subfields*), campos (*fields*) e domínios (*domains*), sendo que toda essa estrutura é baseada na *All Science Journal Classification*⁵² da Scopus. Logo, cada tópico pertencerá a um subcampo, que pertencerá a um campo,

⁵¹ Disponível em: <https://flourish.studio/>

⁵² O ASJC **classifica os periódicos** indexados na Scopus a partir de uma análise realizada por especialistas do objetivo, do escopo e do conteúdo que o periódico publica (Elsevier, c2024a). No anexo A está descrito o esquema de classificação de assuntos da ASJC.

que pertencerá a um domínio (Portenoy, 2024). A Figura 11, ilustra a estrutura mencionada.

Figura 11 - Estrutura de classificação de tópicos no OpenAlex



Fonte: Portenoy (2024)

Nesta pesquisa para facilitar a análise dos dados, convencionou-se utilizar **assunto** para substituir o **termo tópico** (*topic*), **tema** para se referir aos **subcampos** (*subfields*), **áreas** para se referir aos **campos** (*fields*) e **grandes áreas** para se referir aos **domínios** (*domains*).

A análise quantitativa nessa categoria se deu por meio do uso da frequência (F) para somar as publicações de *data papers* por ano, comparar seu crescimento nas grandes áreas do conhecimento e para quantificar os assuntos mais explorados nos *data papers*.

3.4.2 Características das instituições de filiação dos autores e das agências de financiamento

Esta categoria teve como objetivo identificar as características das instituições brasileiras e das instituições internacionais que colaboram com o Brasil, as quais os autores informaram estarem filiados no momento da publicação dos *data papers*. Além disso, nessa categoria foram analisadas as agências de financiamento que foram informadas pelos autores como fomentadoras das pesquisas.

A partir do ROR, que é um registro global para atribuir identificadores persistentes para organizações e assim individualizá-las (ROR, [2025?]), foram

identificados os países (“Location”) e as tipologias das instituições (“Organization types”), tanto de filiação dos autores quanto das agências de financiamento. Tendo em vista contribuir com as análises, a partir dos países foram identificados os continentes e os estados e regiões brasileiras. As tipologias identificadas foram classificadas como aquelas vinculadas à educação, órgãos do governo, *Non-Governmental Organization* (ONGs) ou Organização Não Governamental, empresas privadas, instituições de pesquisa e de saúde, arquivos e outros.

Para analisar os financiamentos foram utilizadas duas classificações: quanto à sua nacionalidade e quanto à sua governança.

Quanto à nacionalidade a análise foi atrelada à localização das agências de financiamento, sendo a classificação baseada na proposta de Alvarez e Caregnato (2021) que classificou as agências como nacional, internacional ou mista. Nesta pesquisa, trocamos termo misto por combinado, pois o primeiro termo já foi utilizado para caracterizar os periódicos que publicam artigos científicos e *data papers* (Candela *et al.*, 2015). Alteramos o termo internacional, para estrangeiro.

Com isso, atreladas à sua localização, as agências nacionais são aquelas localizadas no Brasil, as estrangeiras são aquelas localizadas em outros países e as combinadas representam as agências localizadas no Brasil e no exterior. Portanto, os financiamentos foram considerados nacionais quando foi mencionado apenas uma ou mais agências localizadas no país, estrangeiros quando mencionada uma ou mais agências do exterior e combinados quando mencionada uma ou mais agências nacionais e estrangeiras.

Quanto à sua governança, a classificação se baseou na proposta de Bogado, Hilário e Grácio (2024) na qual os financiamentos foram classificados como públicos, quando advindos por entes públicos, privados, quando oriundos de empresas privadas e financiamentos por ONGs quando oriundos de organizações não-governamentais. Quando mencionadas diferentes agências essa classificação se mesclou, como por exemplo: financiamento público e privado, público e ONG, privado e ONG, dentre outros.

A análise quantitativa dos países que colaboram com o Brasil na publicação dos *data papers* se deu a partir da frequência de instituições em cada país. A análise qualitativa foi relacionada ao conceito de capacidade científica e tecnológica dos países e continentes, conforme classificação utilizada por Chen, Li e Jiao (2022).

A análise quantitativa da frequência das instituições se deu a partir do número de vezes em que ela foi mencionada por cada autor e a porcentagem de cada instituição se deu a partir da divisão do total de menções da instituição pelo total de menções de filiação de todas as instituições identificadas. A análise qualitativa visou associar as instituições que mais se destacaram às práticas voltadas para a abertura da Ciência, bem como uma análise da sua tipologia.

A caracterização das instituições brasileiras buscou identificar os estados e regiões que mais as concentraram, onde a partir da sua frequência pôde-se conhecer aqueles que mais possuíam instituições e a tipologia predominante. Além disso, ao somar as frequências dos estados, foi possível identificar os mesmos padrões por região.

Para identificar as características dos financiamentos, inicialmente foi verificado quantos *data papers* receberam algum tipo de financiamento e quantos não receberam. Para tanto foram identificados aqueles que informaram as agências de financiamento e aqueles que não informaram. Portanto, para essa pesquisa a menção do nome do financiador significa que houve o recebimento de financiamento. Nos demais casos, onde não foi identificado o financiador a partir dos registros extraídos das bases de dados, foi considerado não houve o recebimento de financiamento.

Em seguida, a partir da frequência de *data papers* que receberam financiamento, foi realizada a associação entre a tipologia da agência (financiamento Público, ONG ou Privado) e a nacionalidade das agências de financiamentos (financiamento nacional, internacional ou combinado).

Para identificar o índice de financiamento por grande área e área, conforme Ünal e Akbulut (2025), foi utilizada a proporção, a qual foi calculada a partir do número total de *data papers* dividido pela quantidade de *data papers* que receberam financiamento, ou seja, que informou o nome da agência. A análise qualitativa se deu a partir das áreas onde cada tipologia de financiamento é mais comum, bem como, quais as temáticas que mais receberam atenção das agências financiadoras.

A distribuição das agências de financiamento pelo mundo se deu a partir da soma da quantidade de agências por país, bem como a apresentação da sua tipologia predominante. As agências mais mencionadas foram identificadas a partir da quantidade de vezes que foram mencionadas pelo autor em cada *data paper* e, assim, num mesmo *data paper* a agência pôde ser mencionada mais de uma vez. A análise

qualitativa se deu pelos países que mais financiam pesquisas pelo mundo e a partir dos mandatos das agências de financiamento que mais acumularam menções.

3.4.3 Características dos periódicos

Esta categoria teve como o objetivo identificar as características associadas aos periódicos em que os *data papers* foram publicados.

Os periódicos foram classificados a partir de Candela *et al.* (2015) que os dividiu entre puros e mistos, sendo que os primeiros são aqueles voltados exclusivamente para a publicação de *data papers* e o segundo aqueles que publicam, dentre os *data papers*, artigos resultantes de pesquisas e outros tipos de documentos. No entanto, a nomenclatura “puro”, proposta por Candela *et al.* (2015) não reflete a realidade, pois esses periódicos publicam outros tipos de documentos e não apenas *data papers*. Logo, nesta pesquisa a classificação “puro” foi substituída por *data journal* (Silveira; Dias; Neubert, 2024).

Para a análise dos periódicos, na perspectiva dos *data papers*, foram utilizadas as categorias propostas por Piwowar *et al.*, (2018), na qual, os periódicos:

- a) ouro: são aqueles cujo acesso aos artigos é aberto para leitura, mediante o pagamento de APC pelo autor. Geralmente o periódico ouro está indexado no DOAJ;
- b) verde: são aqueles cujo acesso ao artigo é pago, mas o editor permite que uma cópia seja disponibilizada gratuitamente em um repositório de acesso aberto;
- c) híbrido: são aqueles em os artigos são publicados em uma revista de assinatura, mas são livres para leitura sob uma licença aberta, mediante o pagamento de APC pelos autores
- d) bronze: são aqueles cujo o acesso ao artigo é gratuito para leitura na página do editor, mas não possui uma licença claramente identificável e
- e) fechado: são aqueles cujo o acesso os artigos são realizados mediante assinatura pagas por consórcios ou bibliotecas.

A classificação da tipologia das revisões foi baseada em Santos e Nassi-Caló (2020) onde a “simples cego” é aquela que o revisor tem informações sobre a identificação do autor, mas o autor não sabe quem é o revisor, a “duplo cego” é aquela em que as identidades do revisor e autor são mantidas anônimas e a “aberta” é aquela

que a identidade de autores e/ou revisores é conhecida durante o processo de avaliação.

A análise quantitativa dos periódicos se deu a partir da frequência dos *data papers* publicados por cada um. Já a análise qualitativa se baseou na quantidade de *data papers* por título de periódico, ano de criação e ano publicação e na tipologia dos *publishers* se eram comerciais, ou de associações ou universidades, países de localização desses *publishers*, nas vias de acesso aberto adotadas, bem nas discussões que giram em torno do predomínio de *publishers* comerciais no âmbito da Ciência Aberta.

A estimativa de gastos com APCs foi realizada a partir da multiplicação do valor individual do APC pela quantidade de *data papers* publicados pelo periódico. Cálculo semelhante foi realizado para estimar os gastos com APC por grandes áreas e áreas do conhecimento.

A média do APC por área foi realizada a partir da divisão do valor do APC acumulado pela área pelo total de *data papers* publicados nela. O valor total por grande área é o resultado da soma dos valores dos APCs de cada área do conhecimento individualmente e a média da grande área, é a divisão do valor total do APC acumulado pelo total de *data papers* vinculados às áreas do conhecimento. A partir desses dados, a análise qualitativa se deu com inferências sobre os gastos nas áreas do conhecimento bem como a média cobrada em relação aos artigos tradicionais.

Os indicadores de avaliação e citação da área, SJR, JIF e Qualis, foram analisados de forma qualitativa, atrelando a tipologia do periódico ao impacto nas suas áreas do conhecimento.

O processo de revisão por pares dos periódicos foi analisado a partir da quantidade de periódicos que informaram a tipologia adotada. A análise qualitativa foi realizada a partir do tipo de revisão e dos critérios informados nas suas políticas pelos periódicos mistos e *data journals*.

Os critérios foram categorizados a partir da adaptação do trabalho de Seo e Kim (2020). No entanto, houve uma divisão dos critérios entre aqueles referidos aos dados e ao *data paper* em relação ao estudo adotado, pois alguns *data journals* informaram em suas políticas critérios específicos para esses itens. No Quadro 11 estão descritos os critérios bem como seus significados.

Quadro 11 - Critérios de avaliação dos *data papers* e *datasets* informados pelos periódicos.

Critérios relacionados aos dados	Significado
Metadados	Foram incluídos os critérios referentes à descrição dos dados. Por exemplo: uso de padrões disciplinares, precisão dos metadados, período, latitude/longitude
Adequação metodológica	Na descrição os dados fornecem informações sobre a metodologia da coleta e análises. Por exemplo: ferramentas usadas, scripts, fonte dos dados
Formato dos dados acessível	Refere-se ao formato dados, por exemplo: se foram usados formatos proprietários ou não proprietários
Valores dos dados	Foram incluídos os critérios relacionados a identificação dos dados no arquivo. Por exemplo: cabeçalhos de coluna, descrições em um arquivo de texto leia-me
Utilidade e originalidade	Refere-se à reutilização dos dados, por exemplo: se existe potencial para reutilização em larga escala dentro do campo de interesse
Identificador	Refere-se à exigência de identificadores, tanto para os dados quanto outros elementos associados (versionamento, artigos relacionados)
Adequação do repositório	Refere-se ao depósito de repositórios aceitos pela comunidade (em termos de assunto, sustentabilidade, natureza dos dados)
Licença	Foram incluídos critérios referentes às licenças exigidas. Por exemplo: exigem preferencialmente licenças abertas
Critérios relacionados aos <i>data papers</i>	Significado
Contextualização dos dados	Refere-se à forma como os dados foram coletados. Por exemplo: onde, como, quando
Descrição precisa da metodologia	Foram incluídos os critérios referentes a metodologia para coleta e análise dos dados. Por exemplo: métodos de calibração, protocolos, medidas de controle, se permite a reprodução por terceiros
Fontes de erros	Refere-se a possíveis erros e limitações. Por exemplo o <i>data paper</i> , informa limitações, fontes de erros, algo que impeça a replicação
Referências utilizadas	Refere-se às referências utilizadas no texto
Reutilização	Refere-se a utilidade dos dados: os dados são originais, são úteis para reutilização
Disponibilidade dos dados	Refere-se à forma como os autores disponibilizaram os dados brutos. Por exemplo: repositórios de dados, materiais suplementares
Repositórios	Refere-se às informações sobre os repositórios, por exemplo, adequação, link de acesso aos dados, identificador, nome do repositório

Fonte: adaptado de Seo e Kim (2020)

Por fim, para a validação desta seção junto à área, foi submetida e aceita na 16ª Conferência Lusófona de Ciência Aberta (CONFOA), em 2025, na cidade de Goiânia, uma comunicação com resultados parciais sobre a estimativa de custos para a publicação dos *data papers*⁵³.

⁵³ O trabalho pode ser acessado a partir desse link: <https://recal.rcaap.pt/recal/article/view/31>.

3.4.4 Características dos repositórios

Esta categoria teve como objetivo identificar as características dos repositórios de dados onde dados de pesquisa aos quais os *data papers* se referiram foram depositados.

Inicialmente, foi realizada a identificação da disponibilidade dos dados nos *data papers*, cuja classificação foi adaptada a partir de Jiao, Li e Fang (2022), na qual os dados foram disponibilizados:

- a) em repositórios⁵⁴: essa caracterização se deu a partir da menção do nome do repositório ou do(s) link(s) de acessos aos dados de pesquisa;
- b) como material suplementar: quando os arquivos contendo os conjuntos de dados eram inseridos numa seção específica do *data paper*, como por exemplo: apêndice, informação de suporte, registro e disponibilidade dos dados etc.;
- c) em páginas pessoais ou institucionais: a identificação foi realizada a partir da visita ao link disponibilizado;
- d) não disponibilizado: quando não foi mencionado nenhum conjunto de dados ao longo do *data paper* ou quando o *data paper* informava que o conjunto estava adicionado ao longo do texto e
- e) como material suplementar e repositório: quando mencionava as duas formas.

A partir da identificação dos repositórios citados nos *data papers*, as tipologias das instituições que os mantêm foram identificadas no “Type of institution” do R3data, sendo elas: comerciais, não comerciais e mantidos em parcerias entre instituições comerciais e não-comerciais (Strecker *et al.*, 2023).

Ainda no R3data, a tipologia dos repositórios foi classificada, sendo os multidisciplinares aqueles que abarcam várias disciplinas e disciplinares aqueles que abarcam um conjunto pré-determinado de disciplinas (Strecker *et al.*, 2023).

⁵⁴ Para essa pesquisa foram considerados como repositórios de dados, todos os repositórios que armazenam e preservam os componentes que subsidiam o processo de coleta, tratamento e análise dos dados, como por exemplo, algoritmos, ferramentas, códigos e workflows. Por exemplo: Github um repositório adequado para depositar o código computacional, Zenodo repositório voltado para armazenar de maneira conjunta os dados e os arquivos associados (Aventurier; Alencar, 2016; Sayão; Sales, 2022).

As áreas de cobertura do R3data seguem o esquema de metadados chamado *Metadata Schema for the Description of Research Data Repositories*, versão 4.0 (R3data Coref, 2021). Esse esquema é elaborado pela organização de financiamento a pesquisa da Alemanha, Deutsche Forschungsgemeinschaft (DFG), e fornece os metadados necessários para descrever os repositórios indexados no R3data. No que condiz às áreas temáticas, ele define as quatro principais: Ciências Humanas e Sociais, Ciências da Vida, Ciências Naturais e Ciências da Engenharia, cada qual com suas subdivisões, que são utilizadas quando há a necessidade de descrever em profundidade a temática do repositório (R3data Coref, 2021; Strecker *et al.*, 2023). Nesta análise, foram utilizadas apenas as quatro classes principais.

A análise quantitativa da disponibilidade dos dados utilizou a unidade *data paper* para quantificar seu total em cada categoria, bem como a porcentagem. Em seguida, foi identificada a frequência de periódicos que aderiu a cada categoria também. A análise qualitativa se deu a partir das datas e da consolidação pelos periódicos da tipologia adotada.

Em relação às características dos repositórios, foi identificado em cada *data paper* o quantitativo de conjuntos de dados (*dataset*) mencionado, com isso, a partir do total de *datasets* mencionados pelo total de *data papers* que disponibilizaram os dados em repositórios obteve-se a média de *dataset* por *data paper*.

As discussões em torno das características dos repositórios foram realizadas a partir da frequência de *datasets* depositados e, com isso, obteve-se a média de *datasets* por repositórios. A análise qualitativa delas se deu a partir da concentração de *datasets* em mantenedores comerciais, na localização dos repositórios, na cobrança para submissão e revisão dos dados e nas licenças adotadas.

Foi realizada uma análise da cobertura temática e da tipologia dos repositórios e para isso, foi realizada uma associação entre a quantidade de repositórios mencionados nos *data papers*, e *datasets* depositados. Assim, discutiu-se o quanto a área interfere na quantidade de repositórios e de *datasets* utilizados. A tipologia dos repositórios considerou a média em cada um deles e foi analisada também a influência da área e da consolidação do repositório na decisão dos autores em selecionar o repositório de dados mais adequado.

A análise da cobertura temática relacionou os repositórios às áreas da R3Data e a quantidade de conjuntos de dados depositados neles e os relacionou ao comportamento da área. Por fim, foi analisada a adesão dos autores dos *data papers*

às políticas de depósito de dados. Para verificar essa adesão, foi realizado um cálculo que considerou o total de *datasets* de cada periódico dividido pela quantidade total de depositados em cada tipo de repositório.

Por fim, para a validação desta seção junto à área, foi submetido um trabalho ao ENANCIB⁵⁵, realizado em novembro de 2025, no Rio de Janeiro, o qual foi premiado como o melhor trabalho completo no GT-7 Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação. Também foi aprovado e publicado um trabalho na Conferência Internacional BIREDIAL-ISTEC⁵⁶, realizada no mesmo ano, em Brasília.

3.5 LIMITAÇÕES DA PESQUISA

As limitações são inerentes à prática de pesquisa. Neste caso, a opção pelo uso dos *data papers* para a identificação do objetivo proposto fornece uma noção parcial das práticas de publicação de dados de pesquisa, uma vez, que foram selecionados apenas aqueles que estão indexados nas bases de dados Scopus, OpenAlex e Web of Science.

Durante o desenvolvimento da pesquisa foram observadas limitações em relação à indexação dos *data papers* nas bases de dados, em relação à sua cobertura temporal e, também, foram identificadas limitações referentes aos metadados extraídos das bases utilizadas.

Em relação à indexação da tipologia documental “*data papers*” utilizada pelas bases de dados, a classificação varia de acordo com a base de dados que ora os classifica como artigos genericamente, no caso do OpenAlex, e ora os classifica como artigos e *data papers* de maneira conjunta, no caso da Web of Science e Scopus. Como por exemplo, foi observado que no periódico *Encontros Bibli* não foi recuperado nenhum documento classificado com esta tipologia “*data paper*”, mas este periódico publica *data papers*, logo estes deveriam ser recuperados, uma vez que preenchem os requisitos desta pesquisa. Esse tipo de limitação foi identificado por Candela *et al.* (2015) e Jiao, Li e Fang (2023). Para Candela *et al.* (2015), os editores começaram a

⁵⁵ O trabalho pode ser acessado a partir desse link:

<https://enancib.ancib.org/index.php/enancib/xxvenancib/paper/viewFile/3173/2236>

⁵⁶ O trabalho pode ser acessado a partir desse link: <https://doi.org/10.22477/xiv.biredial.390>

atribuir diversos nomes aos *data papers*, o que pode ter influenciado na indexação desse tipo de documento nas bases de dados.

Houve também uma limitação temporal entre a publicação do *data paper* pelo periódico e sua indexação na base de dados, o que gerou diferenças na quantidade de *data papers* publicados pelo periódico e a quantidade indexada (Jiao; Li; Fang, 2023). Como no caso do periódico *Biodiversity Data Journal* (Pensoft Publishers), no qual foram encontradas discrepâncias entre o total de *data papers* publicados e o total de *data papers indexados*, onde na página do periódico foi possível recuperar 427 *data papers*, na Scopus 21 e na Web of Science 370⁵⁷.

Em termos do ano indexado, Jiao, Li e Fang (2023) explicam que isso ocorre porque algumas publicações foram indexadas por suas primeiras datas publicadas em vez de suas datas formais publicadas. Além disso, as diferenças na adesão da indexação desse tipo de documento pelas bases de dados, também podem explicar essas diferenças na recuperação dos documentos.

Nos relatórios fornecidos pelas bases de dados, foi observado que não foi possível identificar o tipo de apoio de cada órgão de financiamento nas pesquisas. Dessa forma, não foi possível identificar se os financiamentos eram voltados aos pesquisadores, como o fornecimento de bolsas, se eram voltados para o desenvolvimento das pesquisas, ou ainda se contemplavam os dois casos. Para aqueles registros onde não havia indicação de financiamento, foi verificado em cada *data paper*, se havia alguma menção nos agradecimentos e/ou em notas de rodapé.

Ainda nos relatórios, os dados de filiação (especificamente coluna “Affiliations”), em várias linhas foi verificado que a quantidade de autores não coadunava com a quantidade de filiações. Portanto, não foi possível determinar se os autores possuíam apenas uma filiação, se elas foram omitidas ou se foram agrupadas no processo de indexação das bases de dados.

Não foram realizadas análises nos conjuntos de dados em si, e assim, não foi possível precisar nessa pesquisa, se os dados estavam disponíveis de fato e se os conjuntos atendiam aos princípios FAIR, por exemplo.

⁵⁷ Busca realizada em 28 de dezembro de 2023.

4 RESULTADOS E DISCUSSÕES

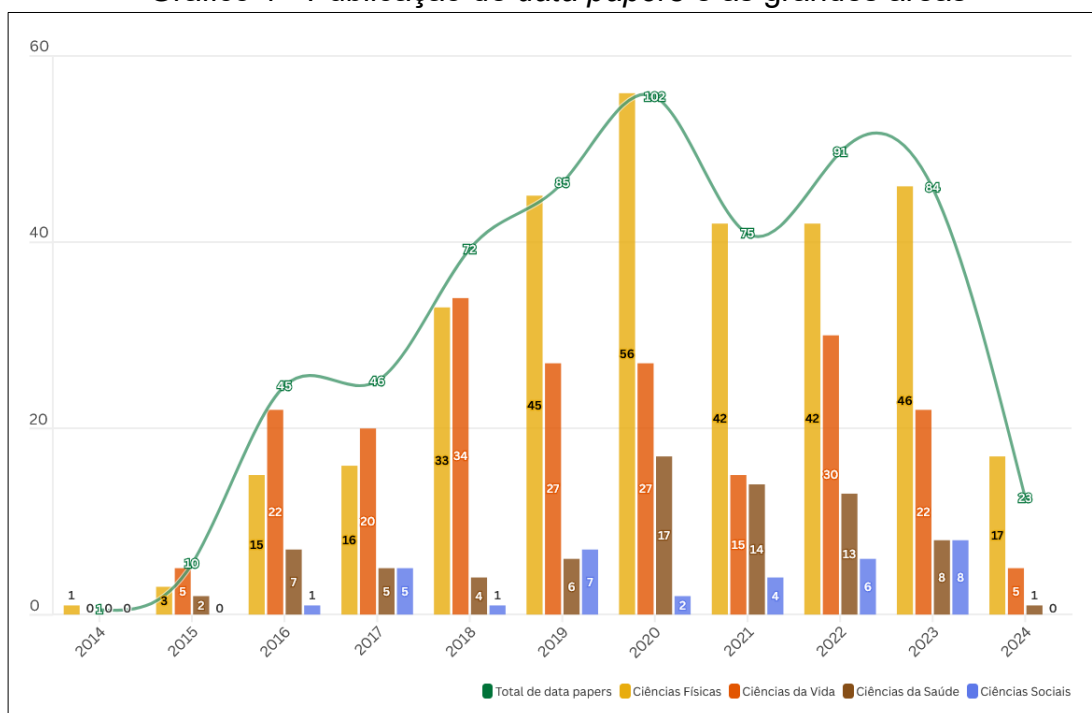
Nesta seção são apresentados os resultados associados aos objetivos específicos, e realizadas as discussões cotejadas com a literatura.

4.1 CARACTERÍSTICAS TEMPORAIS E TEMÁTICAS DOS *DATA PAPERS*

Esta seção visa caracterizar os aspectos temporais da publicação dos *data papers* pelos autores filiados às instituições brasileiras a partir do ano de publicação e os aspectos temáticos a partir da identificação das grandes áreas e áreas do conhecimento, bem como os assuntos abordados.

A coleta dos dados, realizada em abril de 2024, identificou 634 *data papers*. Em maio de 2021, realizando a busca seguindo os mesmos parâmetros desta pesquisa, as autoras Araujo, Príncipe e Alencar (2021) identificaram 366 *data papers*. Com isso, observa-se um aumento de cerca de 57,72% (268) ao longo de três anos.

No Gráfico 1 é possível observar a evolução temporal na publicação de *data papers* conforme as grandes áreas do conhecimento, no qual as barras representam a quantidade de publicações por grande área e a linha refere-se à quantidade total de *data papers* publicados no ano. Cabe ressaltar que a redução na publicação de *data papers* no ano de 2024 está relacionada ao período da coleta que foi até o mês de abril e o *delay* na atualização entre a publicação do *data papers* e sua indexação nas bases de dados.

Gráfico 1 - Publicação de *data papers* e as grandes áreas

Fonte: elaborado pela autora.

Nota: a figura e os dados estão disponíveis em:

<https://public.flourish.studio/visualisation/25833891/>

De maneira geral, observa-se um aumento progressivo na produção dos *data papers* a partir do ano de 2014, com um *data paper* publicado até o ano de 2020, com 102 *data papers* publicados, crescimento também identificado por Araujo, Príncipe e Alencar (2021).

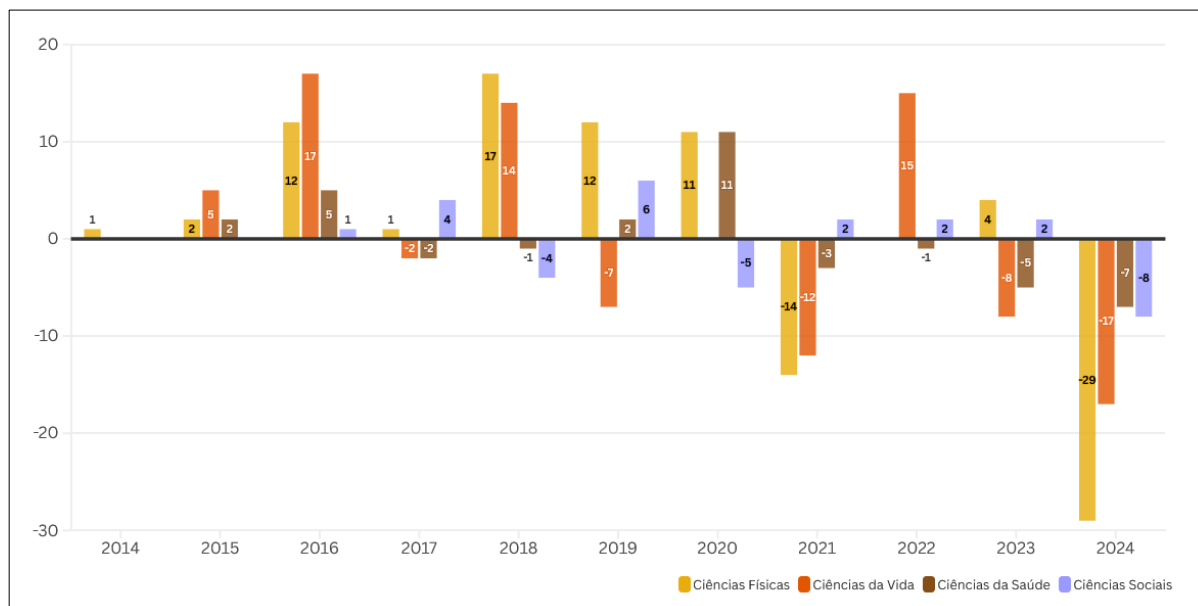
Em outro trabalho que pesquisou o crescimento de *data papers* em nível mundial, Giner-Miguel, Gómez e Cabot (2025) observaram um aumento na produção de *data papers* entre os anos de 2015 e 2023 e um decréscimo em 2021, possivelmente motivado pelo efeito da COVID-19 na produção científica geral. Este decréscimo também foi observado na presente pesquisa, que culminou para uma diminuição de -26,47% (-27) em relação à produção de *data papers* no ano de 2020.

A publicação dos *data papers* é liderada pela grande área das Ciências Físicas com 49,84% (316), com picos de publicação em 2020 (56) e 2023 (46), seguida da grande área das Ciências da Vida que concentrou 32,65% (207) dos *data papers* identificados, com picos de publicação em 2018 (34) e 2022 (30), as Ciências da Saúde que concentraram 12,15% (77) dos *data papers*, com picos em 2020 (17) e 2021 (14) e a grande área das Ciências Sociais que concentrou 5,36% (34) do total de *data papers* publicados, com picos de publicação nos anos 2019 (7) e 2023 (8).

A ascensão da publicação dos *data papers* pelos autores filiados às instituições brasileiras, pode estar atrelada à tendência mundial que vem sendo influenciada pelas práticas preconizadas pela Ciência Aberta, em particular as exigências das agências de financiamento e dos periódicos científicos para o compartilhamento dos dados de pesquisa (Araujo; Príncipe; Alencar, 2021; Kar; Rath, 2025). Também pode estar ligada ao aumento da criação de títulos de *data journals* entre as diferentes áreas do conhecimento e de títulos multidisciplinares, tais como o *Data in Brief* (Elsevier) e o *Scientific Data*, ambos criados em 2014, que vem atraindo a atenção dos estudiosos como um novo local para compartilhamento e publicação de dados (Jiao; Li; Fang, 2023).

Em relação ao crescimento da publicação de *data papers* nas grandes áreas é possível verificar no Gráfico 2 que houve uma oscilação. No gráfico, cada barra representa o saldo de crescimento ou queda de publicações por ano. Valores positivos indicam aumento e negativos indicam redução na publicação.

Gráfico 2 - Variação no crescimento dos *data papers* nas grandes áreas do conhecimento



Fonte: elaborado pela autora.

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/27964036/>

A grande área das Ciências Físicas, que foi a primeira a registrar a publicação de um *data paper*, em 2014, teve um crescimento ininterrupto ao longo dos anos, e um aumento na publicação de *data papers*, em 2016 de 400% (12) em relação a 2015

(2). Em 2021, foi observado um decréscimo significativo de -25% (-14) na publicação desses documentos.

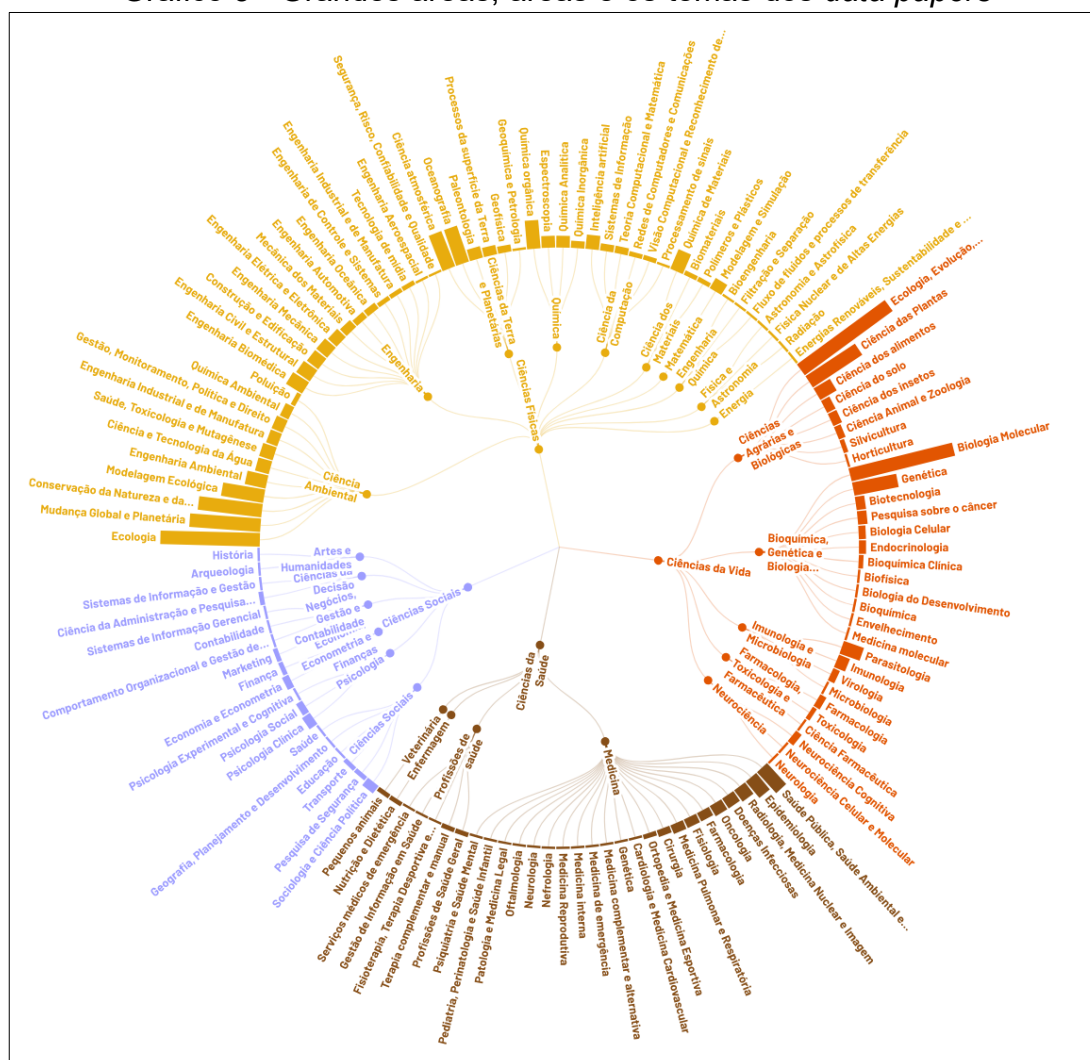
A grande área das Ciências da Vida teve um crescimento semelhante às Ciências Físicas na publicação de *data papers*, com 340% (17) no ano de 2016 e uma redução de 44,44% (-12), em 2021.

As Ciências da Saúde foi a grande área que teve mais variação na publicação de *data papers* ao longo dos anos. Assim como as demais, em 2016, teve seu maior crescimento, de 250% (5) em relação a 2015, e em 2023 teve uma queda de 38,46% (-5) em relação ao ano anterior.

Esse comportamento dessas áreas pode estar associado ao impacto das políticas referentes à disponibilidade dos dados de pesquisa das agências de financiamento (Kar; Rath, 2025; Pinto, 2022), às políticas de periódicos voltadas para o compartilhamento e disponibilidade dos dados (por exemplo, *Plos One*), aos avanços tecnológicos e à ênfase crescente na Ciência Aberta (Kar; Rath, 2025).

Por fim, a área das Ciências Sociais, teve seu primeiro *data paper* publicado em 2016, tendo um crescimento constante ao longo do tempo. Teve aumento de 600% (6) em 2019 em relação ao ano anterior e, esse aumento também foi observado por Kar e Rath (2025) no mesmo período ao mapear a publicação de *data papers* pelas Ciências Sociais. As autoras identificaram que a produção de *data papers* na área ganhou força a partir de 2015 e atribuem o crescimento constante aos mandatos de compartilhamento de dados, avanços dos repositórios digitais, ampla adoção de políticas de acesso aberto e o crescente reconhecimento acadêmico dos *data papers* na área (Kar; Rath, 2025). Esse movimento nas Ciências Sociais, de acordo com Farina *et al.* (2025), está relacionado à ascensão dos métodos computacionais e das humanidades digitais, transformando a maneira como textos antigos, artefatos e fenômenos culturais podem ser analisados, interpretados e compartilhados.

Tendo como objetivo apresentar as grandes áreas, as áreas e os temas identificados nos *data papers* interligados foi elaborado o Gráfico 3. Nos nós centrais estão as grandes áreas e as áreas que se ramificam para a parte externa do círculo onde estão os temas. Os tamanhos das colunas ao lado de cada um representa a quantidade de *data papers* publicados. Detalhes referentes aos quantitativos de cada área estão disponíveis no Apêndice C.

Gráfico 3 - Grandes áreas, áreas e os temas dos *data papers*

Fonte: elaborada pela autora

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/26031291/>

Pode-se observar que a grande área das Ciências Físicas, que mais concentrou a publicação de *data papers*, também foi a que mais concentrou áreas vinculadas, sendo dez no total, bem como a maior variedade de temas estudados. A área das Ciências Ambientais foi a que mais concentrou a publicação de *data papers*, com 24,29% (154), sendo que os temas que envolvem a ecologia com 6,62% (42), mudança global e planetária com 4,73% (30) e a conservação da natureza e da paisagem com 4,26% (27) os que mais se destacaram dentre todos os *data papers* publicados. Portanto, estes foram identificados como os temas de pesquisa mais latentes que originaram os *data papers* publicados pelos autores filiados às instituições brasileiras.

Na grande área da Ciências da Vida, foram identificadas cinco áreas vinculadas. As que mais se destacaram foram as Ciências Agrárias e Biológicas com

14,67% (93) e a Bioquímica, Genética e Biologia Molecular com 13,41% (85) dos *data papers* publicados. Os temas com maior concentração de *data papers* publicados foram a ecologia, evolução, comportamento e sistemática, com 7,26% (46) e a biologia molecular com 7,10% (45).

Kosmopoulos e Schöpfel (2025) argumentam que o montante investido na coleta de dados e a necessidade de tornar os dados amplamente acessíveis contribuiu para o aumento de artigos científicos sobre dados em biologia desde a década de 1990, o que levou pesquisadores internacionais a reuni-los em plataformas online para que todos pudessem acessá-los sem precisar investir grandes somas adicionais. Essa prática pode ajudar a justificar o destaque das áreas citadas anteriormente.

Na grande área das Ciências da Saúde, foram identificadas quatro áreas. A Medicina foi a que mais concentrou a publicação de *data papers* com 10,41% (66), sendo também a área onde ocorreu a maior diversidade dos temas abordados nos *data papers*. O tema que teve maior concentração foi a Saúde Pública, Saúde Ambiental e Ocupacional, com 1,89% (12) dos *data papers* associados.

Na grande área da Ciências Sociais foram identificadas seis áreas, sendo a área homônima aquela que mais concentrou a publicação de *data papers*, 1,74% (11), e os temas que envolvem a Sociologia e a Ciência Política, os que tiveram maior concentração de *data papers* publicados, 0,63% (4).

Cabe ressaltar que a adoção de *data papers* nas Ciências Sociais fica atrás das outras áreas devido à variabilidade das normas disciplinares, complexidades éticas, preocupações com a propriedade intelectual, ausência de diretrizes padronizadas pelos periódicos, adesão inconsistente aos princípios FAIR, e limitação das bases de dados comerciais na indexação de assuntos cobertos pelas áreas das Ciências Sociais e Humanas (Kar; Rath, 2025; Le Fournier; Schöpfel, 2025), além da própria adesão mais tardia a esse tipo de publicação.

Além dos temas, foram identificados também os assuntos, com maior quantidade de *data papers* publicados pelos autores filiados às instituições brasileiras, os quais estão arrolados na Tabela 1. Nessa tabela também é realizada uma comparação com os dez assuntos identificados em âmbito mundial, conforme Giner-Miguel, Gómez e Cabot (2025). O ranqueamento foi realizado pela quantidade de *data papers* vinculados ao assunto e, por isso na coluna assunto existe mais de dez, pois houve coincidência na quantidade de *data papers* publicados.

Tabela 1 - Assuntos com maior frequência nos *data papers*

N	Assuntos identificados nos <i>data papers</i>					10 assuntos identificados por Giner-Miguel; Gómez; Cabot (2025)
	Grande área	Área	Assunto	F data paper	%	Assunto
1	Ciências da Vida	Ciências Agrárias e Biológicas	Impacto do Declínio de Polinizadores nos Ecossistemas e na Agricultura	29	4,57	Sequenciamento de RNA para análise genética
2	Ciências Físicas	Ciência Ambiental	Modelagem da Distribuição de Espécies e Impactos das Mudanças Climáticas	18	2,84	Imagem Médica
3	Ciências da Vida	Bioquímica, Genética e Biologia Molecular	Análise de Dados de Sequenciamento de RNA	12	1,89	Química e física
4	Ciências Físicas	Ciências da Terra e Planetárias	Mudanças Climáticas e Paleoclimatologia	11	1,74	Propriedades do material
5	Ciências Físicas	Ciência Ambiental	Conservação da Biodiversidade e Gestão de Ecossistemas	10	1,58	Análise climática e hidrológica
6	Ciências Físicas	Ciência Ambiental	Biogeografia e Conservação dos Peixes de Água Doce Neotropicais	9	1,42	Conservação biológica
7	Ciências Físicas	Ciências da Terra e Planetárias	Biogeoquímica Marinha e Dinâmica dos Ecossistemas	8	1,26	Agricultura e terras de cultivo
8	Ciências da Vida	Ciências Agrárias e Biológicas	Ecologia e Comportamento de Morcegos	7	1,10	Desenvolvimento de medicamentos
	Ciências Físicas	Ciência Ambiental	Sensoriamento Remoto no Monitoramento da Vegetação e na Fenologia	7	1,10	
9	Ciências Físicas	Ciência Ambiental	Declínio e Extinção Global de Anfíbios	6	0,95	Análise Geológica e Sísmica
	Ciências Físicas	Ciência Ambiental	Ecologia da Vida Selvagem e Biologia da Conservação	6	0,95	
	Ciências Físicas	Ciência Ambiental	Estimativa da Biomassa Florestal e dos Estoques de Carbono	6	0,95	
	Ciências da Vida	Ciências Agrárias e Biológicas	Tecnologias de Agricultura de Precisão	6	0,95	
10	Ciências da Saúde	Medicina	Carga Global da Leishmaniose: Incidência e Tratamento	5	0,79	Reconhecimento de movimento humano
	Ciências da Saúde	Medicina	Epidemiologia e Tratamento da Doença de Chagas	5	0,79	
	Ciências da Vida	Bioquímica, Genética e Biologia Molecular	Evolução do Veneno de Cobra e Toxinologia	5	0,79	
	Ciências da Saúde	Medicina	Impacto Global das Doenças Arbovirais	5	0,79	
	Ciências da Vida	Bioquímica, Genética e Biologia Molecular	Insights Genômicos sobre Insetos Sociais e Simbiose	5	0,79	
	Ciências Físicas	Matemática	Modelagem da Dinâmica da Pandemia de COVID-19	5	0,79	
	Ciências Físicas	Ciência Ambiental	Pesquisa sobre Mudanças Climáticas e Variabilidade	5	0,79	
Total				170	26,81	
Outros 4 ou menos ocorrências				464	73,19	
Total geral				634	100	

Fonte: elaborado pela autora

Notas: F *data paper*: refere-se à quantidade de *data papers* que mencionaram o assunto%: representa a porcentagem de *data papers* que mencionaram o assunto

A análise dos assuntos abordados revela uma grande diversidade. Prova disso é que em 73,19% (464) dos casos, os assuntos foram tratados em quarto ou menos *data papers*. Dentre os assuntos com maior ocorrência, oito estão relacionados à área da Ciência Ambiental, representando 10,57% (67) do total de *data papers*, seguido de três assuntos relacionados às Ciências Agrárias e Biológicas com 6,62% (42) e três assuntos da Bioquímica, Genética e Biologia Molecular, com 3,47% (22) do total de *data papers*.

De maneira geral, é possível observar na coluna assunto que aqueles mais publicados pelos autores filiados às instituições brasileiras estão associados à biodiversidade, mudanças climáticas e conservação, os quais dominam a lista, tais como o impacto do declínio de polinizadores nos ecossistemas e na agricultura, tecnologias voltadas a agriculturas, conservação e sensoriamento remoto.

O destaque para os assuntos atrelados às áreas da Ciência Ambiental pelos autores filiados às instituições brasileiras pode estar relacionado às características ambientais do país, uma vez que em função de sua extensão territorial, diversidade geográfica e climática, apresenta vasta diversidade biológica, o que amplifica o seu papel na produção, armazenamento e disseminação de dados sobre biodiversidade (Silva *et al.*, 2024).

Assuntos interdisciplinares relacionados à genética e à ecologia também obtiveram destaque, tais como: sequenciamento de RNA e insights genômicos sobre insetos sociais e simbiose.

As Ciências da Vida tiveram três assuntos atrelados à área Medicina que correspondem às doenças arbovirais, ou seja, aquelas transmitidas por picadas de insetos ou mosquitos, como a doença de Chagas e a Leishmaniose e com grande potencial epidemiológico.

Ao se realizar uma comparação com os assuntos abordados pelos *data papers* em nível mundial, publicados no trabalho de Giner-Miguel, Gómez e Cabot (2025), é possível verificar que três assuntos identificados nesta pesquisa se aproximam daqueles publicados por autores filiados às instituições brasileiras. São eles: sequenciamento de RNA para análise genética, análise climática e hidrológica e conservação biológica. Embora essa comparação possa ter como limitação a tradução dos termos, ela tende a ser um indicativo de que os assuntos dos *data papers* brasileiros seguem uma agenda de interesses mundial.

4.1.1 Considerações da seção

Ao analisar as características temporais dos *data papers*, é possível observar que sua produção está em ascensão desde 2014, quando identificado o primeiro *data paper* publicado por um autor filiado à uma instituição brasileira. Em cerca de três anos, entre 2021 e 2024, houve um aumento de cerca de 58% na sua publicação. Esse aumento pode estar associado ao aumento em nível mundial, que em 2021 teve um decréscimo, possivelmente influenciado pelo efeito da COVID-19 na produção científica geral.

Ao se associar o crescimento da publicação de *data papers* pelos autores filiados às instituições brasileiras às grandes áreas, houve uma oscilação ao longo dos anos. As grandes áreas das Ciências Físicas e Ciências da Vida tiveram crescimento consistente ao longo dos anos, e um aumento em 2016, sendo que a primeira teve um crescimento de 400% e a segunda 340% e ambas tiveram uma redução em 2021. As Ciências da Saúde foi a grande área que teve mais variação na publicação de *data papers* ao longo dos anos, assim como as demais, em 2016, teve seu maior crescimento, e em 2023 teve uma queda de 38,46% em relação ao ano anterior. A grande área das Ciências Sociais teve seu primeiro *data paper* publicado em 2016, tendo um crescimento constante ao longo do tempo, com um aumento 600% em 2020 em relação ao ano anterior, sendo a área com uma adesão mais tardia à publicação dos *data papers*.

O comportamento dessas áreas pode estar associado ao impacto das políticas referentes à disponibilidade dos dados de pesquisa das agências de financiamento e dos periódicos, além dos avanços tecnológicos e ênfase crescente na Ciência Aberta (Kar; Rath, 2025; Pinto, 2022).

Ao analisar as características temáticas dos *data papers* foi possível observar uma diversidade de temas. As Ciências Físicas foi a que mais concentrou a publicação de *data papers* e, também a que mais concentrou áreas vinculadas, dez no total, e por sua vez a que teve maior variedade de temas estudados. A área Ciências Ambientais foi a que mais concentrou a publicação de *data papers*, sendo que os temas envolvem a ecologia, as mudanças globais e planetárias e a conservação da natureza e da paisagem.

Nas Ciências da Vida foram identificadas cinco áreas vinculadas. As que mais se destacaram foram as Ciências Agrárias e Biológicas e temas de maior relevância foram a ecologia, evolução, comportamento e sistemática e a biologia molecular.

Nas Ciências da Saúde foram identificadas quatro áreas, sendo a Medicina a que mais concentrou a publicação de *data papers* e o tema com maior concentração foi Saúde Pública, Saúde Ambiental e Ocupacional.

Na grande área da Ciências Sociais foram identificadas seis áreas, sendo a área homônima aquela que mais concentrou a publicação de *data papers* e os temas que envolvem a Sociologia e Ciência Políticas com maior destaque.

Dentre os assuntos mais publicados pelos autores filiados às instituições brasileiras estão aqueles associados à biodiversidade, mudanças climáticas e conservação. Assuntos interdisciplinares relacionados à genética e ecologia também obtiveram destaque, tais como sequenciamento de RNA e *insights* genômicos sobre insetos sociais e simbiose. Na área da Medicina, os assuntos relacionados às doenças arbovirais, ou seja, aquelas transmitidas por picadas de insetos ou mosquitos, como a doença de Chagas e a Leishmaniose e com grande potencial epidemiológico, foram os que tiveram mais *data papers* publicados. Por fim, os assuntos dos *data papers* brasileiros seguem uma agenda de interesse mundial.

4.2 CARACTERÍSTICAS DA FILIAÇÃO DOS AUTORES DOS DATA PAPERS E DOS FINANCIAMENTOS

Esta seção tem como objetivo identificar as características de filiação dos autores, sejam aqueles filiados às instituições brasileiras e ou aqueles filiados às instituições estrangeiras, e dos financiamentos mencionados nos *data papers*. Nas características de filiação dos autores foram identificadas as instituições, sua tipologia e localização. Dentre as características dos financiamentos, foram mapeadas as agências, sua tipologia, localização e as áreas vinculadas aos *data papers*.

Foram identificadas 2601 instituições diferentes e em 2,72% (71) dos casos não foi possível identificar as localizações⁵⁸. Com isso a análise das características foi realizada a partir das 2530 instituições cujos países puderam ser localizados.

⁵⁸ A impossibilidade de localizar a instituição se deve à perda de dados no relatório extraído do OpenAlex. Por exemplo: em alguns casos foram recuperados apenas partes dos nomes das instituições, instituições com nomes homônimos, instituições cuja localização não estava disponível no buscador Google.

Essas instituições estavam distribuídas entre 113 países. Aqueles que concentraram mais de 100 instituições diferentes, foram Estados Unidos com 16,13% (408), Brasil com 15,30% (387), França com 5,77% (146), Alemanha com 5,10% (129), Grã-Bretanha com 4,51% (114) e Espanha com 3,99% (101).

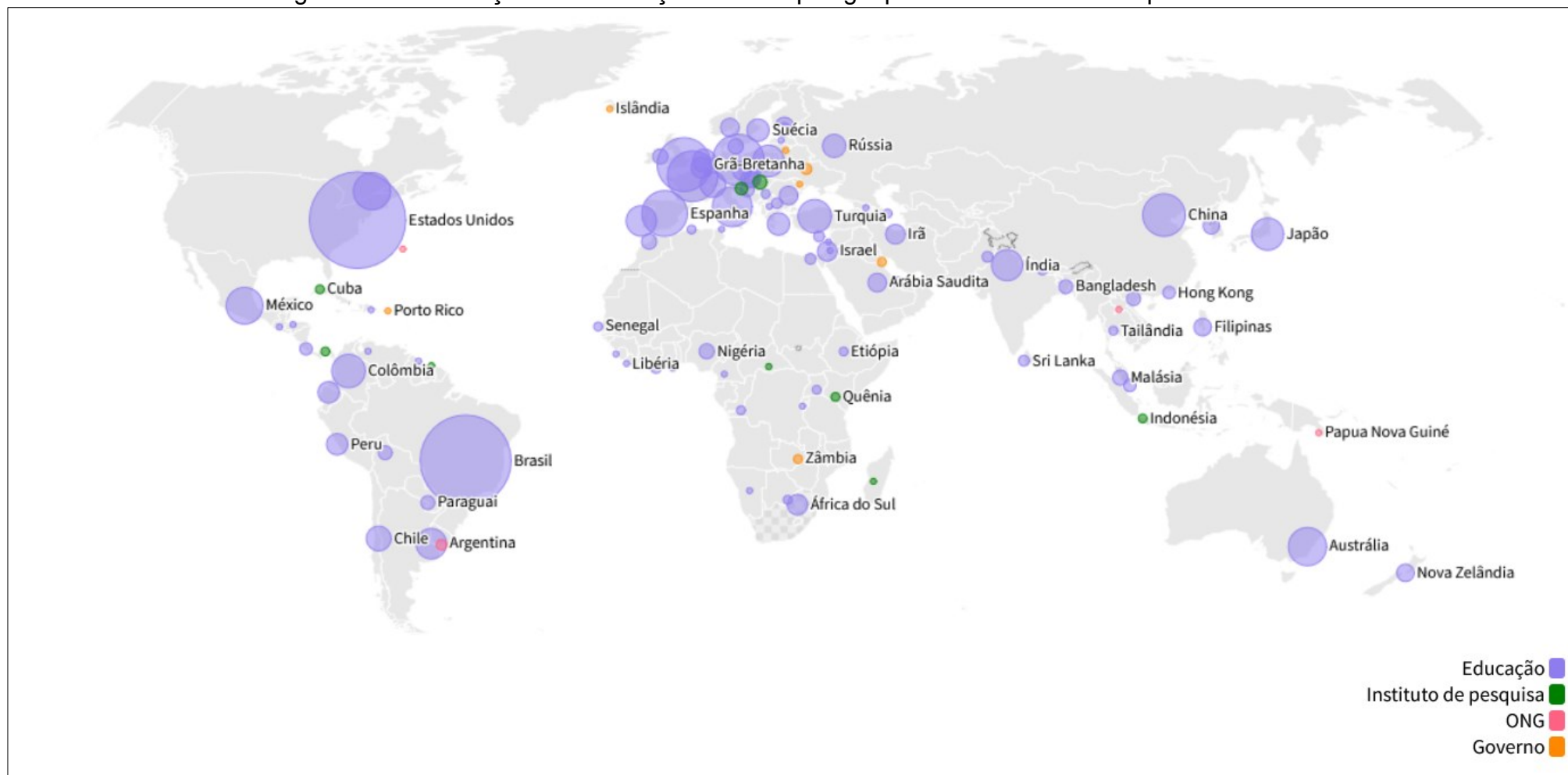
Chen, Li e Jiao (2022) ao pesquisar sobre a geografia da colaboração na publicação de *data papers*, identificaram que os países que mais produzem são os Estados Unidos, China, Itália, Alemanha e Grã-Bretanha. Kar e Rath (2025) ao analisar os *data papers* da grande área das Ciências Sociais, também verificaram que os cinco países que mais publicam *data papers*, são Estados Unidos, China, Alemanha, Reino Unido e Itália.

Nesse sentido, pode-se verificar que esses resultados se assemelham ao que foi identificado nesta pesquisa, com a liderança dos Estados Unidos na publicação de *data papers* e uma concentração de instituições localizadas em países europeus. A diferença entre os achados nos trabalhos de Chen, Li e Jiao (2022) e Kar e Rath (2025) e os desta pesquisa é que, embora a China tenha obtido uma posição de destaque no âmbito mundial, a relação com o Brasil se deu em apenas 2,57% (65) dos *data papers* publicados.

Portanto, em relação aos países com mais instituições, a colaboração internacional do Brasil na produção dos *data papers* se deu com quatro países com maior capacidade de Ciência e Tecnologia, conforme classificação utilizada por Chen, Li e Jiao (2022), que são Estados Unidos, França, Alemanha e Grã-Bretanha. Esses resultados indicam que em relação à quantidade de instituições, o Brasil tende a colaborar com países com maior capacidade científica e tecnológica no que condiz à produção dos *data papers*.

A Figura 12 tem como objetivo ilustrar a distribuição das instituições identificadas e a tipologia da instituição predominante que está destacada por cor. O tamanho dos círculos está relacionado à quantidade de instituições no país. A lista com todos os países, bem como a tipologia das instituições, está disponível no Apêndice D.

Figura 12 - Distribuição das instituições e sua tipologia predominante em cada país



Fonte: elaborada pela autora.

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/21083515/>

Entre os continentes, a maior colaboração institucional se deu entre os países do continente americano, com 43,71% (1106) das instituições identificadas. Dessas, a distribuição se concentra na América do Sul com 22,53% (570), na América do Norte com 20,51% (519) e na América Central, com 0,67% (17).

No continente americano há de se destacar a colaboração realizada entre os países que compõem a América Latina. Foram identificadas 635 instituições diferentes, o que representa 25,1% do total (2530), às quais estavam distribuídas em 16 países. Em ordem decrescente, os países que mais concentraram foram: Brasil com 60,94% (387) das instituições, Argentina com 8,5% (54), México com 8,35% (53), Colômbia com 6,3% (40), Peru com 3,31% (21), Chile com 3,15% (20), Equador com 2,20% (14), Paraguai com 1,73% (11), Bolívia com 1,42% (9), Uruguai com 1,26% (8), Costa Rica e Panamá com 0,79% (5) cada, Cuba com 0,47% (3), Guatemala e Honduras com 0,31% (2) cada e Venezuela com 0,16% (1). Por fim, os países da América Latina que não possuíram nenhuma instituição filiada foram El Salvador, Haiti, Nicarágua e República Dominicana.

Esses dados apontam para uma tendência de colaborações entre instituições localizadas na América Latina, embora essa tendência seja desigual quando analisados os países separadamente. Essa desigualdade pode estar relacionada às barreiras estruturais, econômicas ou políticas, à baixa internacionalização da ciência nesses países, bem como à criação de políticas ou iniciativas voltadas para a Ciência Aberta. Em relação às barreiras econômicas Santin e Caregnato (2019) identificaram que dentre os países da América Latina, o nível médio de investimentos dos países da região em Ciência, Tecnologia e Inovação atinge pouco mais de 0,6% do Produto Interno Bruto nos últimos anos e reforçam uma característica marcante da ciência regional: o baixo investimento, baseado prioritariamente em recursos públicos.

O continente europeu teve 36,09% (913) das colaborações institucionais, com destaque para a França com 5,77% (146), Alemanha com 5,10% (129), Reino Unido com 4,51% (114), Espanha com 3,99% (101) e Itália com 2,85% (72). Juntos esses países contemplam mais da metade das colaborações europeias. Isso reflete não só a capacidade científica desses países, mas também o investimento em infraestrutura de dados e políticas de ciência aberta, há mais de dez anos, sob a liderança da Comissão Europeia (Manco, 2022).

Os outros países do continente europeu, cuja colaboração não teve tanto destaque, sobretudo os países localizados na região sudeste, são eles: Albânia,

Bulgária, Croácia, Grécia, Hungria, Montenegro, Macedônia do Norte, Romênia, Eslovênia, Sérvia. A adesão desigual à Ciência Aberta pode explicar esse resultado. Sevkusic, *et al.* (2022) destacam algumas características dos países que compõem essa região e que tem prejudicado a adesão a Ciência Aberta, são elas:

- a) a diversidade política, social e econômica da região, o que tem um impacto direto na pesquisa;
- b) a relação entre a autoridade ou órgão com a imposição das políticas voltadas à Ciência Aberta;
- c) a mobilização entre as várias partes interessadas e comunidades de pesquisa, tem relação com sucesso das iniciativas voltadas à Ciência Aberta. Por exemplo: os países que não fazem parte da União Europeia estão menos integrados nas redes europeias de Ciência Aberta e o investimento no desenvolvimento da investigação desses é geralmente inferior ao do resto da Europa e
- d) o desenvolvimento desigual de infraestrutura e formação para a Ciência Aberta.

No continente asiático, dentre os fatores que influenciam na divulgação global de resultados de pesquisas e por sua vez impactam nas relações de colaboração científica com outros países, de acordo com Cho (2019), está o fato de que muitos países ainda não possuem políticas de Ciência Aberta robustas em comparação com os países desenvolvidos do hemisfério ocidental, as barreiras linguísticas limitam a circulação das publicações na comunidade internacional e a infraestrutura técnica para compartilhamento dos dados de pesquisa é limitada.

Os países com maior destaque nas iniciativas para a Ciência Aberta, são China, Japão e Índia, os quais possuem um maior número de repositórios cadastrados no R3data (Cho, 2019). Esses fatores podem ter se refletido nesta pesquisa, visto que esses três países juntos possuem a metade das instituições do continente asiático cujos autores colaboram com o Brasil, respectivamente China, com 2,57% (65), Japão com 1,66% (42) e Índia com 1,50 (38).

Em relação aos demais continentes, a Oceania concentrou 4,23% (107) das instituições, com destaque para a Austrália com 3,6% (91), o continente africano concentrou 3,16% (80) instituições, com destaque para a África do Sul com 0,75% (19) e os países que têm seus territórios divididos entre Europa e Ásia como Azerbaijão, Rússia e Turquia que, juntos, somaram 2,49% (63) instituições.

Portanto, os dados sugerem que as assimetrias associadas à capacidade científica e tecnológica dos países pode ser um fator que influencia na colaboração (Chen; Li; Jiao, 2022). Há um destaque para os países europeus e asiáticos, possivelmente motivado pelo maior investimento na infraestrutura e políticas de Ciência aberta (Manco, 2022) e uma forte colaboração entre o Brasil e demais países da América Latina.

A tipologia das instituições foi variada, a partir da classificação obtida no registro global ROR. Aquelas ligadas à educação foram as que mais obtiveram destaque na filiação dos autores com 58,77% (1486) do total de instituições identificadas, seguidas dos institutos de pesquisa com 16,13% (408), órgãos do governo com 10,24% (259), ONGs com 6,21% (157), empresas privadas com 3,36% (85), instituições ligadas à saúde com 2,92% (74), arquivos com 1,97% (50) e outros tipos de instituições com 0,43% (11).

Dentre as instituições ligadas à educação, as universidades majoritariamente foram as que mais se destacaram. Elas se consolidaram como centros de conhecimento, são as principais empregadoras de cientistas financiados com recursos públicos e as principais instituições pelas quais novos conhecimentos são criados (International Association of Universities, 2024). Isso pode indicar que os autores filiados a essas instituições têm maior tendência a publicarem *data papers*, uma vez que os pesquisadores vinculados a essas instituições são os mais impactados pelas agências de financiamento e pelas políticas dos periódicos.

A Tabela 2, que trata da distribuição das instituições, tem como objetivo arrolar os nomes daquelas que mais foram mencionadas nos *data papers*. O número de filiações refere-se à quantidade de menções da instituição, que pode ter sido mencionada por um ou mais autores de um mesmo *data paper*.

Foram identificadas 12221 menções de filiação dos autores a diversas instituições. A maioria delas, 63,06% (7707) são de instituições estrangeiras que se encontram classificadas entre diferentes tipologias.

Tabela 2 - Distribuição das instituições com maior número de autores filiados

Nome da instituição	N. Filiações	%	País* Estado	Continente/ Região	Tipo
Centre National de la Recherche Scientifique (CNRS)	173	1,42	FRA	Europa	Governo
University of California	96	0,79	USA	América do Norte	Educação
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)	75	0,61	ARG	América do Sul	Governo
Institut de Recherche pour le Developpement (IRD)	74	0,61	FRA	Europa	Governo
Consejo Superior de Investigaciones Científicas (CSIC)	72	0,59	ESP	Europa	Governo
Russian Academy of Sciences	56	0,46	RUS	Europa/Ásia	Governo
Chinese Academy of Sciences	54	0,44	CHN	Ásia	Governo
Bordeaux Population Health	52	0,43	FRA	Europa	Instituição de Pesquisa
University of Florida	46	0,38	USA	América do Norte	Educação
French National Research Institute for Agriculture, Food and Environment (INRAE)	40	0,33	FRA	Europa	Governo
Total de instituições estrangeiras com mais filiados	738	6,04			
Outras instituições estrangeiras com menos de 40 filiações	6969	57,02			
Total de instituições estrangeiras	7707	63,06			
Universidade de São Paulo (USP)	459	3,76	SP	Sudeste	Educação
Universidade Federal do Mato Grosso do Sul (UFMS)	219	1,79	MS	Centro-Oeste	Educação
Universidade Estadual Paulista (UNESP)	209	1,71	SP	Sudeste	Educação
Universidade Federal de Minas Gerais (UFMG)	142	1,16	MG	Sudeste	Educação
Instituto Federal de Educação, Ciência e Tecnologia de Sergipe	116	0,95	SE	Nordeste	Educação
Universidade Federal do Rio Grande do Sul (UFRGS)	111	0,91	RS	Sul	Educação
Universidade Estadual de Campinas (UNICAMP)	101	0,83	SP	Sudeste	Educação
Universidade Federal de Pernambuco (UFPE)	100	0,82	PE	Nordeste	Educação
Fundação Oswaldo Cruz (Fiocruz)	97	0,79	RJ	Sudeste	Instituição de Pesquisa
Universidade Federal de Santa Catarina (UFSC)	95	0,78	SC	Sul	Educação
Universidade Federal do Paraná (UFPR)	92	0,75	PR	Sul	Educação
Universidade Federal do Pará (UFPA)	90	0,74	PA	Norte	Educação
Total de instituições nacionais com mais filiados	1831	14,98			
Outras instituições brasileiras com menos de 90 filiações	2683	21,96			
Total de instituições nacionais	4514	36,94			
Total geral	12221	100			

Fonte: dados da pesquisa

Notas: *Abreviação de países conforme ISO 3166-1:alpha 3 codes

Número Filiações: refere-se à quantidade de menções que a instituição obteve

%; refere-se à porcentagem de menções em relação ao total de instituições identificadas

País/Estado: refere-se ao país ou estado onde a instituição está localizada conforme ROR

Continente/região: refere-se ao continente ou região brasileira onde a instituição mencionada está localizada

Tipo: refere-se à classificação da instituição conforme ROR

A França foi o país com maior destaque nas filiações estrangeiras, com cinco instituições entre as dez mais mencionadas. Esse fator pode estar atrelado à sua adesão à Ciência Aberta, que existe desde 2018. Nesse plano foi previsto a ampliação das publicações em acesso aberto e a criação de estruturas que promovam o compartilhamento de dados. O plano francês foca nas universidades e organizações de pesquisa e promove a integração entre o Ministério do Ensino Superior, Pesquisa e Inovação e a comunidade científica (Committee for Open Science, [2025?]).

A Universidade da Califórnia foi a mais mencionada, dentre as instituições estrangeiras que mais colaboram com o Brasil na produção dos *data papers*, com 0,79% (96). Nessa universidade, a política de acesso aberto para as publicações dos pesquisadores filiados a ela, existe desde 2013. As bibliotecas possuem um fundo de acesso aberto para ajudar a cobrir as taxas de publicação, além de iniciativas voltadas ao depósito de dados desde 2022 (University of California, 2025).

De maneira geral, pode-se verificar que as instituições educacionais são as principais protagonistas, com quase 60% das colaborações institucionais entre os autores dos *data papers*. Os institutos de pesquisa e governos também têm papel relevante, especialmente em países com políticas públicas voltadas à Ciência Aberta. E, por fim, a presença de empresas e ONGs ainda é tímida, o que pode indicar oportunidades para maior integração entre ciência e setor privado ou social na publicação dos *data papers*.

No âmbito institucional, a partir da quantidade de filiações pôde-se verificar que as políticas e iniciativas voltadas para a Ciência Aberta podem ser fatores propulsores para a publicação dos *data papers*.

4.2.1 Características das instituições brasileiras

No Brasil, observa-se uma forte colaboração entre instituições doméstica, com 15,30% (387) de todas as instituições identificadas, ficando atrás apenas dos Estados Unidos. Esses dados coadunam com a pesquisa de Chen, Li e Jiao (2022), que identificou que os *data papers* dependem fortemente da colaboração entre instituições do mesmo país para sua produção. Na Tabela 3 é apresentada a distribuição geográfica das instituições localizadas no Brasil, a partir da sua concentração nos estados brasileiros e da tipologia predominante.

Tabela 3 - Distribuição geográfica das instituições brasileiras e suas tipologias

	Educação		Governo		ONGs		Empresa Privada		Instituições de pesquisa		Saúde		Outros		Arquivos		Total	
	F	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%
Sul	48	12,40	10	2,58	7	1,81	10	2,58	1	0,26	3	0,78	-	-	-	-	79	20,41
Rio Grande do Sul	20	5,17	5	1,29	1	0,26	2	0,52	-	-	1	0,26	-	-	-	-	29	7,49
Santa Catarina	14	3,62	3	0,78	3	0,78	7	1,81	-	-	1	0,26	-	-	-	-	28	7,24
Paraná	14	3,62	2	0,52	3	0,78	1	0,26	1	0,26	1	0,26	-	-	-	-	22	5,68
Sudeste	82	21,19	18	4,65	22	5,68	22	5,68	19	4,91	8	2,07	2	0,52	2	0,52	175	45,22
São Paulo	35	9,04	11	2,84	16	4,13	10	2,58	9	2,33	5	1,29	2	0,52	-	-	88	22,74
Minas Gerais	24	6,20	1	0,26	4	1,03	7	1,81	5	1,29	2	0,52	-	-	-	-	43	11,11
Rio de Janeiro	19	4,91	5	1,29	2	0,52	5	1,29	4	1,03	1	0,26	-	-	2	-	38	9,82
Espírito Santo	4	1,03	1	0,26	-	-	-	-	1	0,26	-	-	-	-	-	-	6	1,55
Centro-Oeste	21	5,43	13	3,36	5	1,29	5	1,29	-	-	1	0,26	2	0,52	-	-	47	12,14
Distrito Federal	3	0,78	12	3,10	3	0,78	3	0,78	-	-	1	0,26	1	0,26	-	-	23	5,94
Goiás	8	2,07	1	0,26	-	-	1	0,26	-	-	-	-	-	-	-	-	10	2,58
Mato Grosso do Sul	5	1,29	-	-	1	0,26	1	0,26	-	-	-	-	1	0,26	-	-	8	2,07
Mato Grosso	5	1,29	-	-	1	0,26	-	-	-	-	-	-	-	-	-	-	6	1,55
Nordeste	43	11,11	1	0,26	6	1,55	5	1,29	2	0,52	1	0,26	0	-	-	-	58	14,99
Bahia	12	3,10	1	0,26	1	0,26	-	-	-	-	-	-	-	-	-	-	14	3,62
Rio Grande do Norte	5	1,29	-	-	1	0,26	4	1,03	-	-	-	-	-	-	-	-	10	2,58
Pernambuco	5	1,29	-	-	2	0,52	-	-	1	0,26	-	-	-	-	-	-	8	2,07
Ceará	5	1,29	-	-	-	-	1	0,26	-	-	-	-	-	-	-	-	6	1,55
Piauí	3	0,78	-	-	1	0,26	-	-	1	0,26	-	-	-	-	-	-	5	1,29
Sergipe	4	1,03	-	-	1	0,26	-	-	-	-	-	-	-	-	-	-	5	1,29
Paraíba	4	1,03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	1,03
Alagoas	3	0,78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	0,78
Maranhão	2	0,52	-	-	-	-	-	-	-	-	1	0,26	-	-	-	-	3	0,78
Norte	19	4,91	2	0,52	4	1,03	-	-	1	0,26	1	0,26	-	-	1	0,26	28	7,24
Pará	7	1,81	1	0,26	1	0,26	-	-	-	-	-	-	-	-	1	0,26	10	2,58
Amazonas	3	0,78	-	-	3	0,78	-	-	1	0,26	1	0,26	-	-	-	-	8	2,07
Acre	2	0,52	1	0,26	-	-	-	-	-	-	-	-	-	-	-	-	3	0,78
Roraima	3	0,78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	0,78
Rondônia	2	0,52	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0,52
Amapá	1	0,26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,26
Tocantins	1	0,26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,26
Total Geral	213	55,04	44	11,37	44	11,37	42	10,85	23	5,94	14	3,62	4	1,03	3	0,78	387	100

Fonte: dados da pesquisa

Nota: Os números zeros foram substituídos por um traço (-)

Na Tabela 3 é possível verificar que a distribuição das instituições com autores que publicaram *data papers* pelo país, ficou da seguinte maneira: 45,22% (175) das instituições estavam localizadas na região Sudeste, 20,41% (79) na região Sul, 14,99% (58) na região Nordeste, 12,14% (47) na região Centro-Oeste e 7,24% (28) estavam localizadas na região Norte. A partir desses dados, pode-se verificar uma desigualdade na distribuição das instituições cujos autores publicaram *data papers*.

Em relação ao padrão regional da distribuição das publicações e dos autores, Sidone, Haddad e Mena-Chalco (2016) identificaram que elas estão concentradas na região Sudeste, sendo que a concentração espacial está diretamente relacionada à localização dos *campi* das universidades públicas, estaduais e federais, uma vez que essas são responsáveis pela maior parte da atividade científica, padrão típico de países em desenvolvimento.

Em relação à tipologia das instituições onde os autores estavam filiados, em 55,04% (213) dos casos os autores estavam filiados às instituições ligadas à educação, seguido dos órgãos do governo e ONGs com 11,37% (44) cada, em 10,85% (42) às empresas privadas, em 5,94% (23) às instituições de pesquisa, em 3,62% (14) às instituições ligadas à saúde, em 0,78% (3) aos arquivos e em 1,03% (4) dos casos os autores citaram outras instituições.

O Brasil segue a tendência mundial das instituições ligadas à educação. No entanto, os autores filiados a ONGs e a empresas privadas superam a porcentagem mundial de 6,21% e 3,36% respectivamente. Enquanto os autores filiados a instituições de pesquisa estão abaixo do que foi identificado na porcentagem mundial, de 16,13%.

Esses dados indicam, no caso brasileiro, uma lacuna na parceria entre autores de universidades em detrimento de outros tipos de instituições, sugerindo uma possibilidade de exploração das colaborações na produção dos *data papers* entre universidades, empresas privadas e órgãos do terceiro setor.

Especificamente em relação a quantidade de filiações nacionais, estas representam 36,94% (4514) do total identificado, conforme Tabela 3. Neste contexto, pode-se observar um predomínio das filiações nas universidades. A Fiocruz foi a única instituição de pesquisa, dentre aquelas com maior número de menções, com 97 filiações.

A USP foi a instituição com mais autores filiados, dentre todas as instituições identificadas na pesquisa, com 3,76% (459). Essa universidade possui uma Declaração de Apoio à Ciência Aberta desde 2021 e várias iniciativas neste âmbito, tais como repositórios de dados, ferramentas para a elaboração de plano de gestão de dados de pesquisa e portais de acesso aberto (USP, [2025a?]). Além disso, a USP é uma das mais importantes instituições de ensino superior do Brasil, possuindo cerca de 100 mil alunos, um orçamento estadual de aproximadamente 9 bilhões de reais, e a cooperação internacional representa forte diretriz para a internacionalização (USP, [2025b?]).

Silva *et al.* (2024) ao investigarem os *data papers* de pesquisa da área biodiversidade, indexados na plataforma *Global Biodiversity Information Facility* (GBIF), também identificaram uma contribuição significativa dos autores filiados às universidades brasileiras, particularmente a USP, UFMG e a Unicamp.

Aguirre-Ligüera *et al.* (2023) afirmam que a maior parte da ciência dos países periféricos é desenvolvida nas universidades e, por isso, essas unidades possuem forte influência nas práticas associadas à Ciência Aberta. Sendo o Brasil considerado um país periférico ou cientificamente proficiente, de acordo com Chen, Li e Jiao (2022), este fato pode justificar os achados dessa pesquisa.

Aguirre-Ligüera *et al.* (2023) também atrelam às universidades a influência das práticas da Ciência Aberta, uma vez que elas exercem um papel preponderante para a comunicação e a divulgação da pesquisa científica, o que vai ao encontro de ações voltadas para a disseminação do conhecimento por meio da abertura da ciência (Ribeiro; Oliveira; Santos, 2021). Dessa forma, nesta pesquisa foi associada a filiação dos autores às práticas de abertura da ciência das universidades que se destacaram.

No Brasil, no que condiz às normativas voltadas à Ciência Aberta pelas universidades federais, Dias (2025) identificou apenas seis universidades, sendo elas: a Universidade Federal de Goiás (UFG), Universidade Federal de Alagoas (UFAL), Universidade Federal do Ceará (UFC), Universidade Federal de Universidade Federal de Roraima (UFRR), Universidade Federal de São Paulo (Unifesp) e a Universidade Federal do ABC (UFabc), que possuem menções à Ciência Aberta em seus planos de desenvolvimento institucional ou planos estratégicos. Apenas a Universidade Federal do Estado do Rio de Janeiro (UNIRIO) possui uma normativa exclusiva voltada à Ciência Aberta.

Dentre as universidades estaduais paulistas (USP, Unesp e Unicamp), Ribeiro, Oliveira e Santos (2021) identificaram práticas associadas à Ciência Aberta, como a criação de políticas de acesso aberto, repositórios institucionais a fim de divulgar as publicações produzidas pelas universidades. A USP e a Unicamp possuem um portal de periódicos cada. Na Unesp, cada revista ou conjunto de revistas tem sua própria estrutura de gerenciamento individual. Em relação aos dados de pesquisa, por meio da FAPESP foi construído um repositório para o depósito dos dados produzidos por essas instituições e outras envolvidas, além da exigência para a elaboração de planos de gestão de dados de pesquisa.

Os trabalhos de Ribeiro, Oliveira e Santos (2021) e de Dias (2025), apontam que nos documentos das universidades pesquisadas não foram encontradas menções a mecanismos de incentivo, capacitação e/ou treinamento e articulação e/ou coordenação. Ainda as métricas de avaliação para acompanhar a produção científica ainda estão voltadas para as de natureza formal e as métricas alternativas, com dados extraídos de redes sociais, ainda são incipientes e pouco utilizadas.

Por fim, a partir das evidências identificadas na colaboração interinstitucional entre os autores filiados às instituições brasileiras dos *data papers* os achados se assemelham àqueles identificados por Chen, Li e Jiao (2022) que sugerem que o padrão geral de colaboração entre os autores dos *data papers* se assemelha ao da pesquisa científica geral, especialmente a predominância da colaboração doméstica e a alta concentração de colaboração entre países cientificamente mais avançados no que condiz à Ciência e Tecnologia.

4.2.2 Características dos financiamentos

Na esfera social, de acordo com Ribeiro, Oliveira e Santos (2021) tem sido observado questionamentos quanto ao papel da ciência e sua efetividade, no que condiz ao retorno de seus resultados de pesquisa para a sociedade especialmente quando se trata de financiamento com recursos públicos. As agências de financiamento possuem um papel importante, uma vez que destinam recursos para que as pesquisas possam ser realizadas. Assim, os *data papers* podem ser um instrumento importante para medir a eficácia da investigação de projetos financiados (Ünal; Akbulut, 2025).

Dos 634 *data papers* identificados, em 84,70% (537) deles havia a menção à uma ou mais agências de financiamento. Portanto, para esta pesquisa a menção do nome do financiador significa que houve algum tipo de recebimento de financiamento voltado à pesquisa, cujos dados originaram os *data papers*. Nos demais casos, onde não foi identificada a menção ao nome do financiador, foi considerado que não houve financiamento da pesquisa, o que ocorreu em 15,30% (97) dos *data papers*.

Nesse contexto, o recebimento de financiamento mencionado por cerca de 85% dos *data papers* publicados por autores filiados às instituições brasileiras supera os 66% identificados por Ünal e Akbulut (2025) que realizam uma pesquisa sobre a menção a financiamentos nos *data papers* publicados entre 2006-2017, no âmbito mundial. O valor identificado nesta pesquisa se equipara aos 84%, identificados por Kar e Rath (2025), que pesquisaram a menção de financiamento em *data papers* da grande área das Ciências Sociais.

Esse dado indica que houve um aumento nos financiamentos das pesquisas cujos dados originaram os *data papers* em relação ao que foi identificado nos estudos citados. Isso sugere um possível interesse das agências de financiamento a esse tipo de publicação, uma vez que o compartilhamento de dados de pesquisa vem sendo incentivado por elas (Ünal; Akbulut, 2025).

Com o objetivo de descrever as características dos financiamentos das pesquisas cujos dados originaram os *data papers*, a partir da tipologia da agência mencionada e da sua localização, foi elaborada a Tabela 4.

Tabela 4 - Caracterização da tipologia e nacionalidade das agências de financiamentos dos *data papers*

Tipologia	Nacional		Combinado		Estrangeiro		Total	
	F	%	F	%	F	%	F	%
Público	241	44,88	99	18,44	78	14,53	418	77,84
Público; ONG	2	0,37	31	5,77	16	2,98	49	9,12
Público; Privado	11	2,05	25	4,66	2	0,37	38	7,08
Público; Privado; ONG	-	-	13	2,42	3	0,56	16	2,98
Privado	2	0,37	1	0,19	9	1,68	12	2,23
ONG	-	-	-	-	3	0,56	3	0,56
Privado; ONG	-	-	1	0,19	-	-	1	0,19
Total Geral	256	47,67	170	31,66	111	20,67	537	100

Fonte: elaborado pela autora

Notas: Os números zeros foram substituídos por um traço (-)

Coluna tipologia: representa a tipologia da agência de financiamento a partir do identificador ROR

Quando houver mais de um tipo, significa que diferentes agências financiaram as pesquisas.

A coluna nacional refere-se à frequência e à porcentagem de *data papers* que mencionaram as agências de financiamento localizadas no Brasil

A coluna combinado refere-se à frequência e à porcentagem de *data papers* que mencionaram as agências de financiamento localizadas no Brasil e em países do exterior

A coluna estrangeiro refere-se à frequência e à porcentagem de *data papers* que mencionaram as agências de financiamento localizadas fora do Brasil

Na Tabela 4 é possível observar que em 47,67% (256) dos *data papers* foram mencionadas apenas agências nacionais, o que significa que quase a metade dos financiamentos são nacionais. Em 31,66% (170) os financiamentos foram combinados, ou seja, foram citados nos *data papers* tanto agências nacionais quanto estrangeiras, e em 20,67% (111) foram mencionadas apenas agências internacionais, caracterizando financiamentos apenas estrangeiros.

Distribuição semelhante foi observada no trabalho de Alvarez e Caregnato (2021) que, ao pesquisarem os financiamentos de artigos científicos das áreas da Economia, Hematologia, Horticultura e Matemática, também identificaram esse mesmo perfil na nacionalidade das agências, onde cerca de 62% receberam financiamento apenas de instituições brasileiras (financiamento nacional), 27% de instituições brasileiras e estrangeiras (financiamento combinado) e 11% apenas de instituições estrangeiras (financiamento estrangeiro).

Esse dado sugere que em relação à nacionalidade dos financiamentos das pesquisas que originaram os *data papers*, mesmo entre diferentes áreas do conhecimento, segue-se a tendência dos artigos científicos, na qual, de maneira geral, predominam os financiamentos nacionais para as pesquisas.

Em relação às tipologias dos financiamentos, em 77,84% (418) dos *data papers* os financiamentos são de origem pública, advindos de todos os tipos de

nacionalidades, pública, privada e combinada. As outras tipologias de financiamentos ficaram mais dispersas e com menor representatividade. Assim, esses dados indicam a forte presença pública e do Estado nos financiamentos das pesquisas, cujos dados originaram os *data papers*, o que reforça a necessidade da criação de políticas e infraestruturas de gestão dos dados de pesquisa.

A Tabela 5 tem como objetivo caracterizar o recebimento de financiamento e classificá-los quanto à sua nacionalidade, a partir das grandes áreas e áreas do conhecimento dos *data papers*. A proporção foi calculada a partir do número total de *data papers*, dividido pela quantidade de *data papers* que receberam financiamento, ou seja, que informaram o nome da agência.

Todas as grandes áreas registraram o recebimento de financiamento, sendo que a maior porcentagem de recebimento se concentrou nas Ciências Físicas, com 41,48% (263) e nas Ciências da Vida, com 27,92% (177), enquanto as Ciências Sociais concentraram apenas 4,89%% (31) dos financiamentos. Esses resultados têm relação com a quantidade de *data papers*.

Tabela 5 - Caracterização dos financiamentos a partir das grandes áreas e áreas do conhecimento

Grande área/área	Financiamento					Nacionalidade do financiamento						Total de data papers	
	Não		Sim			Nacional		Combinado		Internacional			
	F	%	F	%	Proporção	F	%	F	%	F	%	F	%
Ciências Físicas	53	8,36	263	41,48	83,23	99	18,44	102	18,99	62	11,55	316	49,84
Ciência Ambiental	28	4,42	126	19,87	81,82	36	6,70	62	11,55	28	5,21	154	24,29
Ciência da Computação	5	0,79	45	7,10	90	28	5,21	7	1,30	10	1,86	50	7,89
Ciência dos Materiais	3	0,47	42	6,62	93,33	7	1,30	20	3,72	15	2,79	45	7,10
Ciências da Terra e Planetárias	9	1,42	16	2,52	64	13	2,42	2	0,37	1	0,19	25	3,94
Energia	1	0,16	16	2,52	94,12	8	1,49	6	1,12	2	0,37	17	2,68
Engenharia	3	0,47	10	1,58	76,92	3	0,56	3	0,56	4	0,74	13	2,05
Engenharia Química	2	0,32	3	0,47	60	-	-	1	0,19	2	0,37	5	0,79
Física e Astronomia	1	0,16	2	0,32	66,67	2	0,37	-	-	-	-	3	0,47
Matemática	1	0,16	2	0,32	66,67	1	0,19	1	0,19	-	-	3	0,47
Química	-	-	1	0,16	100	1	0,19	-	-	-	-	1	0,16
Ciências da Vida	30	4,73	177	27,92	85,51	98	18,25	51	9,50	28	5,21	207	32,65
Ciências Agrárias e Biológicas	14	2,21	79	12,46	84,95	42	7,82	23	4,28	14	2,61	93	14,67
Bioquímica, Genética e Biologia Molecular	14	2,21	71	11,20	83,53	39	7,26	24	4,47	8	1,49	85	13,41
Imunologia e Microbiologia	1	0,16	17	2,68	94,44	11	2,05	2	0,37	4	0,74	18	2,84
Farmacologia, Toxicologia e Farmacêutica	1	0,16	5	0,79	83,33	3	0,56	1	0,19	1	0,19	6	0,95
Neurociência	-	-	5	0,79	100	3	0,56	1	0,19	1	0,19	5	0,79
Ciências da Saúde	11	1,74	66	10,41	85,71	45	8,38	7	1,30	14	2,61	77	12,15
Medicina	9	1,42	57	8,99	86,36	40	7,45	5	0,93	12	2,23	66	10,41
Profissões de saúde	1	0,16	6	0,95	85,71	4	0,74	1	0,19	1	0,19	7	1,10
Enfermagem	1	0,16	1	0,16	50	1	0,19	-	-	-	-	2	0,32
Veterinária	-	-	2	0,32	100	-	-	1	0,19	1	0,19	2	0,32
Ciências Sociais	3	0,47	31	4,89	91,18	14	2,61	10	1,86	7	1,30	34	5,36
Ciências Sociais	1	0,16	10	1,58	90,91	4	0,74	5	0,93	1	0,19	11	1,74
Psicologia	1	0,16	7	1,10	87,50	1	0,19	1	0,19	5	0,93	8	1,26
Economia, Econometria e Finanças	-	-	5	0,79	100	2	0,37	3	0,56	-	-	5	0,79
Negócios, Gestão e Contabilidade	-	-	5	0,79	100	3	0,56	1	0,19	1	0,19	5	0,79
Ciências da Decisão	1	0,16	2	0,32	66,67	2	0,37	-	-	-	-	3	0,47
Artes e Humanidades	-	-	2	0,32	100	2	0,37	-	-	-	-	2	0,32
Total Geral	97	15,30	537	84,70	84,70	256	47,67	170	31,66	111	20,67	634	100

Fonte: elaborado pela autora

Notas: Os números zeros foram substituídos por um traço (-)

As colunas financiamento representam os *data papers* que mencionaram ou não o nome das agências de financiamento

As colunas referentes à nacionalidade classificam a localização de agência de financiamento e, assim, a tipologia de financiamento

Diante desses dados, objetivando identificar a proporção de financiamentos, ou seja, a razão entre os projetos financiados e o total de *data papers* por grande área, foi verificado um recebimento superior a 83%, até mesmo na grande área das Ciências Sociais, que publicou menos *data papers*. Isso pode indicar a influência dos mandatos dos financiadores no compartilhamento dos dados de pesquisa e um maior interesse por parte dos autores em obter reconhecimento pela publicação de seus dados de pesquisa.

Quando analisada a proporção de financiamento por área de conhecimento, o índice foi superior a 50% em quase todas as áreas, exceto na Enfermagem. As áreas da Energia, Neurociência, Veterinária, Artes e Humanidades, Economia, Econometria e Finanças, Negócios, Gestão e Contabilidade e Psicologia foram as que melhor aproveitaram os financiamentos, isto é, tiveram financiamentos para todas as pesquisas cujos dados originaram os *data papers*. É importante ressaltar que essas áreas tiveram uma baixa produção de *data papers* pelos autores filiados às instituições brasileiras, o que pode influenciar diretamente nessa análise. No entanto, isso reforça o papel crucial das agências de financiamento na publicação dos dados de pesquisa (Kar; Rath, 2025).

O destaque para as Ciências Físicas está relacionado com a publicação de *data papers* na área das Ciências Ambientais que concentrou 19,87% (126) dos financiamentos da grande área. Esse destaque difere da pesquisa de Ünal e Akbulut (2025), na qual a área que obteve o maior número de financiamentos foi a Ciências Biológicas.

Ao pesquisar artigos científicos, Scartassini e Moura (2020) identificaram que as áreas que mais recebem financiamentos da Capes, CNPq e das FAPs são as áreas da Química, Física, Engenharia e as subáreas das Ciências Biológicas. No caso dos *data papers*, ao se analisar a proporção dos financiamentos, a Química mencionou o financiamento em todos os *data papers*, a Física em 66,67%, a Engenharia em 76,92% e as Ciências Biológicas em 84,95%. Sendo assim, pode-se inferir que os *data papers* seguem a mesma tendência dos artigos científicos, ou seja, essas áreas têm seus objetos de pesquisa fortemente influenciados pelo interesse do Estado.

As Ciências Ambientais foi a área que mais concentrou o recebimento de financiamento com 19,87% (126) do total identificado e uma proporção superior a 81% de financiamento. Dentre as temáticas abordadas nessa área que mais

receberam financiamento o destaque foi para a Ecologia com 13,29% (42), com temas relacionados às mudanças climáticas com 9,49% (30), conservação da natureza e da paisagem com 8,54% (27), modelagem ecológica com 5,7% (18) e engenharia ambiental com 2,85% (9). Esses dados sugerem que há um interesse por parte dos financiadores em temáticas que envolvam as mudanças climáticas e do meio ambiente, uma vez que a distribuição do financiamento para pesquisa tem sido moldada pelas prioridades políticas dos financiadores, o que pode contribuir para identificar potenciais oportunidades de financiamento específicas para cada área (Ünal; Akbulut, 2025).

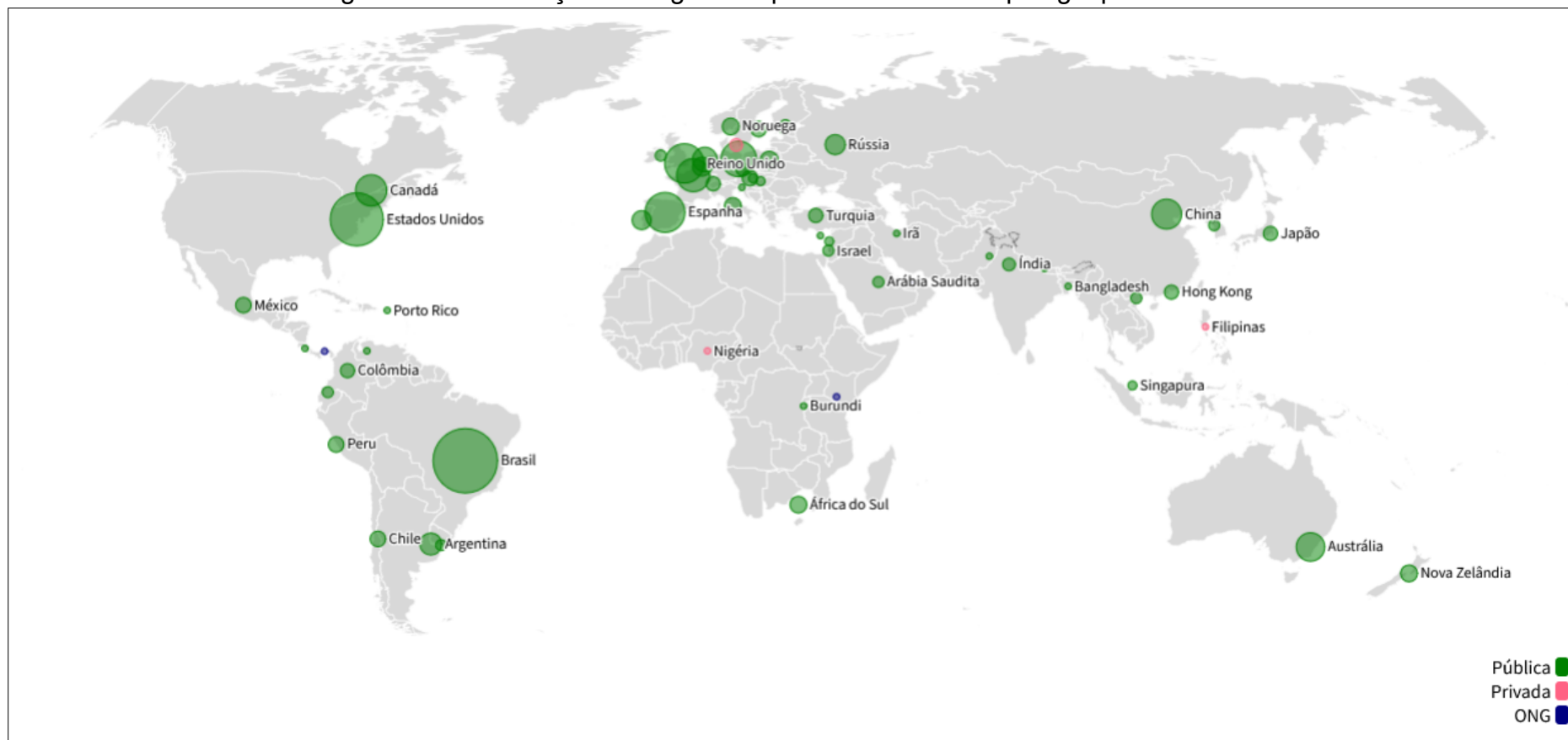
Em relação à nacionalidade dos financiamentos, as áreas com maior diversidade foram: a Ciência Ambiental, com 11,55% (62) de financiamentos combinados, 6,70% (36) de financiamento nacional e 5,21% (28) de financiamento estrangeiro, sendo que essa área foi a que mais concentrou financiamentos combinados e estrangeiros. As demais áreas com maior diversidade na nacionalidade dos financiamentos foram as Ciências Agrárias e Biológicas com 7,82% (42) nacional, 4,28% (23) combinado e 2,61% (14) estrangeiro, a Bioquímica, Genética e Biologia Molecular com 7,26% (39) nacional, 4,47% (24) combinado e 1,49% (8) estrangeiro e a Medicina com 7,45% (40) nacional, 2,23% (12) estrangeiro e 0,93% (5) combinado.

Dentre as pesquisas cujos dados originaram os *data papers* e que receberam financiamento, foram identificadas 838 agências de financiamentos diferentes. Em 10,14% (85) dos casos não foi possível identificar o país⁵⁹ e com isso, as análises se darão a partir de 89,86% (753) das instituições cujas localizações puderam ser identificadas. Além disso, foram identificadas as tipologias de cada agência de financiamento, as quais estão disponíveis no Apêndice E.

A Figura 13 ilustra a distribuição das agências de financiamento pelo mundo. Os tamanhos dos círculos referem-se à quantidade de agências no país e a tipologia predominante em cada país está destacada por cor.

⁵⁹ A impossibilidade de localizar o país das agências de financiamento se deve à perda de dados no relatório extraído do OpenAlex. Por exemplo: em alguns casos foram recuperados apenas partes dos nomes da instituição, instituições com nome homônimo, instituições cuja localização não estava disponível no buscador Google.

Figura 13 - Distribuição das agências pelo mundo e sua tipologia predominante



Fonte: elaborado pela autora

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/21289762/>

Foram identificadas agências de financiamento em 64 países diferentes. As agências públicas foram maioria em grande parte dos países, conforme os círculos na cor verde, exceto na Dinamarca com quatro empresas privadas, Nigéria e Filipinas com uma empresa privada cada e Quênia e Panamá com uma ONG cada.

Conforme pode-se observar pelo tamanho dos círculos, os países com círculos maiores foram aqueles que mais concentraram agências: Estados Unidos com 18,46% (139), Brasil com 17,53% (132), Reino Unido com 6,37% (48), Espanha com 5,71% (43) e Alemanha com 5,58% (42).

A presença dos Estados Unidos, Alemanha e Reino Unido e de países da União Européia como os países que mais possuem agências de financiamento coadunam com a pesquisa Ünal e Akbulut (2025). Na pesquisa desses autores, o Japão foi o segundo país com o maior número de agências com 6,1%, enquanto nesta pesquisa ele concentrou menos de 1% das agências. O que as duas pesquisas têm em comum é o papel dominante das instituições científicas estabelecidas nas nações do Norte Global (Kar; Rath, 2025; Ünal; Akbulut, 2025).

Kar e Rath (2025) destacam que o cenário global de financiamento de pesquisas cujos dados originaram os *data papers* reflete um forte compromisso com a Ciência Aberta e com o avanço de práticas de pesquisa orientadas por dados. Porém, as autoras advertem que apesar dos investimentos do Norte Global, o financiamento no Sul Global permanece limitado, revelando desigualdades em infraestrutura e acesso. Por isso, são necessários modelos de financiamento colaborativo e esforços de capacitação para preencher essas lacunas e promover a participação equitativa na ciência aberta entre todos os países.

Embora na referida essa disparidade, ressaltada pelas autoras, esteja refletida no tamanho menor dos círculos, no caso brasileiro, isso não se reflete, pois pode-se verificar que dentre as menções das agências de financiamento da pesquisa 46,84% (1193) são de agências localizadas no país, enquanto que 53,16% (1354) são agências estrangeiras. Apesar da ligeira diferença, os dados indicam o papel de destaque nacional nos financiamentos das pesquisas.

A Tabela 6 que trata da distribuição das agências de financiamento, tem como objetivo arrolar aquelas que mais foram mencionadas nos *data papers*, bem como sua localização e sua tipologia.

Tabela 6 - Distribuição das agências de financiamentos mais mencionadas

Nome da agência	F	%	País*/ Estado	Continente/Região	Tipo
European Commission (EU)	75	2,94	BEL	Europa	Pública
National Science Foundation (NSF)	61	2,39	USA	América do Norte	Pública
United Kingdom Research and Innovation (UKRI)	57	2,24	GBR	Europa	Pública
Fundação para a Ciência e a Tecnologia (FCT) (Portugal)	34	1,33	PRT	Europa	Pública
Australian Research Council (ARC)	27	1,06	AUS	Oceania	Pública
Deutsche Forschungsgemeinschaft (DFG)	25	0,98	DEU	Europa	ONG
European Research Council (ERC)	22	0,86	BEL	Europa	Pública
National Institutes of Health (NIH)	19	0,75	USA	América do Norte	Pública
Agence Nationale de la Recherche (ANR)	16	0,63	FRA	Europa	Pública
Consejo Nacional de Ciencia y Tecnología (CONACYT)	16	0,63	MEX	América do Norte	Pública
National Aeronautics and Space Administration (NASA)	14	0,55	USA	América do Norte	Pública
Grants-in-Aid for Scientific Research (KAKENHI)	13	0,51	JPN	Ásia	Pública
National Natural Science Foundation of China (NSFC)	13	0,51	CHN	Ásia	Pública
National Oceanic and Atmospheric Administration (NOAA)	12	0,47	USA	América do Norte	Pública
Bundesministerium für Bildung und Forschung (BMBF)	11	0,43	DEU	Europa	Pública
Comisión Nacional de Investigación Científica y Tecnológica (CONICYT)	11	0,43	CHL	América do Sul	Pública
US Department of Defense	11	0,43	USA	América do Norte	Pública
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)	10	0,39	ARG	América do Sul	Pública
Parcial estrangeiras	447	17,55			
Demais instituições com menos de 10 menções	907	41,61			
Total de instituições estrangeiras	1354	53,16			

Nome da agência	F	%	País*/ Estado	Continente/Região	Tipo
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)	370	14,53	DF	Centro-Oeste	Pública
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)	243	9,54	DF	Centro-Oeste	Pública
Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)	154	6,05	SP	Sudeste	Pública
Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ)	43	1,69	RJ	Sudeste	Pública
Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG)	30	1,18	MG	Sudeste	Pública
Institutos Nacionais de Ciência e Tecnologia (INCTs)	20	0,79	DF	Centro-Oeste	Pública
Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS)	18	0,71	RS	Sul	Pública
Universidade de São Paulo (USP)	14	0,55	SP	Sudeste	Pública
Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG)	13	0,51	GO	Centro-Oeste	Pública
Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM)	13	0,51	AM	Norte	Pública
Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES)	13	0,51	ES	Sudeste	Pública
Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE)	12	0,47	PE	Nordeste	Pública
Empresa Brasileira de Pesquisa Agropecuária (Embrapa)	11	0,43	DF	Centro-Oeste	Pública
Financiadora de Estudos e Projetos (FINEP)	11	0,43	RJ	Sudeste	Pública
Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná (FA)	11	0,43	PA	Sul	Pública
Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB)	10	0,39	BA	Nordeste	Pública
Parcial nacional	986	38,71			
Demais instituições brasileiras com menos de 10 menções	207	8,13			
Total de instituições nacionais	1193	46,84			
Total Geral	2547	100			

Fonte: elaborado pela autora

Nota: *Abreviação de países conforme ISO 3166-1:alpha 3 codes

F: refere-se à quantidade de menções que a agência obteve

%; refere-se à porcentagem de menções em relação ao total de agências identificadas

País/Estado: refere-se ao país ou estado onde a agência está localizada conforme ROR

Continente/região: refere-se ao continente ou região brasileira onde a agência mencionada está localizada

Tipo: refere-se à classificação da instituição conforme ROR

Ünal e Akbulut (2025) identificaram que os Estados Unidos ocupam uma posição de liderança, contribuindo significativamente para a pesquisa científica, especialmente por meio do financiamento, destacando-se três agências entre as dez maiores financiadoras. Nesta pesquisa se manteve o destaque para os Estados Unidos, que concentrou cinco agências dentre aquelas com dez ou mais menções.

Enquanto na pesquisa de Ünal e Akbulut (2025) que abrangeu os *data papers* do mundo inteiro, o CNPq foi o único representante da América Latina. Nessa pesquisa, com os *data papers* com pelos menos um autor filiado ao Brasil, dentre as instituições internacionais foi identificada a Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), do Chile, com 0,43% (11) e o Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), da Argentina, com 0,39% (10) do total de menções. Isso ilustra a presença da América Latina nos financiamentos das pesquisas cujos dados originaram os *data papers*, mesmo que ainda incipiente em relação aos países do Norte Global. Apesar desses investimentos, as disparidades de financiamento persistem entre o Norte e o Sul Global, onde recursos limitados dificultam iniciativas de compartilhamento de dados e a participação equitativa na Ciência Aberta (Kar; Rath, 2025).

Tanto nesta pesquisa quando nas pesquisas de Kar e Rath (2025) e Ünal e Akbulut (2025) nenhuma agência africana aparece dentre aquelas que mais financiaram pesquisas, potencialmente refletindo disparidades globais de financiamento em larga escala relevante para *data papers* (Ünal; Akbulut, 2025).

Dentre as agências estrangeiras mencionadas, a Comissão Europeia foi a que mais se destacou, o que pode estar relacionado com a sua consolidação na implementação de políticas voltadas à Ciência Aberta (Kar; Rath, 2025; Ünal; Akbulut, 2025).

Em relação à tipologia, as que mais se destacaram, tanto no âmbito nacional quanto estrangeiro, foram as agências públicas. A DFG, com 0,98% (25) das menções, foi a única ONG com destaque dentre as agências nacionais e estrangeiras mais mencionadas.

Essa agência tem com objetivo promover a pesquisa de alta qualidade em todas as suas formas e disciplinas em universidades e outras instituições de pesquisa alemãs, possuindo mais de 3,9 bilhões de euros disponíveis para investimento, advindos do governo e dos estados (DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 2025). Além disso, coopera com a União Europeia desde 2003 e apóia iniciativas

voltadas ao acesso aberto através da assinatura da Declaração de Berlim. Desde 2022 vem defendendo a criação de uma lei voltada para acesso e disponibilização dos dados de pesquisa (DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 2023). Logo, esse posicionamento em prol da abertura da ciência por meio de financiamentos, é uma exemplo que pode influenciar na publicação dos *data papers*, como forma de angariar fundos para as pesquisas.

No caso das instituições nacionais, o CNPq e a CAPES juntos comportam 24,07% (613) de todas as menções, liderando os financiamentos das pesquisas brasileiras, pois essas são as principais agências de financiamento em território nacional (Scartassini; Moura, 2020). Em seguida, as Fundações de Amparo à Pesquisa (FAPS) mantidas pelos estados também tiveram destaque concentrando 13,58% (346) das menções.

Esses dados sugerem que, mesmo nos *data papers*, se comparado aos artigos tradicionais, essas instituições de fomento à pesquisa são protagonistas para o desenvolvimento da atividade científica, não apenas em nível de disciplina, mas também para a ciência brasileira (Alvarez; Caregnato, 2021).

Embora as fundações tenham se destacado nos financiamentos, em relação às práticas voltadas à Ciência Aberta ou ações mandatórias, Gibbon *et al.* (2024) identificaram que apenas 42,86% (12) das 28 FAPS possuem alguma iniciativa alinhada às práticas de Ciência Aberta em seus websites, sendo que dessas ações, a maioria está voltada à disponibilização de dados abertos, que visam a execução de medidas estabelecidas na legislação federal vigente sobre a transparência de informações públicas e a proteção no uso de dados pessoais.

Em relação aos dados de pesquisa apenas a FAPESP, desde 2018, possui orientações voltadas à elaboração de Plano de Gestão de Dados, ações de apoio a criação de repositórios de dados, além de instruções em seu website descrevendo o compartilhamento de dados como uma prática desejável por parte dos pesquisadores e da instituição (Cunha; Veiga, 2023; Ribeiro; Oliveira; Santos, 2021).

Contudo, as buscas pelas as ações da FAPs voltadas para a Ciência Aberta foram realizadas apenas nos seus websites, o que não impede de haver ações disponibilizadas em editais específicos disponibilizados por elas. Porém, possíveis explicações para a ausência da institucionalização das práticas podem estar atreladas às diferenças das suas criações, sendo que as FAPS mais antigas estão localizadas nas regiões Sul e Sudeste do Brasil, à falta de conhecimento sobre as

vantagens da Ciência Aberta e às disparidades econômicas, na qual cada governo decide a porcentagem que será repassada para as pesquisas por meio das FAPS (Gibbon *et al.*, 2024).

Essas disparidades em relação às FAPS também se refletiram naquelas que possuem maiores menções nesta pesquisa, uma vez que oito das dezesseis de maior destaque estão localizadas nas regiões Sul e Sudeste e as agências localizadas na região Centro-Oeste são vinculadas ao poder executivo federal, exceto a Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG).

Cunha e Veiga (2023) ressaltam que uma política científica de compartilhamento de dados de pesquisa poderá influenciar no papel das FAPS, incluindo como requisito nos seus editais, e como prática a ser exigida para a concessão de financiamento público de pesquisas, além da criação de mecanismos para recompensar o produtor/coletor do dado e a adoção de formas padronizadas para facilitar a citação do produtor, contribuidores e dos repositórios de dados de pesquisa.

Em relação às áreas cobertas pelas agências de financiamento, na pesquisa de Ünal e Akbulut (2025) foi possível identificar que dentre as que mais foram mencionadas, aquelas ligadas à saúde se sobressaíram, tais como: *National Institutes of Health* (USA); *Japan Society for the Promotion of Science* (JPN); *National Institute of General Medical Sciences* (USA) e *National Cancer Institute* (USA). No caso desta pesquisa, observa-se a presença de instituições públicas de caráter multidisciplinar.

Por fim, as agências de financiamento e acadêmicas terão que trabalhar no desenvolvimento de políticas formais para reconhecer os pesquisadores que geram dados de uso para a comunidade científica (Gorgolewski; Margulies; Milham, 2013). Este trabalho deverá ser realizado de maneira conjunta com as instituições as quais os autores estão filiados.

4.2.3 Considerações da seção

Os autores filiados às instituições brasileiras que publicaram *data papers* representam a colaboração entre 2530 instituições diferentes, distribuídas entre 113 países. Os países que tiveram maior colaboração institucional com o Brasil foram os Estados Unidos com 16,13% (408), França com 5,77% (146), Alemanha com 5,10%

(129), Grã-Bretanha com 4,51% (114) e Espanha com 3,99% (101). Logo, a maior colaboração interinstitucional se deu entre os países com maior capacidade de Ciência e Tecnologia e com uma maior tradição em políticas de Ciência Aberta.

Foi identificada também uma tendência de colaborações entre instituições localizadas na América Latina, embora essa tendência seja desigual quando analisados os países separadamente, com destaque para as colaborações entre instituições localizadas na Argentina com 8,5% (54), México com 8,35% (53) e Colômbia com 6,3% (40).

Dentre os países do continente europeu, o destaque das colaborações institucionais, se deu com a França, Alemanha, Reino Unido, Espanha e Itália, os quais contemplam mais da metade das colaborações do continente. Isso reflete não só a capacidade científica e tecnológica desses países, mas também o investimento em infraestrutura de dados e políticas de Ciência Aberta, há mais de dez anos, sob a liderança da Comissão Europeia, embora também existam assimetrias nas colaborações entre os países do sudeste europeu.

Em relação ao continente asiático o destaque se deu com a China, Japão e Índia, no continente africano, as colaborações se deram entre a África do Sul e na Oceania o destaque foi para a Austrália, embora com esses continentes a colaboração tenha sido em menor escala do que com os demais.

Outra característica identificada foi a tipologia dessas instituições, sendo predominantemente universidades, dentre as quais se destacam a USP no âmbito nacional e a Universidade da Califórnia no âmbito internacional. Essas universidades têm em comum a presença de políticas institucionais voltadas ao acesso aberto e o compartilhamento dos dados de pesquisa.

No Brasil, observa-se uma forte colaboração doméstica entre as instituições nacionais na produção dos *data papers*, com destaque para as instituições localizadas na região sudeste. Embora mantenha o padrão de autores filiados às universidades, houve também um destaque em relação à porcentagem mundial de autores filiados às ONGs e às empresas privadas.

Portanto, em relação à autoria pode-se inferir que no caso dos *data papers*, ela segue o padrão da produção do conhecimento dos artigos científicos, onde as colaborações tendem a ser realizadas com países de destaque na ciência e tecnologia, embora tenha-se verificado que os países que mais investem nas políticas

e ações voltadas à Ciência Aberta podem ser influenciadores das colaborações de coautoria dos *data papers*.

Em relação ao financiamento, foi realizada uma análise em nível das grandes áreas e áreas do conhecimento e em nível do financiador, da sua tipologia e nacionalidade. Com isso, foi visto que, em relação à nacionalidade da agência financiadora, os *data papers* seguem a tendência dos artigos científicos, com predomínio das agências nacionais e financiamentos combinados entre agências nacionais e estrangeiras.

O financiamento público das pesquisas foi mencionado em mais de 70% dos *data papers*, o que reforça a necessidade de se haver um retorno à sociedade das pesquisas que recebem esse tipo de financiamento a partir de políticas públicas.

Em relação à proporção de financiamentos, foi verificado um recebimento superior a 83%, até mesmo na grande área das Ciências Sociais, que publicou menos *data papers*.

A área das Ciências Ambientais concentrou a maior parte dos financiamentos dentre todas as áreas identificadas, sendo que os temas mais financiados foram Ecologia, bem como temas relacionados a mudanças climáticas, conservação da natureza e da paisagem, modelagem ecológica e engenharia ambiental. Assim, as áreas relacionadas ao meio ambiente e às mudanças climáticas, tendem a receber maior apoio das agências de financiamento.

Os países do Norte Global, principalmente os Estados Unidos, Reino Unido, Espanha e Alemanha foram os que mais concentraram agências de financiamentos, sendo as instituições públicas as predominantes em todo o mundo.

Dentre as agências estrangeiras mais mencionadas, a Comissão Europeia foi aquela que mais se destacou, o que pode estar relacionado com a sua consolidação na implementação de políticas de compartilhamento de dados (Ünal; Akbulut, 2025). A DFG foi a única ONG com destaque dentre as agências nacionais e estrangeiras mais mencionadas. Dentre as agências nacionais, o CNPq, a CAPES e as FAPs foram as agências mais mencionadas, embora as políticas e práticas voltadas ao compartilhamento de dados e Ciência Aberta ainda sejam incipientes.

4.3 CARACTERÍSTICAS DOS PERIÓDICOS

Esta seção tem como objetivo apresentar as características dos periódicos que publicaram os *data papers* dos autores filiados às instituições brasileiras.

O total de *data papers* publicados (634), por autores filiados a pelo menos uma instituição brasileira, está concentrado em 38 periódicos, uma média de 16,7 *data papers* por periódico. O destaque foi para o periódico *Data in Brief* (Elsevier) que publicou quase a metade deste tipo de documento, 49,68% (315), seguido do *Scientific Data* (Springer) com 16,25% (103). Ambos são periódicos *data journals* (D), ou seja, periódicos especializados em publicar dados de pesquisa e além disso, são multidisciplinares. Esse resultado indica que os autores filiados às instituições brasileiras vêm seguindo a tendência mundial na seleção de *data journals* para publicar seus *data papers*, visto que essa concentração da publicação de *data papers* nesses periódicos, também foi observada por Le Fournier e Schöpfel (2025).

Pinto (2022) também observou o destaque para as publicações no *Data in Brief* (Elsevier) na área da Enfermagem. A autora pressupõe que esse resultado possa estar relacionado aos fatores relacionados a não disponibilização desta tipologia de artigo em periódicos do escopo da Enfermagem, à escassez de periódicos especializados que ofereçam ao pesquisador a possibilidade de comunicar sua pesquisa via *data paper* e às pesquisas de natureza interdisciplinar e/ou de colaboração multidisciplinar, onde os resultados podem ser publicados em fontes de escopo amplo, como por exemplo a Nature. Assim, tais fatores podem estar redesenhando o escoamento da produção científica da área investigada (Pinto, 2022).

Ao se analisar os periódicos que mais publicam *data papers* na grande área das Ciências Sociais, Kar e Rath (2025) observaram que os cinco periódicos que mais concentraram foram todos *data journals*, sendo eles: *Scientific Data* (Elsevier), *Data* (MDPI), *Journal of Open Humanities Data* (Ubiquity Press), *Journal of Open Archaeology Data* (Ubiquity Press) e *Journal of Open Psychology Data* (Ubiquity Press). Desses periódicos, apenas três foram identificados na presente pesquisa.

Na Tabela 7 são apresentadas as características dos periódicos que publicaram os *data papers* dos autores filiados às instituições brasileiras.

Tabela 7 - Características dos periódicos

Nome do periódico	F	%	T	Ano 1º N.	Ano public. 1º dp	T. dp index.	<i>Publisher</i>	País *	Indexação no DOAJ	APC (USD) **	Estimativa (USD)
Data in Brief	315	49,68	D	2014	2015	10468	Elsevier	USA	Sim	1.010,00	318.150,00
Scientific Data	103	16,25	D	2014	2014	3036	Springer Nature	GBR	Sim	2.490,00	256.470,00
Ecology	51	8,04	M	1920	2017	177	Wiley	USA	Não	3.970,00	202.470,00
Earth System Science Data	36	5,68	D	2009	2012	829	Copernicus Publications	DEU	Sim	1.521,66	54.779,76
Biodiversity Data Journal	26	4,10	D	2013	2013	389	Pensoft Publishers	BGR	Sim	865,28	22.497,28
Chemical Data Collections	22	3,47	D	2016	2016	1093	Elsevier	NLD	Não	1.330,00	29.260,00
Data	17	2,68	D	2016	2016	408	Multidisciplinary Digital Publishing Institute (MDPI)	CHE	Sim	1.805,66	30.696,22
Global Ecology and Biogeography	9	1,42	M	1998	2018	44	Wiley	GBR	Não	5.000,00	45.000,00
BMC Research Notes	5	0,79	M	2008	2008	90	Elsevier	GBR	Sim	1.690,00	8.450,00
Gigascience	5	0,79	D	2012	2012	236	Oxford University Press	GBR	Sim	2.638,00	13.190,00
Proteomics	4	0,63	M	2001	2016	37	Wiley	DEU	Não	4.650,00	18.600,00
Frontiers in Ecology And Evolution	3	0,47	M	2013	2016	14	Frontiers Media	CHE	Sim	512,00	1.536,00
Frontiers in Marine Science	3	0,47	M	2014	2016	47	Frontiers Media	CHE	Sim	512,00	1.536,00
Geoscience Data Journal	3	0,47	D	2014	2014	113	Wiley	GBR	Sim	2.290,00	6.870,00
BMC Bioinformatics	2	0,32	M	2000	2018	10	Elsevier	GBR	Sim	2.990,00	5.980,00
Frontiers in Earth Science	2	0,32	M	2013	2018	9	Frontiers Media	CHE	Sim	512,00	1.024,00
Frontiers in Genetics	2	0,32	M	2010	2017	46	Frontiers Media	CHE	Sim	512,00	1.024,00
Frontiers in Neuroscience	2	0,32	M	2007	2015	14	Frontiers Media	CHE	Sim	512,00	1.024,00
Frontiers in Psychology	2	0,32	M	2010	2015	33	Frontiers Media	CHE	Sim	512,00	1.024,00
Marine Genomics	2	0,32	M	2008	2021	156	Elsevier	NLD	Não	2.960,00	5.920,00
Microbiology Resource Announcements	2	0,32	M	2018	2021	102	American Society for Microbiology	USA	Não	1.400,00	2.800,00
Zookeys	2	0,32	M	2008	2011	39	Pensoft Publishers	BGR	Sim	978,21	1.956,42
Annals of Forest Science	1	0,16	M	1999	2016	31	Elsevier	GBR	Sim	2.390,00	2.390,00
Big Earth Data	1	0,16	D	2017	2018	25	Taylor and Francis	GBR	Sim	815,00	815,00
Cahiers Agricultures	1	0,16	M	2006	2018	5	EDP Sciences	FRA	Sim	-	-
Comparative Cytogenetics	1	0,16	M	2010	2017	2	Pensoft Publishers	BGR	Sim	378,56	378,56
Ecosistemas	1	0,16	M	2001	2021	4	Asociacion Espanola de Ecologia Terrestre	ESP	Sim	-	-
Frontiers in Bioinformatics	1	0,16	M	2007	2023	2	Frontiers Media	CHE	Sim	512,00	512,00
Frontiers in Microbiology	1	0,16	M	2010	2015	7	Frontiers Media	CHE	Sim	512,00	512,00
Frontiers in Molecular Biosciences	1	0,16	M	2014	2017	3	Frontiers Media	CHE	Sim	512,00	512,00

Nome do periódico	F	%	T	Ano 1º N.	Ano public. 1º dp	T. dp index.	Publisher	País *	Indexação no DOAJ	APC (USD) **	Estimativa (USD)
Frontiers in Public Health	1	0,16	M	2013	2017	14	Frontiers Media	CHE	Sim	512,00	512,00
Human Mutation	1	0,16	M	1992	2019	10	Wiley	GBR	Não	2.490,00	2.490,00
Journal of Biogeography	1	0,16	M	1974	2020	7	Wiley	GBR	Não	4.830,00	4.830,00
Journal of Open Humanities Data	1	0,16	D	2015	2015	52	Ubiquity Press	GBR	Sim	646,15	646,15
Nature Conservation	1	0,16	M	2012	2013	5	Pensoft Publishers	BGR	Sim	1.086,00	1.086,00
Science of Nature	1	0,16	M	1913	2017	2	Springer Nature	DEU	Não	3.790,00	3.790,00
Scientific Reports	1	0,16	M	2011	2019	1	Springer Nature	GBR	Sim	2.690,00	2.690,00
Taxon	1	0,16	M	1951	2021	4	International Association for Plant Taxonomy	AUT	Não	2.630,00	2.630,00
Total Geral	634	100				17.564					1.054.051,39

Fonte: elaborado pela autora

Notas: * Abreviação de países conforme ISO 3166-1:alpha 3 codes

** Valor da APC de 2025: quando necessário os valores foram convertidos para dólares americanos (USD), em 07 de março de 2025, no Conversor de Moedas do Banco Central do Brasil, disponível em <https://www.bcb.gov.br/conversao>

Coluna F: refere-se à quantidade de *data papers* publicados pelo periódico, no qual pelo menos um autor é filiado a uma instituição brasileira

Coluna %: refere-se à porcentagem de *data papers* concentradas no periódico em relação ao total publicado

Coluna tipo: tipo do periódico: D *data journal* e M misto

Coluna Ano 1º N.: refere-se ao ano de publicação do primeiro número do periódico

Coluna Ano public. 1º dp: refere-se ao ano de publicação do primeiro *data paper* do periódico

Coluna T. dp index.: refere-se ao total de *data papers* indexados nas bases de dados Scopus e WoS. Houve divergências nas indexações das duas bases. Nesta coluna foram considerados a maior quantidade de *data papers* indexadas. Apenas o Chemical data collections, Data in brief, Earth system science data, Ecossistemas, Frontiers in neuroscience, Human mutation, Journal of open humanities data, Marine genomics e o Scientific data estavam em maior quantidade indexados na Scopus, os demais estavam indexados em maior quantidade na WoS.

Coluna *publisher*: refere-se ao nome da empresa editora do periódico

Coluna país: refere-se ao país de localização do periódico conforme Scimago

Coluna indexação no DOAJ: refere-se à verificação se o periódico está indexado ou não neste diretório

Coluna APC (USD): refere-se ao valor da APC cobrada pelo periódico no ano de 2025.

Coluna Estimativa (USD): refere-se à estimativa investida no periódico, resultado da multiplicação da quantidade de *data papers* e o valor da APC individual.

Dos 38 títulos identificados, 26,32% (10) se caracterizam por *data journals*, periódicos especializados em publicar dados de pesquisa, não necessariamente *data papers*. Por exemplo, o *Biodiversity Data Journal* (Pensoft Publishers) publica artigos de pesquisa, taxonomias e inventários, perfil de conservação de espécies (Pensoft Publishers, c2024), o *Earth System Science Data* (Copernicus Publications) publica artigos de revisão e perspectivas⁶⁰ (Copernicus Publication, [2024]) e o *Data in Brief* (Elsevier) publica cartas ao editor, revisões e perspectivas (Elsevier, c2024b; c2024c). Embora a quantidade de títulos seja menor para esse tipo de periódico, ao somar o total de *data papers* publicados, esses concentram 83,44% (529) dos *data papers* publicados. Isso pode indicar que os autores filiados às instituições brasileiras reconhecem esse tipo de periódico como veículos para a publicação dos seus dados de pesquisa.

No entanto, ao analisar especificamente os *data papers* da área da Biodiversidade, Silveira, Dias e Neubert (2023b) verificaram que a área tende a publicar seus *data papers* em periódicos mistos. Esse fato pode indicar que a área tem influência na seleção do tipo de periódico adotado, embora estudos complementares sobre esse comportamento precisem ser realizados.

Os *data journals* se diferem dos periódicos mistos, ou seja, aqueles cuja ênfase é publicar artigos científicos tradicionais, pelo seu escopo. Por exemplo, no caso do *Biodiversity Data Journal* (Pensoft Publishers), seu escopo é “acelerar a publicação, disseminação e compartilhamento de dados relacionados à biodiversidade de qualquer tipo” (Pensoft, c2024, p. 1, tradução nossa) e no caso do *Data in Brief* (Elsevier) é um periódico multidisciplinar que publica principalmente *data papers* curtos que descrevem e fornecem acesso a dados de pesquisa (Elsevier, c2024c). Dessa maneira, pode-se sintetizar que os *data journals* têm como objetivo publicar e compartilhar dados de pesquisa seja no formato de *data paper* ou outro formato.

Embora na década de 1950 e 1970 já existissem periódicos especializados em publicar dados de pesquisa, os *data journals* começam a ganhar evidência, a partir de 2010, impulsionados pelo movimento da Ciência Aberta e vem tendo seu impacto percebido na comunicação científica pelo crescimento exponencial em

⁶⁰ Perspectiva: usada no início de uma edição, destina-se a artigos que fornecem uma visão geral sintética, prática, comentários críticos ou perspectiva histórica e são escritos principalmente por um eminente pesquisador.

diversas áreas do conhecimento (Aventurier; Alencar, 2016; Pinto; Pinheiro, 2019; Santos; Nassi-Caló, 2020; Thelwall, 2020; Walters, 2020).

No caso desta pesquisa, com exceção do *Earth System Science Data (Copernicus Publications)*, cujo primeiro número foi publicado em 2009, os demais *data journals* da amostra começaram a ser publicados a partir de 2012. Tratando-se dos periódicos mistos, pode-se verificar que há uma concentração da publicação de *data papers* entre 2015 e 2017.

Diante disso, pode-se afirmar que os periódicos mistos, conforme pode-se observar na coluna de início da publicação de *data papers*, tiveram uma adesão paulatina entre sua criação, isto é, a publicação do primeiro número, até a publicação dos *data papers*. Em alguns casos, como do periódico *Science of Nature (Springer)*, a variação foi de 104 anos, e no caso, do *Ecology (Wiley)*, a variação foi de 97 anos. Possivelmente, no caso dos periódicos mistos, foi havendo a criação de seções especiais para que eles pudessem acompanhar os mandatos dos financiadores pelo compartilhamento dos dados de pesquisa pois, de acordo com Zhang *et al.* (2022) este mercado da publicação científica é influenciado pelos governos e organizações de financiamento em prol do movimento de acesso aberto.

Independente do tipo de periódico, é possível verificar que a adesão dos periódicos à publicação de *data papers* vem crescendo. Candela *et al.* (2015) identificaram 116 periódicos que publicaram *data papers* e, desses, apenas sete eram considerados *data journals*. Cinco anos depois, Walters (2020) identificou 169 periódicos, sendo que desses 13 publicaram *data papers* regularmente. Nesta pesquisa, utilizando o recorte de autores com vínculo nacional e indexados nas bases WoS e Scopus, foram identificados 38 periódicos, desses 10 são *data journals*.

Os periódicos *Ecosistemas*, publicado pela *Asociacion Española de Ecología Terrestre*, o *Microbiology Resource Announcements*, publicado pela *American Society for Microbiology*, e o *Taxon*, publicado pela *International Association for Plant Taxonomy*, juntos correspondem por 0,63% (4) dos *data papers* publicados por periódicos editados por associações e o *Gigascience*, publicado pela *Oxford University Press*, corresponde a 0,79% (5), sendo editado por uma universidade. Os demais, 98,58% (625) são *data papers* publicados por periódicos editados por *publishers* comerciais.

Em relação aos periódicos mantidos pelos *publishers* comerciais (35), o *Frontiers Media* concentra a maior parte dos títulos, dez, seguido da *Elsevier* e da

Wiley com seis títulos cada, a *Pensoft Publishers* com quatro, a *Springer Nature* com três e os demais editores com um título cada.

Atualmente quase todas as principais editoras científicas comerciais lançaram seus *data journals* e desses mais da metade pertence às cinco maiores editoras comerciais (*Elsevier*, *Springer Nature*, *Wiley*, *Taylor & Francis*, *SAGE*) (Le Fournier; Schöpfel, 2025). Com exceção da *SAGE*, nesta pesquisa os *data journals* (10) foram publicados por pelo menos um desses *publishers*, o que indica que esses editores vêm investindo nesse segmento (García-García; López-Borrull; Peset, 2015). Ao se analisar o *publisher* que mais publicou *data papers*, os cinco que mais concentraram foram: *Elsevier* com 54,73% (347) *Springer Nature* com 16,56% (105), *Wiley* com 10,88% (69), *Copernicus Gesellschaft MBH* com 5,68% (36) e a *Pensoft Publishers* com 4,73% (30).

Nesse sentido, no âmbito da publicação dos *data papers*, há o predomínio de *publishers* comerciais, sendo que os três que mais concentram a publicação desses documentos, estão entre as cinco grandes editoras comerciais, *Elsevier*, *Sage*, *Springer Nature*, *Taylor & Francis* e *Wiley*, que controlam a publicação de artigos em acesso aberto, conforme identificado por Butler *et al.* (2023).

Esses *publishers* mantêm um oligopólio da publicação científica (Butler *et al.*, 2023). Logo, a publicação dos dados de pesquisa, incentivado pelas instituições de fomento no contexto dos *data papers*, vem coadunando com as práticas de publicação dos artigos científicos, na qual o cenário tem sido “[...] marcado pelo crescente controle e exploração econômica do acesso ao conhecimento científico publicado por oligopólios formados por poucos editores comerciais [...]” (Weitzel, 2022, p. 174).

Os *publishers* comerciais se apropriam mais rapidamente do aparato tecnológico para oferecer produtos e serviços atrelados ao acesso aberto, uma vez que cobram altos preços e podem contratar mais funcionários e, portanto, os processos editoriais (gestão de revisões, correções, diagramação, publicação, etc.) são mais ágeis e rápidos. Além disso, as plataformas de publicação desses periódicos possuem melhores funções para gestão de revisões, layouts mais atraentes, a possibilidade de utilização de repositórios de pré-prints e a inclusão de métricas complementares para cada um dos artigos (número de citações, downloads e presença em redes sociais etc.) (Anglada; Abadal, 2023; Weitzel, 2022).

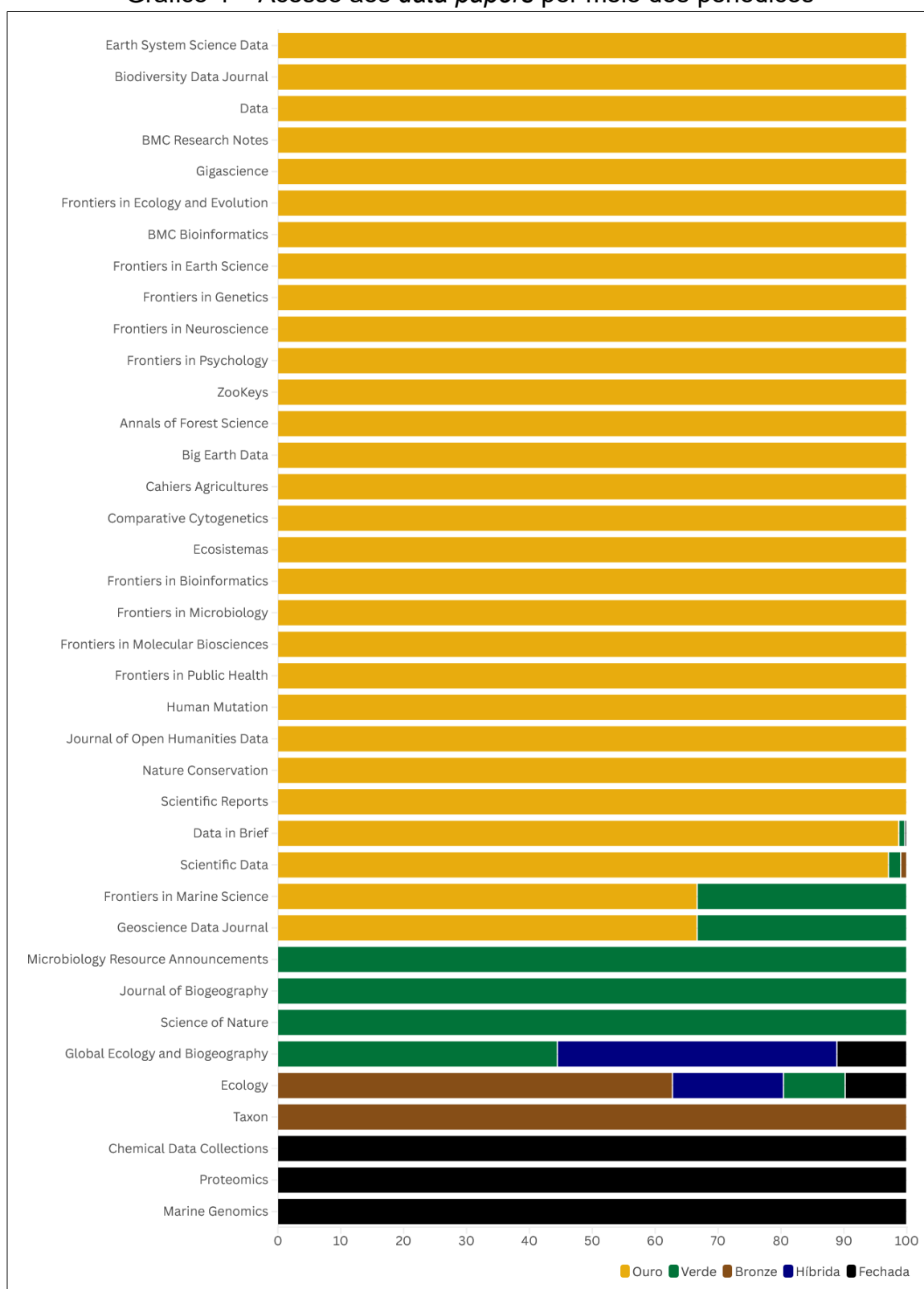
Pinto (2022) identificou que a publicação dos *data papers* em periódicos estrangeiros ou internacionais vem sendo uma saída para os autores filiados às instituições brasileiras, uma vez que o país se encontra como um participante inicial do processo de publicação de dados de pesquisa em periódicos científicos nacionais em algumas áreas do conhecimento.

Essa identificação de Pinto (2022) coaduna com os achados nesta pesquisa, pois em relação a nacionalidade dos periódicos, 12 estão localizados na Grã-Bretanha (GBR), 11 na Suíça (CHE), quatro na Bulgária (BGR), três na Alemanha (DEU) e nos Estados Unidos (USA), dois na Holanda (NDL) e um na França (FRA), na Espanha (ESP) e na Áustria (AUT). A partir disso, pode-se verificar que há uma concentração de periódicos localizados no continente europeu que publicaram os *data papers* da amostra, e esse dado pode estar relacionado às iniciativas voltadas à Ciência Aberta na região, sobretudo por parte da União Europeia.

Nenhum periódico é editado no Brasil, o que pode indicar a necessidade da criação de periódicos especializados nesse tipo de publicação (*data journals*) ou na criação ou maior ênfase de seções específicas para *data papers* nos periódicos já existentes, conforme sugestão de Torino, Roa-Martínez e Vidotti (2020). Logo, esse dado indica a necessidade de criação de vias de financiamento para esse tipo de periódico e de publicação, no caso dos *data papers*, caso o Brasil deseje acompanhar as mudanças impulsionadas pelo movimento da Ciência Aberta.

Visando identificar a adesão dos periódicos ao acesso aberto e considerando o DOAJ como parte vital desse processo (DOAJ, c2024), foi verificado que dos 38 periódicos identificados, dez não estavam indexados nesse diretório, são eles: *Chemical Data Collections* (Elsevier), *Ecology* (Wiley), *Global Ecology and Biogeography* (Wiley), *Human Mutation* (Wiley), *Journal of Biogeography* (Wiley), *Marine Genomics* (Elsevier), *Microbiology Resource Announcements* (American Society for Microbiology), *Proteomics* (Wiley), *Science of Nature* (Springer) e o *Taxon* (International Association for Plant Taxonomy).

No Gráfico 4 é possível analisar a forma de acesso ao *data paper* a partir de cada periódico.

Gráfico 4 – Acesso aos *data papers* por meio dos periódicos

Fonte: elaborado pela autora

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/28266518/>

Nos periódicos *Ecology* (Wiley) (32), *Scientific Data* (Springer) (1) e *Taxon* (International Association for Plant Taxonomy) (1) pode-se observar que há um

predomínio da via bronze para acesso aos *data papers*, ou seja, nestes periódicos não é possível precisar por quanto tempo este tipo de documentos ficará disponível para os leitores.

Nos periódicos *Chemical Data Collections* (Elsevier) (22), *Proteomics* (Wiley) (4) e *Marine Genomics* (Elsevier) (2) pode-se verificar que o acesso aos *data papers* é fechado, sugerindo que o acesso a esse tipo de documento é realizado somente via assinatura do periódico.

Nesses casos, a falta de clareza quanto ao tempo de acesso ao *data paper* ou o acesso mediante assinatura do periódico contrasta com o objetivo da Ciência Aberta que é tornar os dados mais acessíveis aos possíveis reutilizadores (Walters, 2020).

O acesso aos *data papers* pela via verde foi predominante por meio dos periódicos *Microbiology Resource Announcements* (American Society for Microbiology) (2) e *Journal of Biogeography* (Wiley) (1) e o *Science of Nature* (Springer) (1). Sendo possível ser acessado também nos periódicos *Global Ecology and Biogeography* (Wiley) (4), *Ecology* (Wiley) (2), no *Data in Brief* (Elsevier) (3), *Scientific Data* (Springer) (2), *Frontiers in Marine Science* (Frontiers) (1) e *Geoscience Data Journal* (Wiley) (1)

Contudo, houve o predomínio da via ouro no acesso *data papers*, pela qual todos os *publishers* cobram taxas de APC, exceto o EDP Sciences, *publisher* do Cahiers Agricultures e a Asociacion Espanola de Ecologia Terrestre, *publisher* do Ecosistemas. Com isso, a publicação dos *data papers* segue o mesmo rito dos artigos tradicionais, na qual “[...] os editores comerciais cobram uma taxa dos autores para manter o artigo disponível de forma perpétua em acesso “aberto”. Quando essa taxa não é paga, a comunidade científica, através de suas instituições de origem, arca com os custos do acesso a esse conteúdo através das assinaturas anuais dos títulos desses periódicos” (Weitzel, 2022, p. 177).

Embora o Acesso Aberto tenha como objetivo fornecer acesso gratuito e irrestrito à pesquisa, os editores encontraram um novo fluxo de receita com a cobrança de APCs, por meio da via ouro, além do modelo tradicional de negócios de taxas de assinatura pagas pelas bibliotecas acadêmicas, criando assim uma barreira econômica (Butler *et al.*, 2023). Com isso, em vez de tornar a publicação acadêmica sustentável e acessível para todos, os altos APCs visam preservar o *status quo*, onde

a publicação acadêmica é um negócio altamente lucrativo para algumas corporações (Butler *et al.*, 2023).

Candela *et al.* (2015) identificaram que a cobrança de APC para *data journals* é uma abordagem padrão na pesquisa realizada. Os achados de Le Fournier e Schöpfel (2025) coadunam com essa linha, visto que os autores afirmam que a cobrança de APC é o modelo de negócio dominante entre os periódicos de acesso aberto que publicam *data papers*. O predomínio da cobrança de APC, tanto pelos *data journals* quanto pelos periódicos mistos, se confirmou nessa pesquisa, onde apenas dois periódicos, o *Cahiers Agricultures (EDP Sciences)* e o *Ecosistemas (Asociacion Espanola de Ecologia Terrestre)*, não cobram taxas para a publicação dos *data papers*.

As maiores taxas de APC são cobradas pelos periódicos mistos, cujas taxas variam de US\$ 382,64 a US\$ 4.760,00. Os periódicos *Big Earth Data (Taylor and Francis)*, *Comparative Cytogenetics (Pensoft Publishers)*, *Gigascience (Oxford University Press)* e todos os periódicos da *Frontiers* identificados na amostra cobram taxas de APC diferenciadas para a publicação de *data papers*. Em todos os demais periódicos o valor é o mesmo tanto para artigos científicos quanto para *data papers*.

Ao realizar a soma dos valores das taxas de APC dos periódicos mistos, o valor corresponde a 75,91% (US\$ 45.974,82) do valor total estimado com o gasto na publicação dos *data papers*. Além disso, esses periódicos são mais antigos e, com isso, sua reputação pode influenciar os valores cobrados para a publicação de *data papers*, uma vez que estes têm o mesmo status de um artigo tradicional. Björk e Solomon (2015) identificaram que periódicos com uma relação de melhor qualidade, ou seja, maior prestígio na sua área, tendem a atrair maiores volumes de publicação, embora suas APCs sejam altas.

Anglada e Abadal (2023) complementam que a demanda por parte dos pesquisadores para publicar em periódicos com maior fator de impacto, se deve ao fato de que estes oferecem as oportunidades de ascensão acadêmica que buscam, sugerindo que os autores não são sensíveis ao aumento de preços nem à diversidade existente na oferta de periódicos.

Pinto (2022) verificou que o fator econômico não se apresentou como barreira para a escolha por este tipo de publicação pelos autores da área da Enfermagem, que selecionaram *Data in Brief (Elsevier)* para publicar seus *data papers*. Os achados nesta pesquisa sugerem que este fator também não é preponderante, uma vez que

houve uma concentração de *data papers* neste periódico em detrimento de outros que cobravam taxas de APC mais baixas ou não cobravam.

Na Tabela 8 é apresentada uma estimativa de gastos com a APC por área do conhecimento.

Tabela 8 - Estimativa de gastos com APC por áreas, conforme esquema de classificação da All Science Journal Classification (ASJC) (USD)⁶¹

Grande área	Área	F	Valor APC por área (USD)	% APC por Área (USD)	Média APC por área (USD)
Ciências Físicas	Ciência Ambiental	154	330.391,09	31,34	2.145,40
	Matemática	5	8.010,00	0,76	1.602,00
	Ciências da Terra e Planetárias	45	71.341,12	6,77	1.585,36
	Ciência dos Materiais	13	17.570,00	1,67	1.351,54
	Química	25	32.890,00	3,12	1.315,60
	Engenharia	50	64.851,96	6,15	1.297,04
	Ciência da Computação	17	20.881,81	1,98	1.228,34
	Engenharia Química	3	3.670,00	0,35	1.223,33
	Física e Astronomia	3	3.030,00	0,29	1.010,00
	Energia	1	1.010,00	0,10	1.010,00
Total por área		316	553.645,98	52,53	1.752,04
Ciências da Vida	Ciências Agrárias e Biológicas	93	168.365,62	15,97	1.810,38
	Bioquímica, Genética e Biologia Molecular	85	143.611,88	13,62	1.689,55
	Imunologia e Microbiologia	18	24.072,00	2,28	1.337,33
	Farmacologia, Toxicologia e Farmacêutica	6	7.860,00	0,75	1.310,00
	Neurociência	5	6.530,00	0,62	1.306,00
	Total por área	207	350.439,50	33,25	1.692,94
Ciências da Saúde	Veterinária	2	2.815,66	0,27	1.407,83
	Medicina	66	90.226,21	8,56	1.367,06
	Profissões de saúde	7	7.252,00	0,69	1.036,00
	Enfermagem	2	2.020,00	0,19	1.010,00
	Total por área	77	102.313,87	9,71	1.328,75
Ciências Sociais	Psicologia	8	16.462,00	1,56	2.057,75
	Ciências Sociais	11	14.710,00	1,40	1.337,27
	Ciências da Decisão	3	3.825,66	0,36	1.275,22
	Economia, Econometria e Finanças	5	5.730,00	0,54	1.146,00
	Negócios, Gestão e Contabilidade	5	5.050,00	0,48	1.010,00
	Artes e Humanidades	2	1.875,28	0,18	937,64
Total por área		34	47.652,94	4,52	1.401,56
Total Geral		634	1.054.052,29	100	1.662,54

Fonte: elaborado pela autora

Notas: Coluna F: refere-se à quantidade de *data papers* publicados por autores filiados às instituições brasileiras por área do conhecimento

Coluna Valor da APC: refere-se ao valor da APC cobrada pelo periódico no ano de 2025. Quando necessário os valores foram convertidos para dólares americanos (USD), em 07 de março de 2025, no Conversor de Moedas do Banco Central do Brasil, disponível em <https://www.bcb.gov.br/conversao>

Coluna % APC por Área (USD): refere-se à porcentagem acumulada pela área em relação ao valor total gasto com a APC

Coluna Média APC por área (USD): refere-se ao valor médio gasto com APC pela área

⁶¹ É importante ressaltar que essa estimativa considera os valores de APC do ano de 2025, com isso, pode haver distorções no valor gasto por área.

Pode-se verificar que o valor da taxa de APC para publicação de *data papers* custa em média US\$ 1.662,54. Esse valor é quase três vezes superior ao valor médio identificado por Walters (2020), que foi de US\$574,00, em 2020. Esse dado indica que os valores de APC dos *data papers* estão aumentando ao longo do tempo, o que pode indicar que eles estão acompanhando as mudanças globais no mercado de publicação científica, causadas principalmente pela agenda europeia de planos e tratados em prol do acesso aberto e pelas políticas estadunidenses, conforme Zhang *et al.* (2022).

Ainda em relação aos valores, a média geral identificada é superior ao valor de taxa de APC considerada justa pela *Fair Open Access Alliance*, de US\$1.000 (Fair Open Access Alliance, 2021) para a publicação de artigos originais, se comparado ao valor médio desta taxa para publicação de artigos pagos por autores filiados às instituições brasileiras, que varia entre US\$ 1.946,20 (Anselmo; Rodrigues; Mugnaini, 2022) e US\$ 2.868,00 (Neubert *et al.*, 2024), e ainda se equiparando ao valor médio identificado por Morrison *et al.* (2022) que foi de US\$ 1.626,00. Dessa forma, pode-se perceber uma tendência na equiparação dos custos desses dois tipos de documentos sem que efetivamente se observe uma diferença considerável no valor praticado pelos títulos em relação ao *data paper*.

As áreas das Ciências Físicas e Ciências da Vida são aquelas que abarcam os maiores custos com pagamento das APCs, concentrando respectivamente 52,53% e 33,25% do custo total. O alto custo com APCs nessas áreas, além de possuir uma relação direta com concentração de *data papers* publicados por autores filiados às instituições brasileiras, pode ter relação com o prestígio dos periódicos nos quais os *data papers* foram publicados, uma vez que revistas com longa história o tem incorporado como modalidade de publicação (Pérez-Luque; Ros-Candeira, 2019). Além disso, são áreas com forte influência dos *publishers* comerciais.

Apesar de a grande área das Ciências Sociais ter publicado menos *data papers* do que as outras, a média variou significativamente, tendo Psicologia com um dos valores médios mais altos, US\$ 2.057,75 e o mais baixo, com a área das Artes e Humanidades, US\$ 937,64. O que pode justificar as diferenças na variação dos custos de APC nessa grande área é a diversidade de objetos de pesquisa, cujas áreas possuem diferentes características. Além disso, pode-se apontar a criação de *data journals*, tais como o *Journal of Open Humanities Data*, lançado em 2015 e o *Research Data Journal for the Humanities and Social Sciences*, lançado em 2016,

bem como os recursos financeiros escassos que podem ser fatores que explicam o número reduzido de *data papers* publicados (McGillivray *et al.*, 2022; Wigdorowitz *et al.*, 2024).

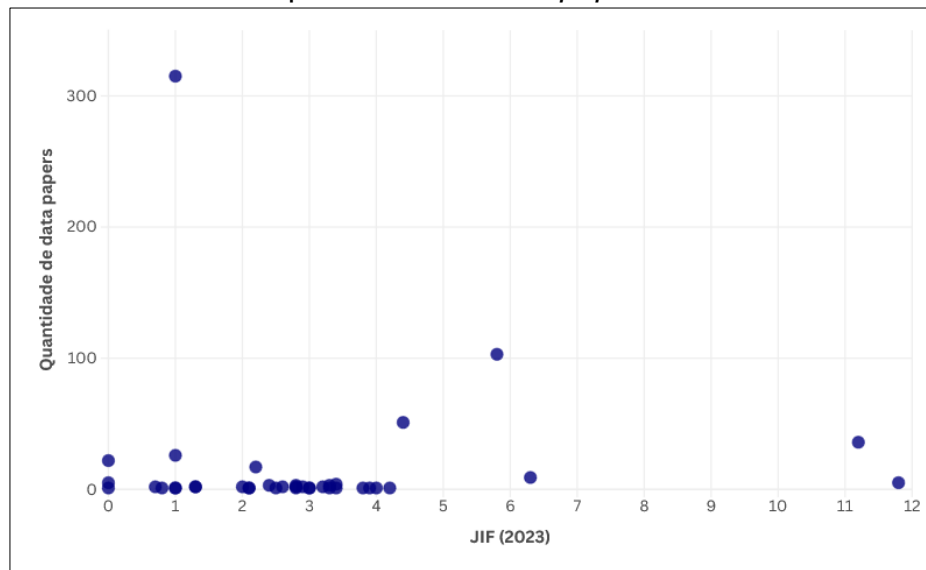
Björk e Solomon (2015) afirmam que normalmente, os periódicos Biomédicos tendem a ser mais caros para publicar artigos do que, por exemplo, periódicos das Ciências Sociais e Humanidades. Mas, curiosamente, nesta pesquisa essa afirmação não encontrou respaldo, embora os periódicos da grande área da Ciência da Saúde, tenham publicado mais *data papers* do que os da grande área das Ciências Sociais, sendo que o custo médio dos APCs dos primeiros foi menor, US\$ 1.328,75. Embora sejam necessários estudos mais aprofundados, esses dados podem indicar que entre essas áreas o reconhecimento dos periódicos que publicam *data papers* é desigual.

4.3.1 Avaliação dos periódicos a partir do Journal Impact Factor (JIF) e do SCImago Journal Rank (SJR) de 2023.

A avaliação visa demonstrar o impacto do periódico na área, mas é importante ressaltar que ela está limitada às citações dos documentos indexados nas bases de dados pertencentes aos seus desenvolvedores, o JIF, no caso da *Web of Science (Clarivate)* e o SJR, no caso da *Scopus (Elsevier)*. Além disso, eles também são influenciados pelo comportamento de publicação e citação das áreas do conhecimento. No Apêndice F, está disponível um quadro relacionando esses identificadores aos periódicos analisados.

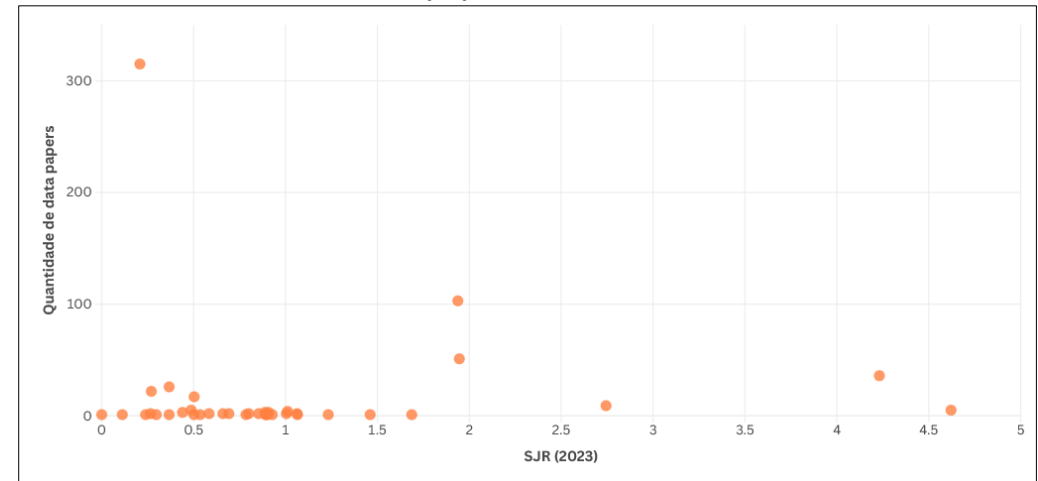
Tendo como objetivo verificar se há relação entre o impacto do periódico a partir dos indicadores mencionados e a quantidade de *data papers* publicados por eles, foram elaborados os Gráficos 5 e 6. O Gráfico 5 relaciona o Fator de Impacto (JIF) e a quantidade de *data papers* publicados e o Gráfico 6 relaciona o SJR e a quantidade de *data papers* publicados.

Gráfico 5 - Relação entre o Fator de Impacto (2023) e a quantidade de *data papers*



Fonte: elaborado pela autora

Gráfico 6 - Relação entre o SJR (2023) e a quantidade de *data papers*



Fonte: elaborado pela autora

O Fator de Impacto variou entre 0, que são os periódicos que não o possuem, que foi o caso do *BMC Research Notes*, do *Chemical Data Collections* e do *Journal of Open Humanities Data*, até 11.8, que foi o caso do *Gigascience* e 11.2 do *Earth System Science Data (Copernicus Publications)*.

No Gráfico 5 pode-se verificar pela concentração de círculos azuis que a quantidade de *data papers* não está necessariamente associada ao alto Fator de Impacto, pois apenas 7,89% (50) *data papers* foram publicados em três periódicos com Fator de Impacto entre 6 e 12, 24 periódicos com impacto entre 2 e 5.8, publicaram 32,49% (206) dos *data papers*, oito periódicos entre 1 a 0.8 publicaram 55,21% (350) dos *data papers* e três periódicos sem fator de impacto publicaram 4,42% (28).

O SJR é um indicador que mede a influência científica de periódicos a partir do número de citações recebidas. Ele variou entre 0, que são os periódicos que não possuem esse indicador, nesse caso o *Frontiers in Bioinformatics*, até o impacto de 4.621 do periódico *Gigascience* e 4.231 do periódico *Earth System Science Data (Copernicus Publications)*.

Conforme Gráfico 6, no caso do SJR, pode-se verificar pela concentração de círculos laranjas, que o impacto do periódico também não influenciou na quantidade de *data papers* produzidos pelos autores filiados às instituições brasileiras. Apenas três periódicos se concentram entre 4.621 a 2.744 no SJR e juntos publicaram 7,89% (50) *data papers*, nove periódicos se concentraram entre 1.945 a 1.005, e juntos publicaram 26,18% (166), 25 periódicos se concentraram entre 0.928 a 0.112, publicando 65,77% (417) *data papers* e um periódico não foi considerado no SJR, publicando apenas 0,16% (1) *data paper*.

Apesar de serem identificadores diferentes, pode-se observar que os periódicos que mais se destacaram em ambos foram os *data journals Gigascience (Oxford University Press)*, com o primeiro número publicado em 2012 e com 0,79% (5) dos *data papers* identificados na pesquisa, e o *Earth System Science Data (Copernicus Publications)*, publicado em 2009 e com 5,68% (36) dos *data papers* de autores filiados às instituições brasileiras.

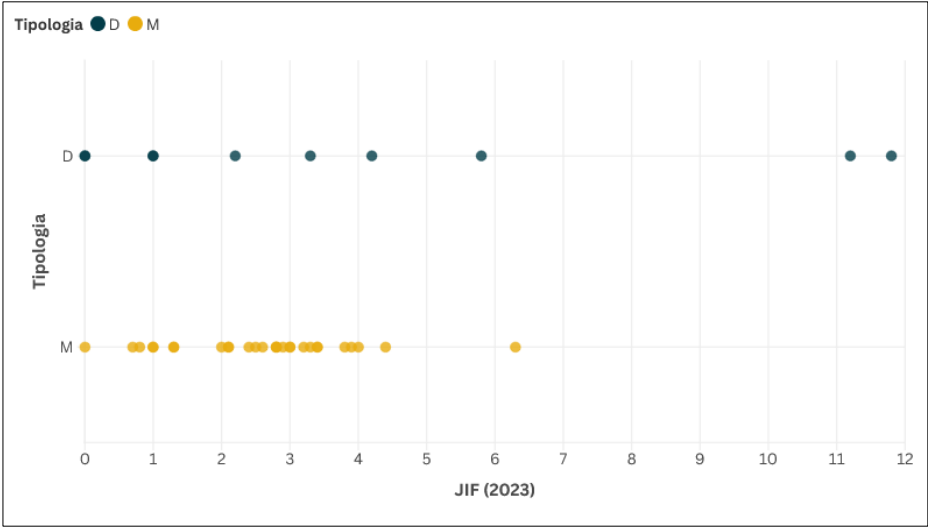
A partir desses dados, pode-se observar que a relação entre os indicadores JIF e o SJR e a quantidade de *data papers* publicadas, de maneira geral, não influenciou na seleção do periódico para publicar *data papers*, pois os dados indicam

que os *data papers* dos autores filiados às instituições brasileiras foram publicados em periódicos de impacto baixo a moderado nessa amostra.

Buscando identificar se há relação entre o impacto do periódico, a partir dos indicadores mencionados, e a tipologia do periódico, foram elaborados os Gráficos 7 e 8. O Gráfico 7 relaciona o Fator de Impacto no eixo X, à tipologia dos periódicos, *data journals* (D) e mistos (M), no eixo Y. O Gráfico 8 relaciona o SJR no eixo X à tipologia dos periódicos, *data journals* (D) e mistos (M), no eixo Y.

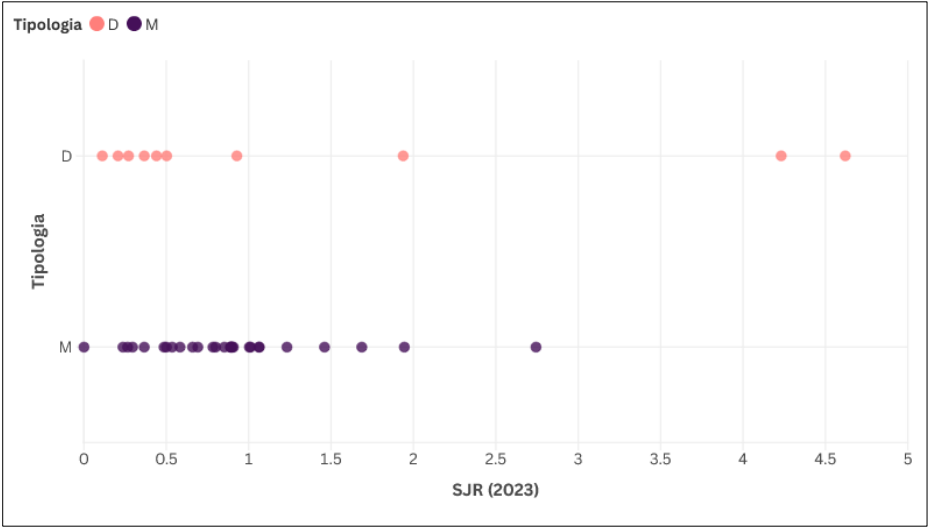
Em relação à tipologia dos periódicos, é importante ressaltar que a maioria 83,44% (529), dos *data papers* foram publicados por dez *data journals* e o restante dos *data papers*, 16,56% (105) foi publicado por 28 periódicos mistos (M).

Gráfico 7 - Relação entre o Fator de Impacto (2023) e a tipologia do periódico



Fonte: elaborado pela autora

Gráfico 8 - Relação entre o SJR (2023) e a tipologia do periódico



Fonte: elaborado pela autora

No Gráfico 7, ao relacionar a tipologia dos periódicos ao Fator de Impacto, é possível observar que houve uma variação entre 1 a 11.8 nesse indicador, sendo que oito dos dez *data journals* estão dispersos nessa variação, representados pelos círculos verdes, e dois deles não possuem esse indicador. Dois periódicos multidisciplinares concentraram 65,93% (418) dos *data papers* publicados o *Scientific Data (Springer)*, com Fator de Impacto 5.8, concentrou 16,25% (103) dos *data papers* e o *Data in Brief (Elsevier)*, com Fator de Impacto 1.0, concentrou 49,68% (349) dos *data papers*.

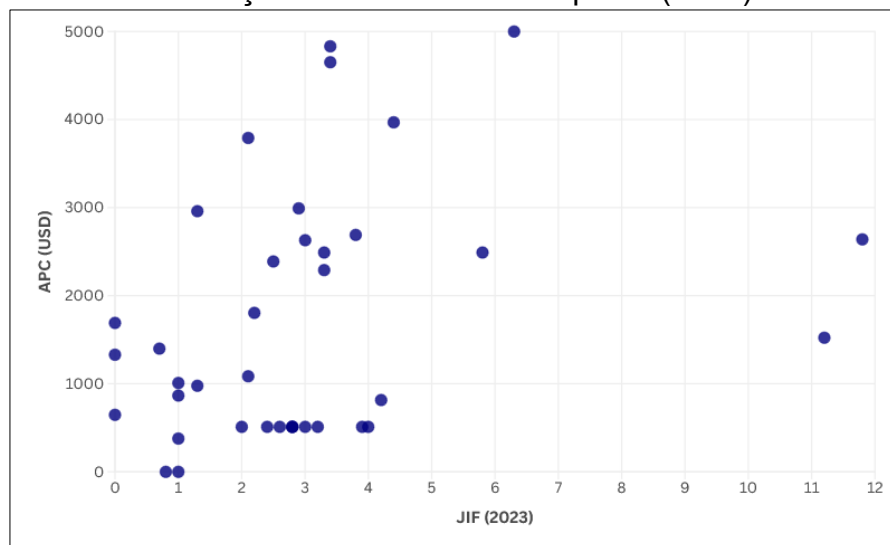
Dentre os periódicos mistos, representados pelos círculos amarelos, pode-se observar que há uma concentração nos valores baixos de JIF entre 0 a 4, com alguns pontos entre 5 e 6, o que pode indicar que os periódicos mistos que aderiram à publicação de *data papers* possuem um menor impacto nesse indicador.

No Gráfico 8, ao relacionar o SJR e a tipologia dos periódicos, foi possível observar que a variação entre os *data journals*, representados pelos círculos rosas, foi alta, entre 0.112 a 4.621. Apenas o *Data in Brief (Elsevier)*, classificado como periódico multidisciplinar nesse indicador, possui um dos menores impactos com 0.208, e uma maior concentração de *data papers*. Dentre os periódicos mistos, representados pelos círculos roxos, a variação foi menor se comparada aos *data journals*, entre 0.238 a 2.744, o que pode indicar um menor prestígio acadêmico, embora estes também tenham publicado uma quantidade menor de *data papers*.

Com isso, os dados permitem inferir que, dada a heterogeneidade entre os indicadores JIF e SJR, pode-se supor que eles não foram um fator preditivo na seleção dos autores filiados às instituições brasileiras para publicarem seus *data papers*. Porém o quantitativo de *data papers* publicados em *data journals* sugere que essa tipologia pode ser um fator preditivo pelos autores.

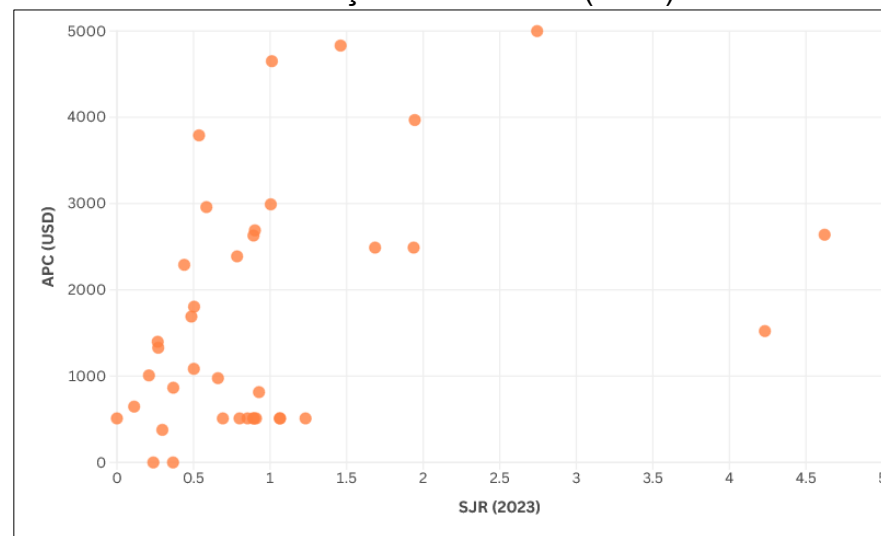
Tendo como objetivo identificar se há relação entre o impacto do periódico, a partir dos indicadores mencionados, e a as taxas de APC, foram elaborados os Gráficos 9 e 10. O Gráfico 9 relaciona o Fator de Impacto, no eixo X, à faixa de valor do APC, no eixo Y. O Gráfico 10 relaciona o SJR, no eixo X, à faixa de valor do APC, no eixo Y.

Gráfico 9 - Relação entre o Fator de Impacto (2023) e a APC



Fonte: elaborado pela autora

Gráfico 10 - Relação entre o SJR (2023) e a APC



Fonte: elaborado pela autora

O Gráfico 9 buscou relacionar o Fator de Impacto à taxa de APC, com base na dispersão dos pontos azuis, sendo possível observar que cinco periódicos entre que aqueles que não possuíam JIF e aqueles com impacto de 0.8, a taxa de publicação variou entre 0 a \$1.690,00 dólares, esses concentraram, 4,89% (31) dos *data papers*. A maior parte dos periódicos (29) se concentrou entre o Fator de Impacto 1 a 4 e entre esses periódicos o valor da APC variou entre aqueles que não cobram a taxa de publicação até aqueles em que as taxas foram de \$4.830,00 dólares, esses concentraram 62,78% (398) *data papers*. Nos periódicos que concentraram o JIF entre 4.2 a 11.8, as taxas de APC variaram entre \$815,00 a \$5.000,00 dólares e esses periódicos publicaram 32,33% (205) dos *data papers*.

Os dados indicam que os autores filiados às instituições brasileiras tendem a publicar os *data papers* em periódicos com Fator de Impacto de baixo a médio, com taxas de APC entre zero a \$4.830,00 dólares.

No Gráfico 10, que buscou relacionar o SJR e a taxa de APC, com base na dispersão dos círculos laranjas, pode-se verificar que a publicação se concentrou em 26 periódicos que concentram 65,93% (418) dos *data papers* publicados. Entre eles o SJR variou entre aqueles que não possuíam até 0.928 e a cobrança do APC variou entre aqueles que não cobram APC ao valor de \$3.790,00 dólares. Em dez periódicos, que concentraram 27,60% (175) *data papers*, o SJR variou de 1.005 a 2.744 e o APC variou de \$512,00 a \$5.000,00 a dólares. Dois periódicos possuem SJR de 4.231 a 4.621, possuem taxa de APC de \$1.521,66 a \$2.638,00 dólares, esses concentraram 6,47% (41) dos *data papers*.

Os dados indicam que os autores filiados às instituições brasileiras tendem a publicar os *data papers* em periódicos com SJR baixo e com taxas de APC no valor de zero a \$3.790,00 dólares.

Embora esta análise precise ser complementada entre esses indicadores e a sua área específica, de maneira geral os dados indicam que as maiores quantidades de *data papers* foram publicadas entre periódicos que não possuíam os indicadores analisados a indicadores de valor médio e o custo de publicação aproximado a \$5.000 dólares. Com isso, pode-se inferir que no caso desta pesquisa, o custo de publicação não parece estar diretamente relacionado ao impacto do periódico, logo esse não parece ser um fator que motivou a seleção dos periódicos pelos autores filiados às instituições brasileiras para a publicação dos *data papers*.

Assim como no caso da publicação dos artigos tradicionais, no caso dos *data papers* a reputação parece ser a *commodity* que leva os pesquisadores a selecionar uma determinada revista, e os editores definem sua taxa APC (Budzinski *et al.*, 2020). Com isso as editoras comerciais conseguem facilmente se apropriar do produto final do trabalho do pesquisador que busca publicar seu artigo nos periódicos de maior prestígio, que possuem maior fator de impacto, pois o sistema de comunicação científica e o produtivismo imposto aos pesquisadores são, portanto, os principais mecanismos que entram em ação quando se trata de induzir os pesquisadores a buscar o almejado prestígio e reconhecimento acadêmicos, tornando-os dependentes dessa indústria editorial (Weitzel, 2022).

Todavia, no caso desta pesquisa, de maneira geral, os dados sugerem que esses indicadores não interferiram na seleção dos periódicos para publicar os *data papers*. Embora o prestígio seja importante, a quantidade de *data papers* publicadas pelos autores filiados às instituições brasileiras está mais ligada à tipologia da revista do que ao seu impacto. Em relação aos periódicos mistos, cabe ressaltar que as citações e o prestígio da revista não estão ligados apenas aos *data papers*, uma vez que nesses casos eles são influenciados pelas taxas de citação dos artigos científicos.

4.3.1.1 Avaliação dos periódicos a partir do Qualis Periódicos (2017-2020)

No caso do Brasil, a relevância da produção científica, está atrelada à avaliação do Sistema Nacional de Pós-Graduação, na qual a entrada dos novos cursos e a sua permanência são aferidas a partir de critérios definidos em regulamentações mais amplas que orientem a avaliação em nível nacional. No entanto, cada área do conhecimento tem alguma liberdade para determinar critérios e indicadores específicos em suas análises (Brasil, 2023b; CAPES, 2021).

Diante da necessidade de considerar a avaliação dos programas *strictu sensu* e considerando a qualidade dos veículos utilizados pelos pesquisadores para divulgar os resultados das suas pesquisas, surge o Qualis Periódicos, que foi instituído pela CAPES, em 1998, como forma de classificar os periódicos científicos nos quais era divulgada a produção científica de docentes e discentes dos programas de pós-graduação brasileiros (Barata, 2023).

O Qualis consiste na qualificação da produção intelectual a partir da análise da qualidade dos periódicos de acordo com as áreas do conhecimento, resultando numa lista com a classificação dos veículos utilizados pelos programas de pós-graduação para a divulgação da sua produção (CAPES, 2022; 2023). Essa listagem é sempre um retrato a posteriori, uma vez que é referente aos anos anteriores, cujos dados já foram informados para a CAPES por meio da Plataforma Sucupira (CAPES, 2022)

Essa classificação se dá a partir do enquadramento dos periódicos em estratos indicativos da qualidade, sendo A1 o mais elevado, seguido em ordem decrescente de A2, A3, A4, B1, B2, B3, B4 e C, tendo esse último peso zero na avaliação. A classificação única é dada pela área mãe em que o periódico possua maior uso, ou seja, aquela que mais registrou publicações entre os anos de 2013 a 2019, no caso desta pesquisa que utiliza o Qualis do quadriênio 2017-2020. Com essa definição, são atribuídos os valores a partir dos indicadores bibliométricos das bases de dados e, em seguida, são realizados cálculos que classificam os periódicos nos estratos mencionados anteriormente (CAPES, 2022).

No Quadro 12 é apresentada a estratificação dos periódicos, considerando os *data journals* e os periódicos mistos que publicaram os *data papers* de autores filiados às instituições brasileiras, a partir da área mãe.

No entanto, é preciso ressaltar que no caso dos periódicos mistos foram considerados não apenas pela publicação de *data papers*, mas sim, por todos os artigos científicos publicados pelos docentes e discentes. Logo, essa estratificação não pode ser generalizável para a publicação de dados de pesquisa, pois a unidade considerada pelos instrumentos de avaliação é o periódico e não a unidade artigo científico ou outras tipologias de documentos publicadas por ele.

Quadro 12 - Estratificação dos periódicos no Qualis (2017-2020)

Área Mãe	Periódico	Estrato	F*	Tipologia	Valor APC (USD)
Biodiversidade	Ecology (Wiley)	A1	51	M	3.970,00
	Global Ecology and Biogeography (Wiley)	A1	9	M	5.000,00
	Gigascience (Oxford University Press)	A1	5	D	2.638,00
	Journal of Biogeography (Wiley)	A1	1	M	4.830,00
	Frontiers in Marine Science (Frontiers Media)	A2	3	M	512,00
	Taxon (International Association for Plant Taxonomy)	A2	1	M	2.630,00

Área Mãe	Periódico	Estrato	F*	Tipologia	Valor APC (USD)
	Frontiers in Ecology and Evolution (Frontiers Media)	A3	3	M	512,00
	Marine Genomics (Elsevier)	A3	2	M	2.960,00
	Comparative cytogenetics (Pensoft Publishers)	A3	1	M	378,56
	Biodiversity Data Journal (Pensoft Publishers)	A4	26	D	865,28
	Data (MDPI)	A4	17	D	1.805,66
	ZooKeys (Pensoft Publishers)	A4	2	M	978,21
	Ecosistemas (Asociacion Espanola de Ecologia Terrestre)	B1	1	M	-
Ciências Agrárias I	Annals of forest Science (Elsevier)	A2	1	M	2.390,00
Ciências Biológicas I	Scientific Reports (Springer)	A1	1	M	2.690,00
	Frontiers in Genetics (Frontiers Media)	A4	2	M	512,00
Ciências Biológicas II	Proteomics (Wiley)	A2	4	M	4.650,00
	Frontiers in Molecular Biosciences (Frontiers Media)	A2	1	M	512,00
	Frontiers in Neuroscience (Frontiers Media)	A3	2	M	512,00
Ciências Biológicas III	Frontiers in Microbiology (Frontiers Media)	A2	1	M	512,00
	Microbiology Resource Announcements (American Society for Microbiology)	B3	2	M	1.400,00
Computação	BMC Bioinformatics (Elsevier)	A1	2	M	2.990,00
Engenharias I	Earth System Science Data (Copernicus Publications)	A1	36	D	1.521,66
	Big Earth Data (Taylor and Francis)	B1	1	D	815,00
Farmácia	Chemical Data Collections (Elsevier)	B2	22	D	1.330,00
Geociências	Frontiers in Earth Science (Frontiers Media)	A2	2	M	512,00
Medicina I	Human Mutation (Wiley)	A2	1	M	2.490,00
	BMC Research Notes (Elsevier)	B1	5	M	1.690,00
Psicologia	Frontiers in Psychology (Frontiers Media)	A1	2	M	512,00
Saúde Coletiva	Frontiers in Public Health (Frontiers Media)	A1	1	M	512,00
Sociologia	Cahiers Agricultures (EDP Sciences)	A2	1	M	-
Sem Qualis	Data in Brief (Elsevier)	-	315	D	1.010,00
	Scientific Data (Springer)		103	D	2.490,00
	Geoscience Data Journal (Wiley)		3	D	2.290,00
	Frontiers in Bioinformatics (Frontiers Media)		1	M	512,00
	Journal of Open Humanities Data (Ubiquity Press)		1	D	646,15
	Nature Conservation (Pensoft Publishers)		1	M	1.086,00
	Science of Nature (Springer)		1	M	3.790,00

Fonte: elaborado pela autora

Notas: Coluna estrato: refere-se à estratificação atribuída pelas áreas da CAPES

Coluna F*: refere-se à quantidade de *data papers* publicados no periódico por autores filiados às instituições brasileiras

Coluna Tipologia: refere-se ao tipo de periódico M (misto) e D (data journal)

Coluna Valor APC (USD): refere-se ao valor da APC cobrada pelo periódico no ano de 2025

Pode-se observar que sete periódicos, dentre eles quatro *data journals* (D) não foram considerados pelas áreas na avaliação do período 2017-2020. Esses

periódicos concentraram a publicação de 67,03% (425) dos *data papers*, com taxas da APC variando entre \$512,00 a \$2.490,00 dólares. Foram considerados 31 periódicos na avaliação, os quais concentram apenas 32,97% (209) dos *data papers* publicados.

Dos 31 periódicos que foram classificados pelos Qualis, nove foram classificados nos estratos A1 e A2, concentrando respectivamente 17,03% (108) e 2,37% (15) dos *data papers*, oito estão concentrados nos estratos A3 e A4, concentrando 1,26% (8) e 7,41% (47) *data papers* respectivamente, três periódicos no estrato B1 com 1,10% (7) dos *data papers* e dois concentrados em B2 e B3, concentrando respectivamente 3,47% (22) e 0,32% (2) dos *data papers*.

Os dados indicam que os *data papers* publicados pelos autores filiados às instituições brasileiras estão sendo publicados em periódicos localizados nos estratos A1 e A2, o que sugere que os autores consideram os periódicos mais reconhecidos nas suas áreas para publicar esses documentos, sendo que isso pode ser um fator preditivo. Contudo, é importante ressaltar que a classificação dos periódicos no Qualis é realizada a posteriori.

Ressalta-se que nem sempre esses periódicos são *data journals*, pois dos 18 periódicos concentrados, nos estratos A1 e A2 apenas dois são periódicos destinados a publicar exclusivamente dados de pesquisa, que são o *Gigascience* e o *Earth System Science Data (Copernicus Publications)*, sendo o primeiro considerado pela Biodiversidade e o segundo pela área da Engenharia 1. Logo, nesse caso, o Qualis parece ser um fator preditivo em detrimento da tipologia do periódico.

Ao analisar as áreas, a Biodiversidade foi a que mais considerou periódicos que publicam *data papers*, sendo 13 periódicos no total, que variaram entre os Qualis A1 a B1. Do total de periódicos, três são *data journals* e concentraram a publicação de 48 *data papers*, sendo o *Gigascience*, *Data* e *Biodiversity Data Journal (Pensoft Publishers)*. Essa área e a Odontologia, foram as únicas que consideraram o *data paper* em seu documento de área para no Qualis do quadriênio 2017-2020 (Silveira; Dias; Neubert, 2023a).

Dentre os periódicos estratificados no Qualis e cuja taxa de APC é superior a \$2.000 dólares, o que corresponde a 11 periódicos, todos estão concentrados nos estratos A, sendo dez mistos e um *data journal*. Com isso, pode-se dizer que o valor do APC pode ser influenciado pelo prestígio do periódico, conforme Neubert *et al.* (2024).

Embora os *data journals* pareçam ser um fator preditivo para a publicação dos *data papers* pelos autores filiados às instituições brasileiras, conforme seção 4.3,1, no caso da avaliação da CAPES, os *data journals* ainda carecem de consolidação no sistema de avaliação dos canais de publicação na pós-graduação brasileira, uma vez que dos dez identificados quatro não possuíam Qualis e esses concentraram a publicação de 66,56% (422) dos *data papers* publicados pelos referidos autores.

Por fim, considerando a emergência dos dados de pesquisa na comunicação científica, que vem culminando para a criação de outros tipos de documentos, dentre eles os *data papers*, é preciso refletir se o sistema de avaliação da produção científica brasileira está preparado para considerá-los nos processos de avaliação, uma vez que os pesquisadores vêm utilizando esses canais para divulgar seus dados de pesquisa e atrelando-os à obtenção de reconhecimento acadêmico.

As atualizações previstas para a avaliação da permanência dos programas de pós-graduação brasileira, do período de 2025-2028, a ser implementada em 2029, passarão a considerar ações que envolvem a Ciência Aberta, dentre elas o compartilhamento de dados de pesquisa, nas dimensões que envolvem a avaliação dos programas (CAPES, 2025). Com isso, espera-se que haja algum impacto na publicação dos dados de pesquisa por meio de *data papers* e *data journals*.

4.3.2 O processo de revisão por pares

A disponibilidade de dados ao público está alinhada ao movimento da Ciência Aberta. No entanto, conjuntos de dados depositados em repositórios e bases de dados públicos podem não passar por verificações de qualidade e com isso, torna-se importante verificar o processo de revisão por pares realizado pelos periódicos, a fim de que esses possam contribuir para a garantia da confiabilidade dos dados (Wang; Xu, 2025). Por isso, para discutir sobre as tipologias de revisões realizadas pelos periódicos que publicaram os *data papers* dos autores filiados às instituições brasileiras, foi elaborado o Quadro 13, no qual são apresentados os periódicos, os tipos de revisões adotados e os aspectos revisados.

Quadro 13 - Tipo de revisão por pares e aspectos revisados

Nome do periódico	Tipo	F	Tipo de revisão					Aspectos revisados			
			Simples Cego	Duplo cego	Aberta	A critério do revisor*	Combinado	Data paper e data set	Data paper	Data set	Não informa
Data In Brief	D	315	X					X			
Scientific Data	D	103		X					X		
Ecology	M	51	X								X
Earth System Science Data	D	36			X			X			
Biodiversity Data Journal	D	26				X		X			
Chemical Data Collections	D	22	X								X
Data	D	17	X					X			
Global Ecology and Biogeography	M	9		X							X
BMC Research Notes	M	5				X					X
Gigascience	D	5				X		X			
Proteomics	M	4	X								X
Frontiers in Ecology And Evolution	M	3					X				X
Frontiers in Marine Science	M	3					X				X
Geoscience Data Journal	D	3	X								X
BMC Bioinformatics	M	2				X					X
Frontiers in Earth Science	M	2					X				X
Frontiers in Genetics	M	2					X				X
Frontiers in Neuroscience	M	2					X				X
Frontiers in Psychology	M	2					X				X
Marine Genomics	M	2	X								X
Microbiology Resource Announcements	M	2	X								X
Zookeys	M	2	X								X
Annals of Forest Science	M	1		X							X
Big Earth Data	D	1		X					X		
Cahiers Agricultures	M	1		X							X
Comparative Cytogenetics	M	1				X					X
Ecosistemas	M	1		X							X
Frontiers in Bioinformatics	M	1					X				X
Frontiers in Microbiology	M	1					X				X
Frontiers in Molecular Biosciences	M	1					X				X
Frontiers in Public Health	M	1					X				X
Human Mutation	M	1	X								X

Nome do periódico	Tipo	F	Tipo de revisão					Aspectos revisados			
			Simplex Cego	Duplo cego	Aberta	A critério do revisor*	Combinado	Data paper e data set	Data paper	Data set	Não informa
Journal of Biogeography	M	1	X								X
Journal of Open Humanities Data	D	1	X						X		
Nature Conservation	M	1	X							X	
Science of Nature	M	1	X								X
Scientific Reports	M	1	X								X
Taxon	M	1	X								X
Total Geral		634	16	6	1	5	10	5	3	1	29

Fonte: elaborado pela autora

Notas: Coluna Tipo: refere-se à tipologia do periódico Data journal (D) ou Periódico Misto (M)

Coluna F: refere-se à quantidade de *data papers* publicados pelo periódico, no qual pelo menos um autor é filiado à uma instituição brasileira

Colunas Tipo de revisão: refere-se ao tipo de revisão informada pelo periódico

A critério do revisor*: quando a avaliação pode ser aberta ou fechada, dependendo do revisor decidir abrir o nome e/ou o parecer

Colunas aspectos revisados: refere-se à declaração do periódico sobre o que é revisado

Todos os periódicos informaram que realizam algum tipo de revisão por pares. Ainda, Pinto (2022) na amostra que verificou a revisão por pares em *data papers* na área da Enfermagem, identificou que dos 12 periódicos analisados, apenas um não realizava a revisão por pares. Esses dados podem indicar que a revisão dos *data papers* por partes dos periódicos vem seguindo as práticas da comunicação científica já consolidada.

O movimento da Ciência Aberta tem se esforçado para aumentar a transparência da produção do conhecimento científico e para tornar os produtos da investigação científica mais amplamente disponíveis. A partir disso, estão sendo realizados esforços para tornar a revisão por pares aberta, onde aspectos do processo de revisão, que tradicionalmente foram ocultos ou anônimos, são tornados públicos, como a divulgação dos nomes dos revisores e/ou seus pareceres (Wolfram *et al.*, 2020).

No entanto, embora haja um movimento a favor da abertura da revisão por pares (Candela *et al.*, 2015; Walters, 2020; Wang; Xu, 2025), observou-se que em 57,9% (22) dos casos, ou seja, na maioria dos periódicos identificados na amostra, segue-se o padrão convencional de avaliação, na qual uma das partes ou ambas não é conhecida. Ainda em 42,11% (16) dos periódicos a revisão adotada é a simples cego e em 15,79% (6) a revisão adotada é a duplo cego. Quando analisados apenas os *data journals*, esse padrão de revisão convencional se mantém, o que condiz com os resultados encontrados por Seo e Kim (2020) e Wang e Xu (2025).

Foi identificado também que em 13,16% (5) dos periódicos o revisor pode decidir pela abertura ou não dos nomes e/ou dos pareceres, o que sugere que a adoção e a aceitação da revisão por pares aberta por autores e revisores ainda são limitadas (Wang; Xu, 2025).

Em 28,95% (11) dos casos, os periódicos adotam práticas abertas e/ou combinadas na revisão por pares. Dentre as práticas combinadas destacam-se os periódicos do *publisher Frontiers*, que realizam a revisão dos *data papers* em duas etapas, sendo que na primeira é realizada a revisão simples cego e a segunda é aberta para a comunidade participar do processo de revisão.

Dessa forma, os dados coadunam com os resultados identificados por Wang e Xu (2025), na qual a maioria dos periódicos de dados utiliza um modelo único e anônimo para revisão por pares, enquanto a revisão transparente e aberta representa apenas uma porcentagem relativamente pequena.

Candela *et al.* (2015) destacam que a divulgação do processo de revisão deve ser incentivada como um meio de comprovar a qualidade do *data paper*. A revisão por pares aberta dos *datasets* e dos *data papers* pode contribuir para a confiabilidade e a replicabilidade da pesquisa, uma vez que os problemas com dados podem ser descobertos depois de estarem disponíveis para uso externo, quando artigos subsequentes tentam replicar ou verificar achados específicos. Nesse sentido, o verdadeiro valor de um conjunto de dados nem sempre será estabelecido após a publicação, mas pode dar-se a partir da ampla distribuição e de um período de tempo suficiente de uso (Mayernik *et al.*, 2015).

Ainda que todos os periódicos identificados na amostra informaram que realizam a revisão por pares dos *data papers*, apenas 23,68% (9) informam nas suas políticas que realizam a revisão desses documentos e/ou dos *datasets*. Os demais, 76,32% (29), não descrevem critérios específicos para avaliar os *data papers* e seus dados adjacentes, isto é, os periódicos não deixam claro se os *data papers* seguem os mesmos critérios de avaliação dos artigos tradicionais. Essa ausência de políticas pode prejudicar a transparência e a confiabilidade dos *data papers* e dos *datasets*.

Kong *et al.* (2019) afirmam que a revisão por pares em *data journals*, é mais complexa do que nos periódicos tradicionais, uma vez que envolve revisão dos *data papers*, dos conjuntos de dados e sua correlação num processo único. Essa complexidade pode justificar o porquê apenas cinco *data journals* dos 38 periódicos identificados informaram realizar a revisão dos *data papers* e *datasets* de maneira conjunta.

O ponto central para o sucesso dos *data papers* é conduzir uma revisão por pares de alta qualidade, de modo, que seja equivalente aos formatos de revisões dos artigos de pesquisa. A falha em garantir que os objetivos e a descrição da amostra que está sendo relatada no *data paper*, não fique clara, prejudicará o processo de compartilhamento de dados e a credibilidade dos *data papers*, por isso, se um revisor não puder avaliar diretamente os dados preparados pelo autor e verificar se a amostra compartilhada está pronta para o compartilhamento, o reuso ficará comprometido (Gorgolewski; Margulies; Milham, 2013).

Em relação à divulgação dos critérios utilizados para a revisão, apenas oito periódicos descreveram tal informação e esse número coincide com o quantitativo identificado por Wang e Xu (2025). Embora o periódico *Nature Conservation* tenha informado que realiza a revisão apenas dos *datasets*, os critérios não foram descritos

na sua política de revisão por pares. Esse resultado pode estar atrelado ao estágio ainda inicial da revisão por pares dos *data papers* e, especialmente dos próprios dados, onde não há um processo bem definido (Kong *et al.*, 2019; Mayernik *et al.*, 2015). Somado a isso os processos e softwares para avaliação de dados, que muitas vezes são específicos para tipos de dados, projetos experimentais e sistemas e, por isso, poucos revisores seriam qualificados para revisar todos os aspectos de um conjunto de dados (Mayernik *et al.*, 2015).

Dos nove periódicos que informam nas suas políticas que realizam a revisão desses documentos e/ou dos *datasets*, apenas sete informaram os critérios utilizados, e desses todos são *data journals*. Esses dados podem indicar que o tipo de periódico também segue as práticas de revisão adotadas pelos periódicos tradicionais, embora os critérios estejam voltados para as características dos *data papers* e as dimensões de avaliação dos dados de pesquisa (Wang; Xu, 2025).

A partir dos critérios identificados nos *data journals* e com base na síntese de Seo e Kim (2020) foi elaborado o Quadro 14, no qual estão especificados os critérios utilizados, tanto para *data papers* como para *datasets*, pois observou-se que os *data journals* definem critérios diferentes para cada parte avaliada, embora sejam complementares no processo de revisão como um todo.

Quadro 14 - Critérios identificados para avaliação dos *data papers* e *datasets*

Critérios relacionados ao dataset	Data In Brief*	Scientific Data	Earth System Science Data	Biodiversity Data Journal	Data	Gigascience	Journal of Open Humanities Data**
Metadados			X	X	X	X	
Adequação metodológica				X	X	X	
Formato dos dados acessível	X		X	X	X		
Valores e originalidade dos dados	X			X			
Identificador					X		
Disponibilidade dos dados							
Adequação do repositório				X		X	X
Licença				X	X	X	X

Crítérios relacionados ao data paper	Data In Brief*	Scientific Data	Earth System Science Data	Biodiversity Data Journal	Data	Gigascience	Journal of Open Humanities Data**
Contextualização dos dados	X	X	X	X	X		
Descrição precisa da metodologia	X	X	X	X	X	X	X
Fonte de erros				X	X		
Referências utilizadas			X				
Potencial de reutilização	X		X	X		X	X
Disponibilidade dos dados	X		X	X		X	
Repositórios	X						
Licença							

Fonte: elaborado pela autora a partir da adaptação de Seo e Kim (2020)

Nota:*Os critérios são centrados no conjunto de dados, embora informe que avalie tanto o conjunto de dados quanto o *data paper*.

**Embora informe que avalie apenas o *data paper* nos critérios, requer informações específicas que estão atreladas ao depósito dos dados.

No Quadro 14 é possível observar que os critérios visam cobrir aspectos referentes aos *datasets* e aos *data papers*, mas é possível identificar uma predominância deles nos *data papers*. A quantidade de critérios não preenchidos pelos periódicos, sugere que falta uma padronização nos critérios de avaliação a serem adotados para os *data papers* e *data sets*, apesar de que essas diretrizes possam variar devido a diferenças entre editoras e disciplinas (Wang; Xu, 2025).

Nos *data papers*, a descrição dos dados é o objetivo primordial, isso fica explícito quando verificado que esse quesito é avaliado por todos os periódicos. Portanto, os *data journals* dão ênfase à revisão dos processos relacionados ao tratamento e processamento dos resultados da coleta de dados (Lee; Kim, 2021).

Esses dados vão ao encontro do que foi observado por Seo e Kim (2020), onde as ênfases comuns nos critérios de revisão por pares de dados foram a adequação da metodologia de produção de dados e uma descrição detalhada da metodologia. Além da metodologia, outro critério que recebeu destaque pelos periódicos, foi a contextualização dos dados, referindo-se a onde os dados foram coletados, como a cobertura geográfica e temporal e se foram coletados a partir de outros bancos de dados.

Esses critérios são particularmente importantes para facilitar a reprodutibilidade da pesquisa e a reutilização de dados (Seo; Kim, 2020; Wang; Xu, 2025). Não é por acaso que o critério relacionado à reutilização dos dados, foi o segundo mais indicado pelos periódicos. No entanto, Jiao e Darch (2025) ao

analisarem os pareceres abertos de revisão dos *data papers* do periódico *Earth System Science Data (Copernicus Publications)*, observaram que apenas um pequeno número de revisores comentou sobre a utilidade potencial do conjunto de dados.

A avaliação de um *data paper* não pode se concentrar apenas na qualidade do manuscrito em si, mas precisa ir além. Precisa ser avaliada a qualidade dos dados a seleção dos repositórios e a garantia de proteção e arquivamento dos dados a longo prazo, embora isso seja menos frequente (Le Fournier; Schöpfel, 2025). Nesta pesquisa apenas três periódicos verificam a adequação do repositório à comunidade.

A avaliação dos *datasets* recebeu menos atenção pelos periódicos, sendo que apenas o *Biodiversity Data Journal (Pensoft Publishers)* foi o que mais atendeu aos requisitos relacionados a eles. Assim, é preciso que os periódicos considerem os *datasets* como parte essencial do processo de avaliação por pares, estabelecendo diretrizes claras sobre como realizar uma revisão de dados e quais características de um conjunto de dados devem ser examinadas (Mayernik *et al.*, 2015). Os conjuntos de dados são o sentido da publicação dos dados e se a avaliação não ocorre de maneira transparente, a credibilidade e a confiabilidade dos dados ficam abaladas, o que prejudica seu reúso por outros pesquisadores.

Para Wang e Xu (2025) a revisão por pares desempenha um papel fundamental no estabelecimento da credibilidade e da transparência dos *data papers*, ao mesmo tempo em que aprimora sua qualidade para enfatizar efetivamente a importância e o valor de reutilização dos dados. Portanto, analisá-la pode contribuir para que os autores dos *data papers* melhorem seus manuscritos e conjuntos de dados, além de garantir a sua confiabilidade e atração de mais autores para a publicação dos seus dados.

4.3.3 Considerações da seção

Esta pesquisa identificou que os 634 *data papers* publicados pelos autores filiados às instituições brasileiras, estão concentrados em 38 periódicos, dos quais dez são *data journals* e 28 são periódicos mistos. Embora os periódicos mistos estejam em maior número, cerca de 84% dos *data papers* se concentram em *data journals*.

Esse tipo de periódico vem crescendo em termos de criação de títulos e na quantidade de *data papers* publicados a partir de 2010, enquanto dentre os periódicos mistos a concentração da publicação *data papers* tenha tido um maior impacto entre 2015 e 2017.

Dentre as características dos periódicos, três são publicados por associações científicas, um por universidade e os demais (34) por *publishers* comerciais, ou seja, 98,58% (625) dos *data papers* foram publicados por esses editores, sendo que os cinco que mais se destacaram foram: *Elsevier* com 54,73% (347), *Springer Nature* com 16,56% (105), *Wiley* com 10,88% (69), *Copernicus Gesellschaft MBH* com 5,68% (36) e a *Pensoft Publishers* com 4,73% (30).

Todos os periódicos, onde os *data papers* foram publicados estão localizados fora do Brasil, sendo que 35 estão localizados em países do continente europeu e três nos Estados Unidos. Assim, foi possível observar que os periódicos estrangeiros têm sido uma saída para o déficit nacional na publicação de periódicos que aceitam *data papers* e somado a isso, os periódicos multidisciplinares se destacaram como canais de publicação dos *data papers*.

A cobrança de taxas APC por periódicos da via Ouro foi o modelo de negócio dominante entre os periódicos de acesso aberto que publicam *data paper* dos autores filiados às instituições brasileiras e esses dados coadunam com o encontrado em pesquisas de âmbito mundial. Em relação ao valor da APC dos *data papers*, observou-se que ele foi crescendo ao longo dos anos, estando quase equiparado aos valores cobrados pelos artigos de pesquisa.

As áreas das Ciências Físicas e Ciências da Vida, são as que abarcam o maior montante do custo com pagamento das APCs e, também, concentram a maior quantidade de *data papers* publicados. Além disso, os periódicos dessas áreas têm incorporado os *data papers* como modalidade de publicação, tendo forte influência dos *publishers* comerciais. As Ciências Sociais, embora tenham publicado menos *data papers*, foi a área onde a média do APC mais variou entre as áreas do conhecimento.

Foi verificado, também o impacto dos periódicos que publicaram os *data papers* a partir do JIF e do SJR. Esses indicadores variaram bastante e os dados sugerem que eles não parecem ser um fator que motivou a seleção dos periódicos pelos autores filiados às instituições brasileiras para a publicação dos *data papers*, mas sim a tipologia do periódico, uma vez que os *data journals*, principalmente aquele

publicado pelo *Data in Brief* (Elsevier) que tiveram maior concentração de *data papers*.

Em relação ao Qualis, embora os *data journals* pareçam ser um fator preditivo para a publicação dos *data papers* pelos autores filiados às instituições brasileiras, sete dos dez *data journals* identificados na pesquisa não o possuíam, o que indica que eles ainda carecem de consolidação no sistema de avaliação dos canais de publicação da pós-graduação brasileira.

Cerca de 20% dos *data papers* foram publicados em periódicos estratificados entre A1 e A2, o que sugere que autores consideram os periódicos mais reconhecidos nas suas áreas para publicar esses documentos. A área da Biodiversidade foi a que mais considerou periódicos que publicam *data papers*, o que vai ao encontro às temáticas identificadas nos *data papers*, além de ser a área onde começaram as primeiras iniciativas de compartilhamento dos dados de pesquisa.

Em relação ao processo de revisão por pares, a tipologia predominante foi a combinada, realizada pelos periódicos dos *publisher Frontiers*, que realizam a revisão dos *data papers* em duas etapas, sendo que na primeira é realizada a revisão simples cego e a segunda é aberta para a comunidade participar do processo de revisão. Ainda assim, observou-se que predominam os modelos tradicionais de revisão por pares, o simples cego e o duplo cego. Todos os periódicos informam que realizam a revisão por pares, mas apenas nove informam nas suas políticas que realizam a revisão desses documentos e/ou dos *datasets* e, desses, sete *data journals* informam os critérios de revisão utilizados.

4.4 A DISPONIBILIDADE DOS DADOS DE PESQUISA E AS CARACTERÍSTICAS DOS REPOSITÓRIOS DE DADOS

O objetivo dessa seção é verificar de que maneira os autores que publicaram os *data papers* disponibilizaram os dados associados, bem como as características dos repositórios de dados utilizados e sua cobertura temática. Por fim, busca-se verificar se os autores seguiram as orientações fornecidas pelos periódicos para disponibilizar os dados de pesquisa.

A disponibilidade dos dados de pesquisa em *data papers* refere-se à maneira como os autores compartilharam os dados de pesquisa associados aos *data papers*. Por isso, antes de se analisar as características dos repositórios, torna-se importante

verificar se essa foi a única forma adotada pelos autores filiados às instituições brasileiras.

A partir da análise individual dos 634 *data papers* publicados, em 38 periódicos, as maneiras utilizadas para disponibilizar os dados foram:

- a) no caso de 55,05% (349) dos *data paper* a maneira adotada foi a disponibilização a partir de **repositórios**;
- b) em 18,14% (115) dos casos os conjuntos de dados foram disponibilizados no **próprio *data paper***;
- c) em 8,68% (55) dos *data papers* os conjuntos foram disponibilizados como **material suplementar**;
- d) em 8,52% (54) dos *data papers* os **dados não foram disponibilizados**;
- e) em 8,36% (53) dos *data papers* os conjuntos foram disponibilizados de maneira combinada, **tanto em repositórios como em material suplementar e**
- f) em 1,26% (8) dos *data papers* os dados foram disponibilizados em **páginas pessoais ou institucionais**.

A partir dessas informações, pode-se observar que existem inconsistências na disponibilidade dos dados de pesquisa. Walters (2020) também já havia identificado esse padrão nas práticas adotadas pelos periódicos. Esses diferentes mecanismos adotados para disponibilizar os dados de pesquisa impedem que os dados cumpram seus objetivos na comunicação acadêmica, que é possibilitar a verificabilidade, a reprodutibilidade e a reutilização (Jiao; Li; Fang, 2022).

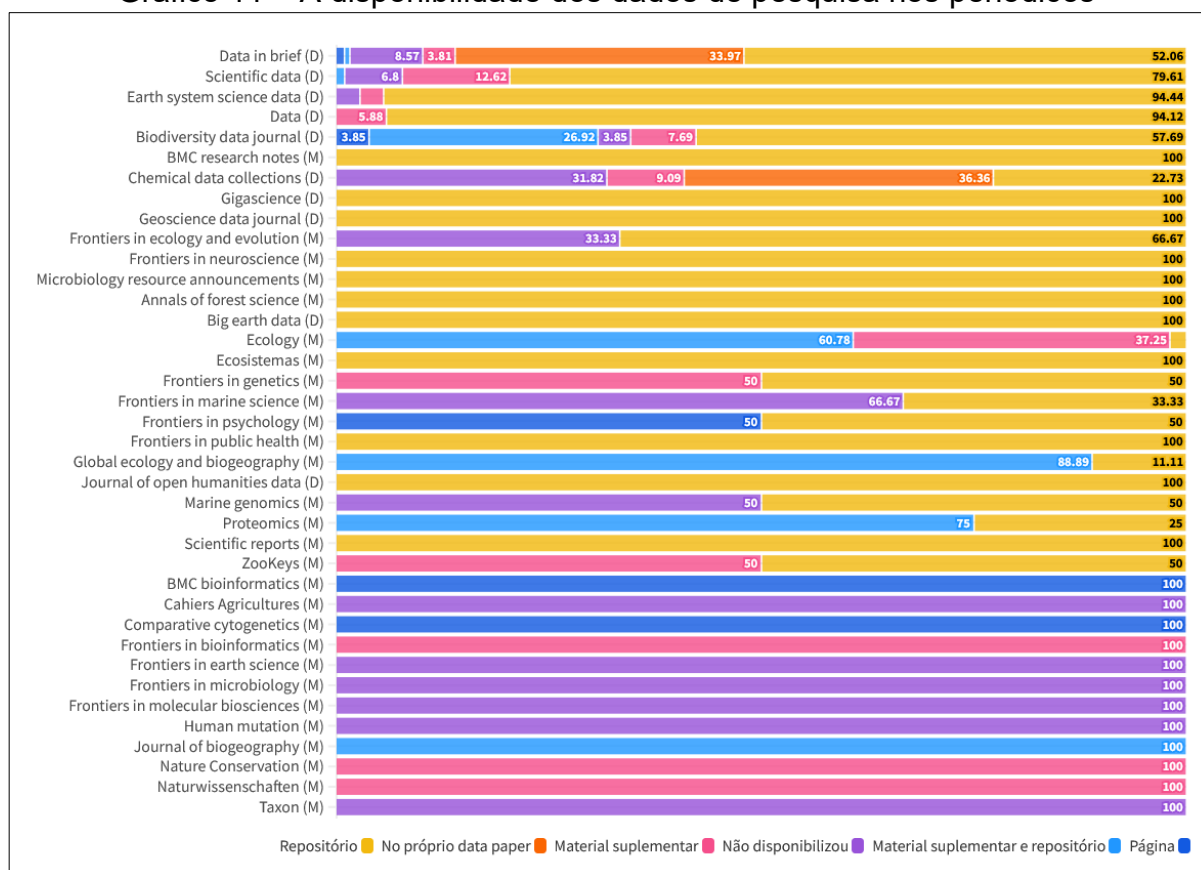
O compartilhamento dentro do artigo é insuficiente porque os dados compartilhados tendem a ser dados resumidos quando analisados, não permitindo o acesso aos dados brutos (Federer *et al.*, 2018). A disponibilização dos dados de pesquisa como material suplementar prejudica a curadoria e a preservação, além de não permitir que os leitores encontrem o seu *link* independentemente da publicação principal (Candela *et al.*, 2015).

A disponibilização dos dados como parte integrante dos *data papers* e a não disponibilização dos dados pesquisa foram aspectos observados em 14 dos 38 periódicos identificados, embora essa prática esteja regredindo ao longo dos anos, conforme Figura 14. Assim, pode-se dizer que, em 26,66% (169) dos *data papers*, os conjuntos de dados não foram disponibilizados de fato, uma vez que esse tipo de comportamento não permite o acesso à totalidade dos dados de pesquisa. Nesses

casos, os pesquisadores interessados acabam não tendo acesso aos conjuntos de dados, exceto se entrarem em contato direto com os autores e se forem atendidos, o que acaba se tornando uma barreira para a verificabilidade e o reuso dos dados de pesquisa (Federer *et al.*, 2018; Federer, 2022).

O Gráfico 11 apresenta a disponibilidade dos dados identificadas nos 38 periódicos que publicaram os *data papers*.

Gráfico 11 – A disponibilidade dos dados de pesquisa nos periódicos



Fonte: elaborado pela autora.

Nota: a figura e os dados estão disponíveis em: <https://public.flourish.studio/visualisation/24856780/>

A partir da análise do Gráfico 11, pode-se verificar padrões distintos de disponibilidade de dados entre periódicos mistos (M) e os *data journals* (D). Dentre os *data journals* houve um maior percentual de dados disponíveis em repositórios, à exceção do *Chemical Data Collections (Elsevier)* cujo destaque foi para a disponibilização no próprio *data paper* com 36,36%(8) e a não disponibilização com 31,82% (7). Embora o destaque para o depósito em repositórios tenha se mantido entre os *data journals* *Data in Brief (Elsevier)* e *Biodiversity Data Journal (Pensoft*

Publishers), nesses periódicos também foi possível verificar uma diversidade de maneiras utilizadas para disponibilizar os dados de pesquisa.

Dentre os periódicos mistos (M) foi possível identificar uma maior heterogeneidade na disponibilização dos conjuntos de dados, mesmo que em sete deles o uso dos repositórios tenha representado a totalidade e em oito deles a não disponibilização e o uso de páginas pessoais compromete o acesso aos dados brutos utilizados.

O que pode explicar essa heterogeneidade na maneira como os dados foram sendo disponibilizados ao longo do tempo nos periódicos, é a adesão irregular por parte dos *publishers* às práticas de abertura da ciência. Candela *et al.* (2015) identificaram que as primeiras tentativas das editoras para apoiar a publicação de dados foram a inserção dos dados como parte integrante do artigo ou anexá-los como arquivos suplementares.

O primeiro marco identificado na literatura que pode ajudar a explicar a heterogeneidade foi a obrigatoriedade de os artigos científicos possuírem uma Declaração de Disponibilidade de Dados (DAS), sendo que nos periódicos da PLOS a obrigatoriedade se iniciou a partir de 2014 (Jiao; Li; Fang, 2022) e nos periódicos da Wiley essa obrigatoriedade se deu a partir de 2018 (Graf *et al.*, 2020).

Além disso, em 2015, foram publicadas as *Transparency and Openness Promotion* (TOP) ou Diretrizes de Promoção de Transparência e Abertura, patrocinadas e coordenadas pelo *Center for Open Science*, cujo objetivo foi desenvolver padrões compartilhados para práticas abertas em periódicos, onde estes puderam selecionar o nível de padrão de transparência das práticas de pesquisa que desejavam implementar e o nível de implementação para cada um desses padrões (Center for Open Science, [2025]), sendo esse mais um marco que pode justificar os diferentes padrões de disponibilidades de dados nos periódicos.

Outro motivo que pode influenciar a disponibilidade dos dados, são os dados sobre sujeitos humanos identificáveis que geralmente não serão de domínio público, devido à necessidade de proteger a privacidade dos participantes da pesquisa, e com isso, os autores optam por compartilhar seus dados mediante solicitação (Hrynaskiewicz *et al.*, 2016).

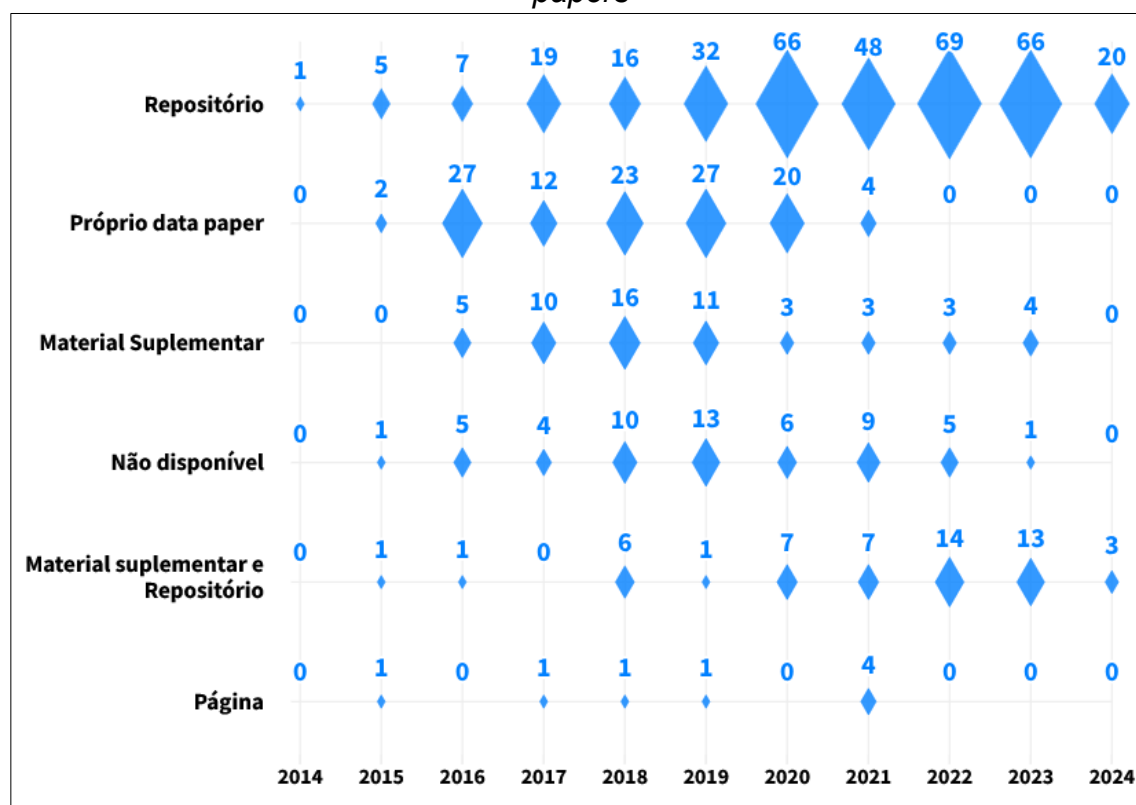
No entanto, McGuinness e Sheppard (2021) alertam que exigir declarações de disponibilidade de dados é um primeiro passo, porém a falta de revisão do

conteúdo dessas declarações leva a problemas, que comprometem a acessibilidade dos dados de pesquisa.

É importante ressaltar que esses marcos são direcionados aos periódicos científicos, mas com a ascensão dos *data journals* e das publicações de *data papers*, entre os anos de 2010 a 2012, conforme Tabela 7, essas práticas também foram sendo incorporadas nessas publicações.

Ao se analisar a disponibilidade dos dados de pesquisa associados aos *data papers*, ao longo dos anos, na Figura 14, foi possível identificar que à medida que a publicação de *data papers* foi crescendo a forma de disponibilização dos dados variou. Por exemplo, o destaque para a disponibilização no próprio *data paper* em 2016 foi de 27 *data papers*, o que corresponde a 60% do total publicado naquele ano, e em 2018 sua proporção diminuiu para 31,94%, o que corresponde a 23 *data papers*, do total publicado naquele ano.

Figura 14 - Formas de publicação dos dados de pesquisa associados aos *data papers*



Fonte: elaborado pela autora

Nota: a figura e os dados estão disponíveis em:

<https://public.flourish.studio/visualisation/20627498/>

Embora o uso dos repositórios para disponibilizar os conjuntos de dados venha ocorrendo desde 2014, sua consolidação por parte dos autores filiados às instituições brasileiras que publicaram *data papers*, se deu de maneira progressiva a partir de 2017 e seu predomínio em relação às outras formas de disponibilização aconteceu apenas em 2019, ocorrendo em consonância com diminuição de outras formas de disponibilização. A tendência de crescimento do uso de repositórios ao longo dos anos em detrimento da diminuição de outras formas também foi identificada nos estudos de Jiao, Li e Fang (2022). Ao comparar o trabalho desses autores com esta pesquisa, os resultados indicam uma tendência crescente na adesão do uso de repositórios de dados de pesquisa por parte dos autores, embora seja importante ressaltar que a tipologia documental analisada pode ter influência direta nos resultados.

Dito de outra forma, em pouco mais da metade dos *data papers* publicados os repositórios foram a opção dos autores filiados às instituições brasileiras para disponibilizar os dados de pesquisa, e isso pode ser explicado pelas diferenças entre as disciplinas (Federer, 2022). No caso dos artigos científicos publicados na *Plos One*, Federer *et al.* (2018) haviam identificado que o uso dos repositórios é o método dominante em cerca de 20% dos artigos que foram analisados.

Portanto, os resultados deste trabalho parecem compatibilizar com os resultados do estudo de Federer (2022) onde simplesmente exigir que os dados sejam compartilhados de alguma forma pode não ter o impacto desejado de tornar os dados de pesquisa FAIR. Sendo assim garantir que os dados estejam realmente disponíveis e permaneçam assim a longo prazo pode exigir requisitos e especificações adicionais sobre como e onde os dados devam ser compartilhados.

4.4.1 Características dos repositórios utilizados pelos autores filiados às instituições brasileiras

Ao se analisar cada *data paper*, foi possível identificar que em 402 deles, o que corresponde a um percentual de 63,40% do total identificado, os dados de pesquisa foram disponibilizados somente em repositórios ou combinados como material suplementar e repositório, totalizando 699 conjuntos de dados (ou *datasets*), refletindo uma média de 1,49% *dataset* por *data paper*, ou seja, em muitos casos o documento em questão fazia menção a mais de um conjunto.

O percentual de 63,40% de *data papers* cujos dados de pesquisa foram disponibilizados em repositórios, supera os 20% identificados por Federer *et al.* (2018), quando avaliaram as formas de compartilhamento de dados de pesquisa, a partir das declarações de disponibilidade de dados dos artigos científicos publicados na *Plos One* entre os anos 2014 e 2016. Essa diferença na porcentagem da disponibilização dos dados em repositórios pode ser influenciada por alguns fatores, tais como a tipologia de documento usado nesta pesquisa, os *data papers*, uma vez que os repositórios já fazem parte desse ecossistema de publicação, e também pelo período analisado, uma vez que a publicação de *data papers* por autores filiados às instituições brasileiras, se inicia em 2014 e vem se consolidando ao longo do tempo, ainda com adesão por parte dos periódicos ao uso de repositórios.

Visando identificar um rol de características dos repositórios cujos dados de pesquisa associados aos *data papers* foram depositados, foi elaborado o Quadro 15. No Apêndice G estão arrolados os mantenedores dos repositórios de dados mencionados.

Quadro 15 - Características dos repositórios

N	Repositório	F dataset	Tipo de Instituição	País	Depósito/ Cobrança	Revisão	Acesso	Licença
1	Mendeley Data	108	C	NLD; GBR	Gratuito até 10GB	S	Aberto; Restrito; Fechado	CC; CC0
2	Zenodo	84	SL	EUE	Gratuito até 50GB	N	Aberto; Restrito; Fechado	CC; CC0; outro
3	Figshare	70	C	ROU; USA; GBR	Gratuito até 20GB	Apenas na versão Plus	Aberto; Embargo	CC0; CC; Apache License 2.0; BSD; outro
4	Sequence Read Archive	69	SL	USA	Gratuito	S	Aberto	copyright; domínio público
5	Genbank	64	SL	USA; GBR	Gratuito	S	Aberto; Restrito; Fechado	CC; copyright
6	Pangaea	31	SL	DEU	Gratuito	S	Aberto	CC
7	Lancaster University Research Directory	24	SL	GBR	Gratuito	NI	Aberto	CC; ODC
8	Open Science Framework	23	SL	USA	Gratuito	N	Aberto; Restrito; Fechado	CC; outro
9	Global Biodiversity Information Facility	20	SL	DNK	Gratuito	S	Aberto; Restrito; Fechado	Outro
10	Github	20	C	USA	Gratuito até 500MG	N	Aberto	CC; CC0; outro
11	BioProject	17	SL	USA	Gratuito	N	Aberto; Restrito	copyright; domínio público
12	Proteomics Identifications Database	16	SL	GBR; EUE	Gratuito	S	Aberto; Restrito	CC
13	Gene Expression Omnibus	12	SL	USA	Gratuito	S	Aberto; Restrito	copyright
14	Dryad	11	SL	USA	Taxa de US\$ 150 para 50GB e mais taxas quando exceder o tamanho	S	Aberto; Embargo	CC0; copyright
15	European Nucleotide Archive	10	SL	GBR; EUE	Gratuito	N	Aberto	CC
16	Mass Spectrometry Interactive Virtual Environment (Massive)	8	SL	USA	Gratuito	S	Aberto	CC
17	Cambridge Crystallographic Data Center	7	SL	USA; GBR	Gratuito	S	Aberto	copyright; domínio público
18	Science Data Bank	7	SL	CHN	Gratuito. Mas não garante que serviços de armazenamento, revisão e publicação de dados terão gratuidade vitalícia.	S	NI	NI
19	GigaScience Database	7	C; SL	HKG	As cobranças para depósito estão cobertas pelas taxa de APC do periódico GigaScience	S	Aberto	CC; domínio público

N	Repositório	F dataset	Tipo de Instituição	País	Depósito/Cobrança	Revisão	Acesso	Licença
20	BioSample	7	SL	USA	Gratuito	N	Aberto; Restrito; Embargo	CC; CC0
21	Synapse	6	SL	USA	Gratuito até 100GB. Quando excedido o tamanho o valor varia de USD\$ 15.000 a USD\$ 30.000	N	Aberto; Restrito; Embargo; Fechado	Outro
22	Dados Abertos do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	5	SL	BRA	Gratuito	NI	Aberto	CC; CC0; outro
23	Plasmodium Genomics Resource	5	SL	USA	Gratuito	S	Aberto; Embargo	Outro
24	Atmospheric Science Data Center (ASDC)	4	SL	USA	Gratuito	NI	Aberto	domínio público; outro
25	Data repository	4	SL	DEU	Gratuito	N	NI	NI
26	National Centers for Environmental Information	4	SL	USA	Gratuito	S	Aberto	copyright; domínio público
27	ArrayExpress	3	SL	USA; EUE; GBR	Gratuito	S	Aberto; Restrito; Embargo; Fechado	Domínio público; outro
28	Harvard DataVerse	3	SL	USA	Gratuito	N	Aberto	CC; CC0
29	Map of Life (MOL)	3	SL	USA	Gratuito	NI	NI	NI
30	RCSB Protein Data Bank	3	SL	USA	Gratuito	S	Aberto; Embargo	Outro
31	Sea Scientific Open Data Publication	3	SL	FRA	Limitado até 100GB	S	Aberto	CC0; outro
32	MetaboLights	3	SL	EUE; GBR	Gratuito	S	Aberto; Restrito; Embargo	CC
33	Environmental Data Initiative Repository	2	SL	USA	Gratuito	S	Aberto; Restrito	CC; CC0
34	Herbarium of Federal University of Mato Grosso do Sul	2	SL	BRA	Gratuito	S	NI	NI
35	NCBI DataSets	2	SL	USA	Gratuito	S	Aberto; Restrito; Embargo	outro
36	Immunology Database and Analysis Portal	2	C; SL	USA	Gratuito	S	Aberto; Embargo	CC0; copyright
37	Ag Data Commons	1	SL	USA	Gratuito	S	Aberto; Embargo	CC0; CC; Domínio público; ODC; outro
38	Archive of Data on Disability to Enable Policy	1	SL	USA	Gratuito	NI	NI	NI
39	British Antarctic Survey	1	SL	GBR	Gratuito	S	Aberto; Embargo	copyright; domínio público

N	Repositório	F dataset	Tipo de Instituição	País	Depósito/Cobrança	Revisão	Acesso	Licença
40	British Oceanographic Data Centre	1	SL	GBR	Gratuito	N	Aberto; Restrito; Fechado	copyright
41	Cirad dataverse	1	SL	FRA	Gratuito	N	Aberto; Restrito	CC; domínio público; ODC
42	Dagshub	1	C	ISR	Gratuito até 20GB	N	NI	NI
43	DataSuds	1	SL	FRA	Não	N	Aberto; Restrito; Embargo; Fechado	CC
44	DataverseNL	1	SL	NLD	Taxa de adesão 5.550 euros e 310 euros por terabyte (TB) por ano para as instituições participantes	N	Aberto; Restrito; Fechado	CC
45	Digital CSIC	1	SL	ESP	Não	N	Aberto; Embargo; Restrito	CC; ODC; outro
46	Digital Rocks Portal	1	SL	USA	Patrocínio para manutenção do repositório que varia de: US\$20 a US\$3000 dólares	NI	Aberto; Embargo	ODC; copyrights
47	Fairdata IDA Research Data Storage Service	1	SL	FIN	Gratuito	NI	Fechado; Restrito	CC; outro
48	Freshwater Research and Environmental Database	1	SL	DEU	Gratuito	NI	Aberto; Restrito	CC; ODC; outro
49	GFZ Data Service	1	SL	DEU	Gratuito	S	Aberto; Embargo	CC; outro
50	Global freshwater biodiversity research	1	SL	AUT; DEU; BEL; EUE	Gratuito	S	Aberto	CC; CC0; ODC; outro
51	Global Health Data Exchange	1	SL	USA	Gratuito	N	Aberto; Restrito; Embargo	ODC
52	ICOS Carbon Portal	1	SL	EUE; SWE; NLD	Gratuito	N	Aberto; Restrito	CC
53	International Neuroimaging Data-Sharing Initiative	1	C; SL	USA	Gratuito	NI	Restrito	copyright
54	IsoArch	1	SL	FRA	Gratuito	S	Aberto; Restrito	CC; copyright
55	Materials Cloud repository	1	SL	CHE	Limitado até 5 GB	S	NI	NI
56	Metagenomics analysis server	1	SL	USA	Gratuito	S	Aberto; Restrito; Embargo; Fechado	CC; copyright; outro
57	NoMaD Repository & Archive	1	SL	DEU	Gratuito	N	Aberto; Restrito; Embargo;	CC
58	OTELo Research Data Repository	1	SL	FRA	Gratuito	S	Aberto; Restrito;	CC

N	Repositório	F dataset	Tipo de Instituição	País	Depósito/ Cobrança	Revisão	Acesso	Licença
							Embargo; Fechado	
59	ProteomeXchange	1	SL	EUE	Gratuito	S	Aberto; Restrito	CC; outro
60	Repositório da UNESP	1	SL	BRA	Gratuito	N	NI	NI
61	Repositório da Universidade de São Paulo	1	SL	BRA	Gratuito	N	NI	NI
62	Repositório de Dados de Pesquisa da Fundação Getulio Vargas	1	SL	BRA	Gratuito	NI	NI	NI
63	Repositório de Dados de Pesquisa da Unicamp (REDU)	1	SL	BRA	Gratuito	N	NI	NI
64	Repository of Open Access Datasets	1	SL	AUS	Limitado até 20GB	NI	NI	NI
65	Saint Mary's University Dataverse	1	SL	CAN	Gratuito	NI	Aberto; Restrito; Fechado	CC; CC0
66	TRY	1	SL	FRA; DEU	Gratuito	N	Aberto; Restrito	outro
	Total	699						

Fonte: dados da pesquisa

Notas: Coluna F: refere-se à quantidade de *data sets* identificados nos *data papers*

Coluna tipo de instituição: refere-se ao tipo de instituições mantenedora do repositório. Sendo SL: sem fins lucrativos; C: comercial; NI: Não foi identificado

Coluna país: refere-se ao país ou países que mantêm os repositórios, dados identificados no R3data. Abreviação de países conforme ISO 3166-1:alpha 3 codes

Coluna depósito/cobrança: refere-se às orientações fornecidas pelo repositório sobre o depósito e/ou cobranças relacionadas ao depósito dos dados

Coluna revisão: refere-se à informação fornecida pelos repositórios se os dados passam por algum processo de validação e/ou revisão. Sendo S: sim; N: Não; NI: não identificado

Coluna acesso: refere-se aos tipos de acesso aos dados que o repositório oferece, dados identificados no R3data

Coluna licença: refere-se às permissões referentes ao reuso e manipulação dos dados de pesquisa, dados identificados no R3data. Sendo: ODC: Open Data Commons; NI: Não foi identificado

Foram identificados 66 repositórios diferentes e, desses, 59 são mantidos por instituições sem fins lucrativos, concentrando 70,10% (490) dos conjuntos de dados e uma média de 8,3 conjuntos por repositório. Ainda, quatro repositórios são mantidos por organizações comerciais que, juntas, concentram 28,47% (199) dos conjuntos de dados e uma média de 49,7 conjuntos, e três são mantidos em parceria por instituições sem fins lucrativos e comerciais que, juntas, possuem 1,43% (10) dos conjuntos de dados e uma média de 3,33 conjuntos. Esses resultados indicam que embora as instituições totalmente comerciais sejam minoria no que condiz à criação e a manutenção dos repositórios, elas concentram uma média de conjuntos muito maior do que as demais.

Dessas instituições, a Elsevier, responsável pelo *Mendeley Data*, detém 15,45% (108) conjuntos de dados depositados, sendo também o *publisher* responsável pela maior parte dos *data papers* publicados. Logo, esses resultados apontam o domínio dessa empresa nos *data papers* e seu ecossistema publicados por autores filiados às instituições brasileiras.

Além do *Mendeley Data*, os repositórios *Zenodo* e *Figshare*, que concentram 12,02% (84) e 10,01% (70) conjuntos de dados respectivamente, foram aqueles com maior destaque entre os autores filiados às instituições brasileiras que publicaram *data papers*. Com relação à área da Enfermagem, Pinto (2022) identificou o *Figshare* como o preferido pela área, seguido do *Zenodo* e do *Dryad*. Contudo, a autora não identificou as razões pelas quais um pesquisador optou por um determinado repositório em detrimento a outro.

Todavia, Pinto (2022) apontou algumas hipóteses que podem justificar essa preferência. A primeira está relacionada à simplicidade para realizar o *upload* dos dados de pesquisa, a segunda relaciona-se com a possibilidade de descrição dos dados e a terceira está associada à gratuidade do serviço, ou seja, o *Figshare* não cobra taxa de curadoria e preservação dos dados com até 20GB ali depositados. Seguindo a tendência dos resultados apresentados pela autora, nesta pesquisa pode-se supor que a gratuidade, atrelada à capacidade de armazenamento desses repositórios, pode ser fator que também levou à seleção dos repositórios que mais concentraram conjuntos de dados pelos autores filiados às instituições brasileiras.

Em relação aos mantenedores, 60,61% (40) dos repositórios são mantidos por mais de uma instituição e 26 repositórios, o que representa 39,39%, são mantidos por apenas uma instituição. A esta análise pode-se adicionar ainda que 11

repositórios são mantidos por instituições localizadas em mais de um país, sendo que esses concentram 40,63% (284) dos conjuntos de dados depositados. Esses resultados podem indicar o êxito das colaborações interinstitucionais entre os países na manutenção e depósito de dados, dada a complexidade na gestão e curadoria dos dados de pesquisa, bem como na manutenção dos repositórios.

Ainda no que tange à localização dos mantenedores dos repositórios, foi possível identificar as iniciativas individuais por países, sendo que:

- a) nos Estados Unidos (USA) estão localizados 24 repositórios com 29,61% (207) dos conjuntos depositados;
- b) no Brasil (BRA) localizam-se seis repositórios com 1,57% (11) conjuntos depositados;
- c) na Alemanha (DEU) localizam-se cinco repositórios com 5,44% (38) dos conjuntos;
- d) na França (FRA) localizam-se quatro repositórios com 1% (7) conjuntos;
- e) na Grã Bretanha (GBR), localizam-se três repositórios com 3,72% (26) conjuntos de dados;
- f) na União Européia (EUE), localizam-se dois repositórios com 12,16% (85) conjuntos;
- g) na Dinamarca (DNK), localiza-se 1 repositório com 2,86% (20) conjuntos de dados;
- h) em Hong Kong (HKG) e na China (CHN), localiza-se um repositório cada e 1% (7) conjuntos depositados cada e
- i) na Austrália (AUS), Canadá (CAN), Espanha (ESP), Finlândia (FIN), Israel (ISR), Holanda (NLD) e Suíça (CHE) também localiza-se um repositório e 0,14% (1) conjuntos de dados cada.

O predomínio dos repositórios em países do Norte Global sugere um monopólio desta porção do Globo no armazenamento dos dados de pesquisa.

Considerando os quantitativos de repositórios de dados brasileiros, foi possível identificar que todos são mantidos por uma única instituição, sendo quatro deles por universidades. Mesmo com as iniciativas brasileiras do 4º plano de ação para a criação de repositórios de dados no país e governança dos dados de pesquisa (Brasil, 2018a), apenas 1,57% dos conjuntos de dados dos autores filiados às instituições brasileiras foram depositados no país. Isso pode indicar a necessidade de criação de repositórios, por meio do estabelecimento de parcerias entre

instituições e a criação de políticas de incentivo ao compartilhamento dos dados de pesquisa.

Em relação ao depósito dos dados de pesquisa, nos quatro repositórios que são mantidos por instituições comerciais e que limitam o depósito dos dados que variam de 500MG até 20GB, bem como nos repositórios que são mantidos em parcerias entre instituições com fins lucrativos e sem fins lucrativos, foi possível identificar que apenas um não limita o tamanho dos dados, um não foi possível identificar se existe cobrança. No caso do repositório *GigaScience Database*, a gratuidade está atrelada ao pagamento da taxa de APC ao periódico, ou seja, o depósito dos dados está condicionado à publicação do *data paper* no periódico específico editado pelo *publisher*.

Tratando-se dos repositórios mantidos por instituições sem fins lucrativos, 44 repositórios não cobram taxas e nem limitam o tamanho dos dados de pesquisa e em quatro repositórios não foi possível identificar algum tipo de limitação para o depósito. Embora o repositório *Science Data Bank*, informe que não há cobranças pelo depósito dos dados, ele também menciona que não garante o serviço gratuito de maneira vitalícia para o armazenamento, revisão e publicação dos dados de pesquisa. Apenas o repositório Dryad informa que cobra uma taxa de \$150 dólares para o depósito dos dados e mais taxas para o tamanho excedente, a taxa de *data publishing charges* (DPCs). Nos demais repositórios há algum tipo de limitação e/ou cobrança para o depósito dos dados de pesquisa.

Candela *et al.* (2015) alertam que em alguns casos, a publicação do conjunto de dados em si em um repositório tem um custo. Portanto, o custo de publicação de um artigo de dados é, na verdade, a soma dos dois custos, o custo pelo depósito no repositório ou pelo excedente no tamanho do conjunto de dados e o custo do APC do periódico. Isso pode gerar uma barreira que afeta o movimento de publicação de dados.

A partir desse contexto, foi possível identificar que a cobrança nos repositórios ocorre tanto em repositórios comerciais quanto naqueles são mantidos por instituições sem fins lucrativos. Essa cobrança está relacionada ao tamanho dos dados de pesquisa, na maioria dos repositórios identificados, e um repositório a cobrança está atrelada à publicação no periódico editado pelo *publisher* que mantém o periódico e o repositório, e por última a cobrança está relacionada a um patrocínio para a manutenção do repositórios. Esses resultados podem indicar uma tendência

de dominação do mercado de repositório de dados pelos *publishers* comerciais, assim como acontece com os periódicos.

Os autores filiados às instituições brasileiras publicaram 47,58% (333) dos dados associados aos *data papers* em repositórios que não cobram e não limitam o tamanho dos conjuntos. Ainda, 22,14% (155) dos conjuntos de dados foram publicados em repositórios que limitam o tamanho dos dados de 20Gb e 50Gb e apenas 1,57% (11) dos conjuntos foram publicados num repositório que cobra DPC.

Mayernik *et al.* (2015) destacam que a distinção nos processos de revisão de dados entre as publicações de dados em periódicos e em repositórios de dados, é que, no caso dos repositórios, ela concentra-se principalmente em aspectos técnicos do conjunto de dados, a fim de garantir que o conjunto de dados possa ser gerenciado e curado corretamente. Assante *et al.* (2016) complementam ainda, que os repositórios fazem uma validação dos dados e não uma revisão por pares, conforme ocorre nos periódicos, uma vez que não há critérios estabelecidos compartilhados sobre como realizar tal revisão.

Com isso, alguns repositórios de dados realizam validação do conjunto de dados, oferecendo práticas e serviços para a pré-publicação do conjunto de dados e a validação pós-publicação. O primeiro é uma espécie de avaliação preventiva com o objetivo de evitar que os conjuntos de dados de má qualidade sejam publicados, sendo uma espécie de garantia de qualidade envolvendo tanto o proprietário dos dados quanto o repositório atuando como editor de dados. O último é uma validação resultante do *feedback* recebido de usuários que tentaram reutilizar o conjunto de dados publicado (Assante *et al.*, 2016).

Em relação à revisão ou validação dos dados de pesquisa depositados, 31 repositórios informaram que a realizam. Os repositórios *Sequence Read Archive*, *NCBI DataSets* e o *RCSB Protein Data Bank* possuem ferramentas que realizam a validação automática dos conjuntos de dados. O repositório *National Centers for Environmental Information* possui um conselho curador para a avaliação dos dados e o repositório *Materials Cloud* possui uma política de moderação que abarca os seguintes critérios: conteúdo impróprio, submissões que não possuam os dados produzidos por pesquisas originais, formulário apropriado, tópico inadequado, conteúdo duplicado, tamanho de submissão excessivo e envio de material protegido por direitos autorais (Materials Cloud..., 2020). Os outros 33 repositórios informam não realizar nenhum tipo de revisão ou validação nas suas políticas.

A revisão dos dados de pesquisa pode ajudar na sua confiabilidade e reusabilidade e este pode ser um atributo a ser requerido no futuro, por parte de governos e agências de financiamento (Richardson; Wolski, 2014).

Os acessos e licenças identificados nesta pesquisa estão associados àquelas suportadas pelos repositórios de dados, ficando a critério dos responsáveis pelo depósito selecionar o tipo de uso e a forma de acesso aos dados de pesquisa associados aos *data papers*. Cabe salientar que a análise individual dos conjuntos de dados, não foi realizada nesta pesquisa. É importante destacar que pode haver dois tipos de licenças na publicação de dados de pesquisa, aquela relacionada ao *data paper* em si e aquela relacionada ao *dataset*.

Os acessos estão relacionados às maneiras utilizadas para se obter os dados de pesquisa a partir dos repositórios. Foi verificado que 38 repositórios associaram ao acesso aberto, a possibilidade de acesso restrito, fechado ou embargado⁶², 14 deles adotam acesso totalmente aberto, 12 não informaram as formas de acesso utilizadas e um informou apenas o acesso restrito ao dado. Na prática, estes dados informam que os repositórios deixam a critério dos depositantes, o tipo de acesso aos seus dados de pesquisa.

As licenças estão relacionadas às maneiras como os dados podem ser reutilizados, sendo atribuídas pelos proprietários dos dados. Foi verificado que os repositórios adotam diversos padrões de licenças e uma tendência de combinações entre elas. As licenças que mais predominaram foram a *Creative Commons* (CC), onde dez repositórios informaram adotar apenas ela, e 26 repositórios informaram usar a CC associada a outras licenças, 12 repositórios não informaram as licenças suportadas, seis repositórios informaram usar outras licenças, mas sem especificar quais, e os demais (12) informaram outras modalidades de licença.

Esses dados indicam uma presença dominante das licenças *Creative Commons* e daquelas ditas de domínio público, além da *Open Data Commons* (ODC). Elas aparecem isoladamente ou em combinação com outras, o que demonstra uma possível flexibilidade por parte dos repositórios na abertura dos dados.

A presença da licença *copyright* sozinha ou associada à uma licença aberta (como CC ou domínio público) indica uma restrição ou uma sobreposição de licenças, o que pode prejudicar a reutilização dos dados de pesquisa, uma vez que não fica

⁶² Embargo: refere-se a um período de tempo em que o acesso ao conteúdo do conjunto de dados é restrito ao acesso público (Assante *et al.*, 2016).

definido se os dados podem ser de fato reutilizados e acaba ferindo o princípio de tornar abertas as pesquisas financiadas com recursos públicos (Assante *et al.*, 2016).

4.4.2 Cobertura temática e tipologia dos repositórios

Para verificar a cobertura temática e a tipologia dos repositórios aos quais os dados associados aos *data papers* são depositados, inicialmente foi identificada a média de repositórios citados nessas publicações para o depósito dos conjuntos de dados. Também foi realizada a análise dos *data papers* individualmente e, com isso, pode-se perceber, conforme Tabela 9, que houve uma variação na quantidade de repositórios, indo de um a oito repositórios por *data paper*.

Tabela 9 - Quantidade de repositórios mencionados nos *data papers*

Quantidade de repositórios	F* <i>data paper</i>	%	F** <i>dataset</i>	%	Média
1	357	88,81	446	63,81	1,25
2	33	8,21	73	10,44	2,21
3	9	2,24	54	7,73	6
4	1	0,25	4	0,57	4
5	1	0,25	85	12,16	85
6	-	-	-	-	-
7	-	-	-	-	-
8	1	0,25	37	5,29	37
Total	402	100	699	100	1,74

Fonte: elaborado pela autora

Notas: os traços (-) significam zeros

Coluna F* *data paper*: refere-se à quantidade de *data papers* que mencionou a utilização de repositórios

Coluna %: refere-se à porcentagem de *data papers* que mencionaram a utilização de repositórios em relação ao total identificado

Coluna F** *datasets*: refere-se à quantidade de *datasets* (ou conjuntos de dados) identificados no repositório específico

Coluna %: refere-se à porcentagem de *datasets* depositados no repositório em relação ao total identificado

Coluna média: refere-se à média de *datasets* por *data paper*

Em relação à quantidade de repositórios mencionados nos *data papers*, foi identificado que 63,81% (446) dos dados de pesquisa associados aos *data papers* foram depositados em apenas um repositório. Dessa forma, pode-se inferir que o comportamento de publicação dos autores filiados às instituições brasileiras no que condiz aos *data papers* é utilizar apenas um repositório de dados para cada conjunto de dados.

Embora haja um predomínio de uso de um único repositório, é possível observar que um *data paper* utilizou oito repositórios diferentes para disponibilizar os

dados de pesquisa, totalizando 37 conjuntos de dados e outro utilizou cinco repositórios diferentes para disponibilizar 85 conjuntos de dados de pesquisa. Essa grande quantidade de repositórios e de conjuntos de dados pode ter relação com o comportamento da área, uma vez que ambos⁶³ pertencem à área da Medicina.

Foi identificado que em média os *data papers* fazem a descrição de 1,74 conjuntos de dados. McGillivray *et al.* (2022) também realizaram essa análise e identificaram um conjunto de dados associado por cada *data paper* na área das Ciências Humanas e Sociais. Embora nesta pesquisa a medida tenha sido um pouco maior, pode-se supor que suas práticas podem estar associadas às necessidades de compartilhamento de dados relacionadas às áreas do conhecimento.

Gorgolewski, Margulies e Milham (2013) alertam que alguns projetos que produzem dados em larga escala buscam determinar se deve-se relatar o trabalho num único artigo ou dividi-lo em dois ou mais estudos menores com o objetivo de aumentar a velocidade ou o número de publicações. Nesses casos, tais práticas devem ser desencorajadas em todos os níveis do processo de publicação, pois apenas a publicação de dados completos fornece uma compreensão clara do projeto experimental pretendido e das decisões tomadas.

Com o objetivo de identificar a tipologia dos repositórios, na Tabela 10 são arrolados todos os repositórios onde os conjuntos de dados foram depositados, bem como sua média e sua tipologia.

Tabela 10 - Tipologia dos repositórios

Nome do Repositório	F <i>datasets</i>	F <i>data</i> <i>paper</i>	%	Média	Tipo
Mendeley Data	108	102	15,45	1,06	M
Zenodo	84	79	12,02	1,06	M
Figshare	70	64	10,01	1,09	M
Sequence Read Archive	69	12	9,87	5,75	D
Genbank	64	23	9,16	2,78	D
Pangaea	31	13	4,43	2,38	D
Lancaster University Research Directory	24	1	3,43	24	M
Open Science Framework	23	7	3,29	3,29	D
Github	20	19	2,86	1,05	D
Global Biodiversity Information Facility	20	19	2,86	1,05	D

⁶³ Almutairi, H. *et al.* (2021). Chromosome-scale genome sequencing, assembly and annotation of six genomes from subfamily *Leishmaniinae*. *Science Data*, 8(234), 1-9. <https://doi.org/10.1038/s41597-021-01017-3>

DeBarry, J.D.. *et al.* (2022). MaHPIC malaria systems biology data from *Plasmodium cynomolgi* sporozoite longitudinal infections in macaques. *Science Data*, 9(722), 1-43. <https://doi.org/10.1038/s41597-022-01755-y>

Nome do Repositório	F datasets	F data paper	%	Média	Tipo
BioProject	17	6	2,43	2,83	D
Proteomics Identifications Database	16	13	2,29	1,23	D
Gene Expression Omnibus	12	6	1,72	2	D
Dryad	11	11	1,57	1	D
European Nucleotide Archive	10	4	1,43	2,50	D
Mass Spectrometry Interactive Virtual Environment (Massive)	8	5	1,14	1,60	D
BioSample	7	2	1	3,50	D
Cambridge Crystallographic Data Center	7	4	1	1,75	NI
GigaScience Database	7	5	1	1,40	D
Science Data Bank	7	6	1	1,17	M
Synapse	6	5	0,86	1,20	D
Dados Abertos do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	5	3	0,72	1,67	D
Plasmodium Genomics Resource	5	1	0,72	5	D
Atmospheric Science Data Center (ASDC)	4	1	0,57	4	D
Data repository	4	4	0,57	1	NI
National Centers for Environmental Information	4	3	0,57	1,33	D
ArrayExpress	3	3	0,43	1	D
Harvard DataVerse	3	3	0,43	1	M
Map of Life (MOL)	3	1	0,43	3	NI
MetaboLights	3	1	0,43	3	D
RCSB Protein Data Bank	3	2	0,43	1,50	D
Sea scientific open data publication	3	3	0,43	1	D
Environmental Data Initiative Repository	2	2	0,29	1	D
herbarium of Federal University of Mato Grosso do Sul	2	2	0,29	1	NI
Immunology Database and Analysis Portal	2	1	0,29	2	D
NCBI DataSets	2	2	0,29	1	D
Ag Data Commons	1	1	0,14	1	D
Archive of Data on Disability to Enable Policy	1	1	0,14	1	NI
British Antarctic Survey	1	1	0,14	1	D
British Oceanographic Data Centre	1	1	0,14	1	D
Cirad dataverse	1	1	0,14	1	M
Dagshub	1	1	0,14	1	NI
DataSuds	1	1	0,14	1	D
DataverseNL	1	1	0,14	1	M
Digital CSIC	1	1	0,14	1	M
Digital Rocks Portal	1	1	0,14	1	D
Fairdata IDA Research Data Storage Service	1	1	0,14	1	M
Freshwater Research and Environmental Database	1	1	0,14	1	D
GFZ Data Service	1	1	0,14	1	D
Global freshwater biodiversity research	1	1	0,14	1	D
Global Health Data Exchange	1	1	0,14	1	D
ICOS Carbon Portal	1	1	0,14	1	D
International Neuroimaging Data-Sharing Initiative	1	1	0,14	1	D
IsoArch	1	1	0,14	1	D
Materials Cloud repository	1	1	0,14	1	NI
Metagenomics analysis server	1	1	0,14	1	D

Nome do Repositório	F <i>datasets</i>	F <i>data</i> <i>paper</i>	%	Média	Tipo
NoMaD Repository & Archive	1	1	0,14	1	NI
OTELo Research Data Repository	1	1	0,14	1	D
ProteomeXchange	1	1	0,14	1	D
Repositório Institucional da UNESP	1	1	0,14	1	M
Repositório da Universidade de São Paulo	1	1	0,14	1	NI
Repositório de Dados de Pesquisa da Fundação Getulio Vargas	1	1	0,14	1	NI
Repositório de Dados de Pesquisa da Unicamp (REDU)	1	1	0,14	1	M
Repository of Open Access Data sets	1	1	0,14	1	NI
Saint Mary's University Dataverse	1	1	0,14	1	M
TRY	1	1	0,14	1	D
Total	699	468*	100	-	

Fonte: dados da pesquisa

Notas: Coluna F *datasets*: refere-se à quantidade de *datasets* (ou conjuntos de dados) publicados no repositório específico

Coluna F *data paper*: refere-se à quantidade de *data papers* que mencionou o repositório específico

Coluna %: refere-se à porcentagem de *data papers* em relação ao total identificado.

Coluna média: refere-se à média de de *datasets* por *data paper*.

Coluna tipo: refere-se ao tipo de cobertura do repositório: D indica que o repositório é disciplinar; M indica que o repositório é multidisciplinar e NI (não identificado) indica que o repositório não estava indexado no R3data.

Total de *data papers*: 468*. Esse quantitativo diverge do número total de *data papers* identificados (402), uma vez que um único *data paper* contemplou múltiplos repositórios. Nessa circunstância, cada repositório foi contabilizado individualmente como um *data paper*.

Foram identificados 66 repositórios distintos, dos quais 42 são disciplinares, concentrando 53,79% (376) dos conjuntos de dados, 13 são multidisciplinares que, apesar da menor quantidade, concentram 42,92% (300) dos conjuntos e 11 repositórios não estavam indexados no R3data, não tendo sua tipologia classificada, sendo que esses concentraram 3,29% (23) dos conjuntos de dados.

Esses resultados sugerem que a concentração de dados em repositórios multidisciplinares ou generalistas pode estar associada à ausência de repositórios específicos para certas comunidades (Assante *et al.*, 2016), à influência das recomendações dos periódicos que priorizam o uso de repositórios específicos, levando autores a depositar dados disciplinares nesses repositórios (Candela *et al.*, 2015) ou à ascensão dos repositórios de dados multidisciplinares, uma vez que esses começaram a ser criados no início dos anos de 2010, têm sido cada vez mais utilizados na amostra do artigo, principalmente impulsionados pelo crescente grau de multidisciplinaridade da revista (Jiao; Li; Fang, 2022).

Embora os repositórios disciplinares estejam em maiores quantidades, quase o dobro dos multidisciplinares, tendem a concentrar um número maior de conjuntos por *data paper*, resultando em médias de depósito que variam de 5,75 a 1. O fato de

os repositórios disciplinares possuírem uma maior média de conjuntos por *data paper* pode ser influenciado pelo comportamento das áreas do conhecimento, embora neste estudo a análise tenha se concentrado apenas na tipologia do repositório e não no repositório especificamente. Hrynaskiewicz *et al.* (2016) destacam que a utilização de repositórios disciplinares, que tendem a usar formatos e padrões de dados aceitos pela comunidade, além de padronizarem os processos de curadoria de dados, podem aumentar a probabilidade de que os dados possam ser reutilizados.

Em relação aos dez repositórios que mais concentraram o depósito dos dados de pesquisa, Federer *et al.* (2018) identificaram o *Figshare* (M), *Gene Expression Omnibus* (D), *Genbank* (D), *Dryad* (D), *Sequence Read Archive* (D), *No-repository web site*, *Institucional repository*, *Github* (D), *Dataverse* (M) e *Protein Databank* (D). Em um estudo mais recente, ainda na revista *Plos One* mas com um recorte temporal de 2014 a 2020, maior do que o estudo anterior de Federer *et al.* (2018), Jiao, Li e Fang (2022) identificaram que os dez repositórios que mais se destacaram foram o *Figshare* (M), *Gene Expression Omnibus* (D), *Dryad* (M), *GenBank* (D), *Sequence Read Archive* (D), *Open Science Framework* (D), *GitHub* (D), *Dataverse* (M), *Zenodo* (M) e *Bioproject* (D).

Ao realizar uma comparação entre esses dois estudos, pode-se observar que seis entre os dez repositórios se repetem, são eles: *Figshare* (M), *Gene Expression Omnibus* (D), *Genbank* (D), *Dryad* (D), *Sequence Read Archive* (D), *Github* (D) e o *Dataverse* (M). Esses resultados podem estar atrelados ao fato de ser o mesmo periódico analisado, no caso o *Plos One*, mas em períodos diferentes e com critérios de inclusão também diferentes para a elaboração do estudo.

Ao comparar os repositórios identificados pelos autores anteriores, com os dez que mais se destacaram neste trabalho, sete deles estão contemplados, sendo eles: *Sequence Read Archive* (D), *Genbank* (D), *Figshare* (M) e *Github* (D), *Open Science Framework* (D), *Zenodo* (M) e *Bioproject* (D). Esse resultado pode demonstrar que os autores reconhecem a importância que esses repositórios têm para as suas áreas, uma vez que os mesmos se mantêm entre os dez que mais concentram o depósito de dados.

Entretanto, embora haja uma constância nos repositórios utilizados neste trabalho e nos estudos de Federer *et al.* (2018) e Jiao, Li e Fang (2022), pode-se observar que o repositório *Mendeley Data* não aparece entre eles, enquanto neste estudo é o repositório que mais concentra conjuntos de dados e é utilizado por cerca

Em relação às áreas de cobertura, as Ciências da Vida concentraram 19 repositórios, com 35,36% (239) dos conjuntos de dados depositados. Os repositórios multidisciplinares concentraram 13 repositórios com 44,82% (303) dos conjuntos de dados. As Ciências da Vida e Ciências Naturais concentraram oito repositórios com 8,28% (56) dos conjuntos de dados. As Ciências Naturais concentraram sete repositórios com 1,92% (7) dos conjuntos de dados. As Ciências Humanas e Sociais, Ciências da Vida e Ciências Naturais concentraram dois repositórios com 3,55% (24) dos conjuntos de dados. Apenas um repositório abarcou a área das Ciências Humanas e Sociais com 1,63% (11) dos conjuntos de dados, sendo que o mesmo ocorreu com o repositório que cobriu as Ciências da Engenharia, o qual concentrou 2,96% (20) dos conjuntos de dados. O repositório que abarca as Ciências da Vida e Ciências da Engenharia concentrou 1,04% (1) dos conjuntos de dados

Jiao, Li e Fang (2022) haviam identificado que os repositórios disciplinares *Bioproject*, *Dryad*, *GenBank*, *Gene Expression Omnibus* e *Sequence Read Archive* (SRA), são predominantemente usados pelas Ciências da Vida e da Terra e pelas Ciências Biomédicas e da Saúde. Neste trabalho, essas duas áreas são concentradas na área das Ciências da Vida e, se considerado o percentual de conjuntos de dados depositados nesses repositórios, 39,73% (209) pode-se afirmar que os autores dessas áreas tendem a utilizar esses repositórios disciplinares.

Esses resultados podem indicar que há comportamentos diferentes entre as áreas do conhecimento, uma vez que as Ciências da Vida, por exemplo, concentram a maior parte dos repositórios, sendo também a área que concentra um quantitativo maior de repositórios por *data paper*. Por outro lado, as Ciências Humanas e Sociais foram representadas por apenas um repositório. Possivelmente para explicar isso, pode-se mencionar o estudo de McGillivray *et al.*, (2022) que destacam que os repositórios para depositar os dados produzidos por disciplinas humanas são escassos e nem sempre são adequados, o que pode indicar o motivo que leva os autores a recorrerem a repositórios multidisciplinares.

Em relação à quantidade de dados e de *data papers* produzidos pelas Ciências Humanas e Sociais, McGillivray *et al.* (2022) destacam que compartilhar conjuntos de dados como resultados da pesquisa nessa área ainda não é uma questão tão difundida quanto nas outras ciências, pois as Ciências Humanas e Sociais abarcam diversas áreas com características de publicação e objetos de estudos bem diversos. Ainda, os dados são frequentemente subvalorizados como

resultados de pesquisa, os *data journals* nem sempre são incluídos nas listas de periódicos reconhecidos pelo crédito acadêmico em muitos países e o fluxo de trabalho dos pesquisadores de humanidades normalmente não inclui a publicação dos dados, o que leva conjuntos de dados a não serem bem documentados e à dificuldade, para alguns pesquisadores, de acessar os recursos financeiros que lhes permitiriam publicar seus dados, sendo que isso se deve à falta de fundos para suportar os custos de armazenamento de dados ou a taxa de processamento de artigos.

Nesse sentido, Jiao, Li e Fang (2022), ao analisarem a relação entre os repositórios específicos e as áreas do conhecimento, verificaram que existem culturas de compartilhamento de dados diferentes entre os domínios de conhecimento, o que pode ser explicado por suas relações com dados e pelo paradigma de pesquisa orientado por dados.

Por fim, concorda-se com Jiao, Li e Fang (2022), que os repositórios de dados são agentes mediadores-chave entre publicações, sejam artigos científicos ou seja *data papers*, e dados de pesquisa e, por isso, é vital entender como eles são usados pelos pesquisadores.

4.4.3 Depósito de dados versus orientações dos periódicos

Os periódicos e seus *publishers* podem desempenhar um papel útil para possibilitar e concretizar os benefícios associados ao compartilhamento dos dados de pesquisa pelos pesquisadores, permitindo estes atendam aos requisitos definidos por seus financiadores e aumentem a capacidade de descoberta de dados compartilhados. Por isso, é preciso que os *publishers* e os periódicos definam políticas e processos robustos para o compartilhamento dos dados de pesquisa (Graf *et al.*, 2020).

Yi *et al.* (2023) revelam que dentre os periódicos coreanos ainda não houve uma ampla adoção de políticas de compartilhamento de dados, pois há um aumento potencial para a carga de trabalho de editores ou equipe editorial após a sua implementação. Esse não foi o caso dessa pesquisa pois, dos 38 periódicos identificados, apenas o *Cahiers Agricultures* não possuía uma política voltada para a disponibilização dos dados de pesquisa em repositórios. Esse dado pode estar relacionado à característica do objeto de análise da presente tese, uma vez que os

data papers e os repositórios fazem parte do ecossistema de publicação dos dados de pesquisa.

Assim, buscando discutir as orientações fornecidas pelos periódicos quanto à obrigatoriedade do depósito dos dados associados aos *data papers* em repositórios, bem como a adesão dos autores às tipologias de repositórios, disciplinar ou multidisciplinar, sugeridas pelos periódicos, foi elaborada a Tabela 11.

Como essa análise está voltada para a disponibilidade dos dados de pesquisa em repositórios, foram excluídos 11 periódicos que não os disponibilizaram dessa maneira, são eles: *BMC Bioinformatics*, *Cahiers Agricultures*, *Comparative cytogenetics*, *Frontiers in bioinformatics*, *Frontiers in Earth Science*, *Frontiers in Microbiology*, *Frontiers in Molecular Biosciences*, *Human mutation*, *Nature Conservation* e *Science of Nature*.

Dentre os 27 periódicos que disponibilizaram os dados em repositórios, os quais estão arrolados na Tabela 11, as orientações variaram entre “encorajar” a “obrigar” o depósito deles em repositórios de dados. Essa variação no nível de orientações realizadas pelos periódicos pode estar atrelada à sua adesão às Diretrizes de Promoção de Transparência e Abertura.

Dos periódicos analisados, apenas o *Gigascience* é editado por uma universidade e todos os demais são editados por *publishers* comerciais, o que sugere o interesse das editoras comerciais, principalmente, em acompanhar o ritmo e liderar mudanças associadas ao compartilhamento dos dados para que possam atender as demandas dos pesquisadores (Graf *et al.*, 2020; Silveira; Silva; Barbosa, 2023).

Tabela 11 - Orientações dos periódicos e a adesão dos autores

Periódico		Orientação do periódico			Total de datasets		Depósito na prática								
			Tipo de repositório				Não identificado o tipo de repositório			Multidisciplinar			Disciplinar		
	Tipo	Tipo de orientação	Multidisciplinar	Disciplinar	F	%	F	%	Adesão	F	%	Adesão	F	%	Adesão
Data In Brief	D	Obriga todos os dados	X	X	206	29,47	5	0,72	2,4	132	18,88	64,08	69	9,87	33,50
Scientific Data	D	Obriga todos os dados	X	X	251	35,91	6	0,86	2,4	83	11,87	33,07	162	23,18	64,54
Earth System Science Data	D	Obriga todos os dados	-	-	45	6,44	-	-	-	22	3,15	48,89	23	3,29	51,11
Ecology	M	Obriga todos os dados	X	X	34	4,86	1	0,14	2,9	26	3,72	76,47	7	1	20,59
Biodiversity Data Journal	D	Obriga todos os dados	X	X	26	3,72	-	-	-	3	0,43	11,54	23	3,29	88,46
Data	D	Obriga todos os dados	X	X	20	2,86	-	-	-	18	2,58	90	2	0,29	10
Global Ecology And Biogeography	M	Obriga todos os dados	X	X	12	1,72	1	0,14	8,3	6	0,86	50	5	0,72	41,67
Gigascience	D	Obriga todos os dados	X	X	24	3,43	-	-	-	-	-	-	24	3,43	100
BMC Research Notes	M	Encoraja a disponibilidade, obriga apenas para dados ômicos	-	X	8	1,14	-	-	-	4	0,57	50	4	0,57	50
Chemical Data Collections	D	Encoraja a disponibilidade	X	-	8	1,14	7	1	87,5	-	-	-	1	0,14	12,50
Proteomics	M	Obriga todos os dados	-	X	6	0,86	-	-	-	-	-	-	6	0,86	100
Geoscience Data Journal*	D	Obriga todos os dados	X	X	21	3	-	-	-	1	0,14	4,76	20	2,86	95,24
Frontiers in Ecology And Evolution	M	Obriga todos os dados	X	X	12	1,72	-	-	-	-	-	-	12	1,72	100
Microbiology Resource Announcements	M	Obriga todos os dados	X	X	9	1,29	-	-	-	-	-	-	9	1,29	100
Frontiers in Neuroscience	M	Obriga todos os dados	X	X	2	0,29	-	-	-	2	0,29	100	-	-	-
Journal of Biogeography	M	Encoraja a disponibilidade	X	-	3	0,43	3	0,43	100	-	-	-	-	-	-
Frontiers in Genetics	M	Obriga todos os dados	X	X	2	0,29	-	-	-	1	0,14	50	1	0,14	50
Annals of Forest Science	M	Encoraja a disponibilidade, obriga apenas para dados ômicos	X	X	1	0,14	-	-	-	1	0,14	100	-	-	-
Big Earth Data	D	Obriga todos os dados	X	X	1	0,14	-	-	-	-	-	-	1	0,14	100
Ecosistemas*	M	Obriga todos os dados	X	X	1	0,14	-	-	-	1	0,14	100	-	-	-
Frontiers in Marine Science	M	Obriga todos os dados	X	X	1	0,14	-	-	-	1	0,14	100	-	-	-

Periódico		Orientação do periódico			Total de <i>datasets</i>		Depósito na prática								
			Tipo de repositório				Não identificado o tipo de repositório			Multidisciplinar			Disciplinar		
	Tipo	Tipo de orientação	Multidisciplinar	Disciplinar	F	%	F	%	Adesão	F	%	Adesão	F	%	Adesão
Frontiers in Psychology	M	Obriga todos os dados	X	X	1	0,14	-	-	-	1	0,14	100	-	-	-
Frontiers in Public Health	M	Obriga todos os dados	X	X	1	0,14	-	-	-	-	-	-	1	0,14	100
Journal of Open Humanities Data	D	Obriga todos os dados	X	X	1	0,14	-	-	-	-	-	-	1	0,14	100
Marine Genomics	M	Obriga todos os dados	-	X	1	0,14	-	-	-	-	-	-	1	0,14	100
Scientific Reports	M	Obriga todos os dados	X	X	1	0,14	-	-	-	1	0,14	100	-	-	-
Zookeys	M	Encoraja a disponibilidade	X	X	1	0,14	-	-	-	-	-	-	1	0,14	100
Total					699	100	23	3,29	3,3	303	43,35	43,35	373	53,36	53,36

Fonte: dados da pesquisa

Notas: **Ecosistemas e Geoscience data journal*, obrigam o depósito apenas para os *data papers*. Nos demais a obrigação de depósito se estende a todos os tipos de documentos publicados

Colunas orientações dos periódicos: referem-se às orientações descritas pelo periódico quanto ao tipo de orientação para o depósito dos dados e quanto à tipologia de repositório que deve ser adotada pelos autores

Colunas total de *datasets*: referem-se ao total de *datasets* identificados nos *data papers* publicados pelos periódicos, isto é, referem-se à soma de *datasets* obtida a partir dos *data papers*

Colunas depósito na prática: referem-se à verificação na prática, se o autor seguiu a recomendação do periódico quanto ao tipo de repositório utilizar para depositar os dados. Verificação realizada a partir da comparação dos nomes dos repositórios citados nos *data papers* e classificados conforme o R3data e a tipologia citada nas políticas dos periódicos.

Os periódicos, *Chemical Data Collections*, *Journal of Biogeography* (Wiley), *ZooKeys* (Pensoft Publishers), *BMC Research Notes* (Elsevier) e *Annals of Forest Science* (Elsevier) encorajam o depósito, sendo que os dois últimos obrigam apenas o depósito de dados ômicos⁶⁴. Os demais periódicos (22) obrigam o depósito dos dados de pesquisa em repositórios.

Dentre os dez periódicos que são *data journals*, apenas o *Chemical Data Collections* (Elsevier) encoraja a disponibilidade dos dados de pesquisa e o *Earth System Science Data* (Copernicus Publications) não indica repositórios. Os demais (8) *data journals* obrigam o depósito dos dados de pesquisa e realizam indicações de repositórios a serem utilizados pelos autores. A maior quantidade de periódicos que apenas encorajam o depósito dos dados está entre os periódicos mistos, onde quatro de 17 periódicos realizam tal orientação, o que pode indicar que esse tipo de periódico ainda vem aderindo às práticas voltadas à abertura dos dados de pesquisa.

Os *data journals* concentram 86,27% (603) dos conjuntos de dados e os periódicos mistos concentram 13,73% (96) dos conjuntos de dados depositados em repositórios. Esses dados sugerem que o tipo de periódico e a orientação realizada influenciam na quantidade de *datasets* depositados, o que leva a inferir que os *data journals* são significativamente mais eficazes na promoção do depósito de conjuntos de dados por autores filiados às instituições brasileiras, enquanto que os periódicos mistos apresentam desafios maiores, mesmo quando adotam políticas editoriais semelhantes.

Em seguida foi verificado também se os periódicos estipulavam a tipologia de repositório específica, no caso dessa pesquisa multidisciplinar ou disciplinar, onde os autores deveriam depositar os conjuntos associados aos *data papers* publicados por eles.

Apenas o periódico *Earth System Science Data* (Copernicus Publications), não indicou nenhum tipo de repositório. Os periódicos *BMC Research Notes* (Elsevier), *Proteomics* (Wiley) e *Marine Genomics* (Elsevier) indicam apenas repositórios disciplinares. Os periódicos *Chemical Data Collections* (Elsevier) e *Journal of Biogeography* (Wiley) indicam apenas repositórios multidisciplinares e todos os demais periódicos (21) indicam ambas as tipologias, embora ressaltem que

⁶⁴ Dados ômicos são dados gerados pelas tecnologias de sequenciamento genético de alto rendimento, que permitem extrair informações genômicas, transcriptômicas e proteômicas em grande escala de maneira rápida, confiável e barata (Souza *et al.*, 2022).

sempre que possível deve-se usar o repositório disciplinar. Isso demonstra o interesse dos *publishers* em estabelecer diretrizes para o depósito em repositórios.

Foi verificado que os periódicos que “obrigam” o depósito de dados e que sugerem ambas as tipologias de repositórios tendem a concentrar uma maior quantidade de *datasets*, 97% (678), do que aqueles que apenas encorajam o depósito dos conjuntos de dados, com 3% (21). Com isso, é possível inferir que a obrigatoriedade do depósito dos dados pode estar associada à maior adesão dos autores filiados às instituições brasileiras às políticas dos periódicos, o que sugere uma efetividade das mesmas, visto que diretrizes do editor são uma ferramenta forte para garantir a qualidade da documentação dos conjuntos de dados, pois a comunidade de pesquisa os segue fortemente (Giner-Miguel; Gómez; Cabot, 2025).

Em relação ao depósito dos dados de pesquisa, 53,36% (373) deles foram depositados em repositórios disciplinares, 43,35% (303) foram depositados em repositórios multidisciplinares e para 3,29% (23) dos conjuntos de dados não foi possível identificar a tipologia do repositório. Houve uma diferença de cerca de 10% (70) entre o depósito em repositórios disciplinares e multidisciplinares, o que leva a três possíveis indicativos: primeiro os autores filiados às instituições brasileiras não priorizaram o depósito em repositórios disciplinares, conforme indicado pelos periódicos; segundo, não existem repositórios específicos para os dados documentados nos *data papers* e terceiro, na prática os periódicos não estão avaliando se os conjuntos foram depositados em repositórios aos quais os dados de pesquisa se referem.

Na sequência foi verificada a adesão dos autores filiados às instituições brasileiras às orientações dos periódicos, segundo a tipologia dos repositórios onde os dados associados aos *data papers* foram depositados. Para verificar essa adesão foi realizado um cálculo que considerou o total de *datasets* de cada periódico dividido pela quantidade total desses *datasets* depositados em cada tipo de repositório, disciplinar ou multidisciplinar. Foi observado que:

- a) dos 19 periódicos que **obrigam** o depósito dos dados e que sugerem o uso de repositórios de **ambas as tipologias**, em 11 os autores depositaram **todos os conjuntos** de dados em um dos repositórios indicados. Dentre os **repositórios multidisciplinares** a adesão variou entre em 4,76% a 76,47% e dentre os **repositórios disciplinares** variou

entre 10% a 95,24%. Esses dados indicam a efetividade da política mandatória, mas ausência de uma indicação clara quanto à especificidade dos repositórios que leva a uma ampla variação na escolha dos autores, com resultados bem variados entre os tipos de repositórios;

- b) dos dois periódicos que **obrigam** o depósito e que **sugerem** repositórios disciplinares, em ambos os autores depositaram os conjuntos na tipologia de repositório recomendada. Pode-se inferir que a indicação específica e obrigatória resulta em adesão total por parte dos autores ao depósito na tipologia do repositório;
- c) no caso do periódico que **obriga**, mas **não sugere** repositórios de nenhum tipo, a adesão variou de 48,89% para um repositório multidisciplinar a 51,11% para um repositório disciplinar. Esses dados associados aos resultados anteriores, reforçam que quando não houve indicações claras para os conjuntos de dados, os autores se dividiram entre os dois tipos de repositórios, embora tenha havido uma ligeira diferença entre eles;
- d) no caso do periódico que **encoraja** o depósito e **sugere** o uso de **repositórios disciplinares**, a adesão foi de 50% para a tipologia sugerida. Diante disso, pode-se verificar que no caso do encorajamento, a sugestão não foi seguida com consistência pelos autores, pois eles se dividiram igualmente entre os tipos, indicando que o incentivo sem obrigatoriedade tem impacto limitado;
- e) no caso dos dois periódicos que **encorajam** o depósito e **indicam repositórios multidisciplinares**, nenhum dos conjuntos foi depositado na tipologia de repositório indicada. Foi verificado que a sugestão dos periódicos não foi seguida, novamente reforçando que apenas o encorajamento, resultou numa baixa adesão por partes por autores filiados às instituições brasileiras e
- f) no caso dos dois periódicos que **encorajam** o depósito, e **sugerem** ambas as **tipologias de repositório**, a efetividade foi de 100% para ambas as tipologias, ou seja, num dos periódicos, todos os conjuntos de dados foram depositados num repositório disciplinar e no outro periódico os conjuntos foram depositado num repositório multidisciplinar. Esses

resultados indicam que, mesmo sem obrigatoriedade, os autores aderiram à sugestão do periódico.

Embora esses resultados sejam limitados pela pequena quantidade de conjuntos de dados depositados no caso de alguns periódicos e pela ausência da análise aos conjuntos de dados propriamente ditos, os resultados levam a inferir que, embora a maioria dos periódicos exija o depósito de todos os conjuntos de dados, a efetividade dessas recomendações entre os periódicos varia significativamente.

Os autores filiados às instituições brasileiras tendem a seguir a obrigatoriedade dos periódicos realizando o depósito dos dados. No entanto, quando o periódico não estipula os repositórios de maneira clara, tende haver uma maior variação dentro a tipologia selecionada pelo autor.

Todos esses dados permitem deduzir que os *data papers* são os principais objetos de preocupação dos periódicos, tornando os conjuntos de dados objetos secundários (Candela *et al.*, 2015). Portanto, é preciso um posicionamento mais efetivo dos periódicos quanto à avaliação se os conjuntos de dados estão sendo depositados nos repositórios aderentes à sua tipologia.

4.4.4 Considerações da seção

A partir da análise, dos 634 *data papers* publicados em 38 periódicos foram identificadas diversas maneiras de disponibilizar os dados de pesquisa na tipologia de documento analisada, embora tenha havido um predomínio dos repositórios de dados exclusivamente em cerca de 55% dos *data papers*. Esses diferentes mecanismos podem estar atrelados à adesão gradual dos *publishers* dos periódicos às práticas abertas da ciência e às diferenças disciplinares.

Em relação às características dos repositórios de dados utilizados, foram identificados 66 repositórios diferentes, sendo a maioria (59) mantida por instituições sem fins lucrativos. Embora os mantenedores comerciais sejam minoria (4), eles concentram quase 30% (199) dos conjuntos de dados depositados.

Apenas seis repositórios estão localizados no Brasil e os demais estão concentrados no Norte Global, sugerindo novamente o monopólio dessa região no que condiz à gestão dos dados de pesquisa, assim como ocorreu com os periódicos identificados nesta pesquisa.

Em geral, os repositórios identificados possuem uma gratuidade limitada quanto ao tamanho dos conjuntos e à revisão dos dados de pesquisa. A partir disso, pode-se presumir que a gratuidade, atrelada à capacidade de armazenamento dos repositórios, pode ser fator que motivou os autores filiados às instituições brasileiras a depositarem seus dados de pesquisa.

Em relação às licenças e acessos, os repositórios foram bastantes flexíveis nos formatos a serem adotados pelos depositantes dos dados, o que por um lado pode ser positivo pois amplia a possibilidade de reuso dos dados e por outro lado, pode ser negativo, pois em casos, onde há não limitações quanto a privacidade dos dados, o reuso pode ser inviabilizado.

Em relação à quantidade de repositórios utilizadas, varia de 1 a 8, sendo que a adoção de apenas um repositório foi a prática mais adotada pelos autores filiados às instituições brasileiras, o que foi identificado em quase 90% dos *data papers*.

Em relação à tipologia dos repositórios, 42 eram disciplinares, concentrando cerca de 54% (376) dos conjuntos de dados, 13 multidisciplinares, que concentraram 42,92% (300) dos conjuntos de dados e 11 repositórios não estavam indexados no R3data, não tendo sua tipologia classificada, sendo que esses concentraram 3,29% (23) dos conjuntos de dados.

Especificamente em relação às áreas dos repositórios, aqueles ligados às Ciências da Vida foram os preferidos dos autores filiados às instituições brasileiras, o que está relacionado às características das áreas e temas identificados nos *data papers*. Esses dados sugerem que, embora haja comportamentos diferentes entre as áreas do conhecimento, os autores tendem a depositar os dados nos repositórios associados a elas.

Foi verificado ainda se os autores filiados às instituições brasileiras seguiram as orientações dos repositórios quanto ao depósito dos dados e à tipologia recomendada pelo periódico. Embora as orientações dos periódicos tenham variado entre “encorajar” a “obrigar”, o depósito dos dados e as políticas quanto às tipologias dos repositórios também foram heterogêneas. Os dados indicam que os autores filiados às instituições brasileiras tendem a seguir a obrigatoriedade dos periódicos, realizando o depósito dos dados. No entanto, quando o periódico não estipula os repositórios de maneira clara, tende a haver uma maior variação dentre a tipologia selecionada pelo autor.

5 CONSIDERAÇÕES FINAIS

O objetivo geral desta tese foi caracterizar a publicação de dados de pesquisa através de *data papers* por parte dos autores filiados às instituições brasileiras. A análise realizada permitiu compreender não apenas o volume e a frequência dessas publicações, mas também a filiação dos autores, fontes de financiamentos, periódicos onde os *data papers* foram publicados e os repositórios onde os conjuntos de dados associados foram depositados.

Com isso pode-se elencar como características da publicação de *data papers*, a jovialidade das publicações brasileiras, sendo os primeiros a serem publicados em 2014, com uma demanda crescente principalmente nas Ciências Físicas, com temas ligados à Ecologia e a Biodiversidade. Essa característica está relacionada a necessidade de se adequar aos mandatos de financiadores, sobretudo os estrangeiros, que vem exigindo o compartilhamento dos dados de pesquisa.

A autoria dos *data papers*, por meio da filiação institucional mostrou que ela se dá em sua maioria entre as instituições nacionais e instituições estrangeiras, localizadas em países com maior capacidade científica, tecnológica e com incentivos voltados à Ciência Aberta, predominantemente localizados no Norte Global. Essas observações sugerem que os *data papers* também pode ser publicações que possibilitam a colaboração científica entre pesquisadores de diferentes instituições do Brasil e do mundo, o que pode ampliar a visibilidade científica do país e das pesquisas realizadas pelos autores filiados.

Os financiamentos se caracterizam por originarem-se de instituições nacionais e estrangeiras, majoritariamente de fundo público. As características identificadas apontam que os *data papers* podem servir para indicar oportunidades de pesquisas nas temáticas que envolvem as mudanças climáticas e do meio ambiente, uma vez que esses foram os temas que mais concentraram publicações pelos pesquisadores filiados às instituições brasileiras.

A publicação dos *data papers* se concentrou em *data journals*, mantidos por *publishers* comerciais, localizados no Norte Global e que cobram taxas de publicação. Esses resultados indicam que os autores filiados às instituições brasileiras, têm procurado periódicos especializados para publicar seus dados de pesquisa, via *data papers*, uma vez que esses realizam práticas consolidadas pela comunicação científica como a revisão por pares e a indexação em bases de dados, sendo que a

maioria deles são considerados pelos principais indicadores de avaliação, como o SJR e JIF. Porém, esse tipo de periódico ainda carece de consolidação no sistema de avaliação da pós-graduação brasileira.

Todas essas características demonstram que os *publishers* comerciais aos poucos foram criando periódicos especializados, os *data journals*, ou criando seções especiais para publicar *data papers*, permitindo, assim, se adequar ao movimento de abertura da ciência e garantindo sua hegemonia nesse contexto. Com isso, os dados indicam que os autores filiados às instituições brasileiras, têm buscado nos *data journals* canais para publicar seus dados e, dessa forma obterem o reconhecimento científico. Todavia, os impactos na avaliação brasileira desses canais ainda são incipientes, visto que o principal *data journal* utilizado pelos autores, o *Data in Brief* (Elsevier), não foi considerado na avaliação do Qualis do triênio 2017-2020.

Ao analisar as características relacionados ao depósito dos dados de pesquisa associados aos *data papers* publicados por autores filiados às instituições brasileiras, o resultado mais surpreendente é que em apenas metade dos *data papers* os dados associados foram disponibilizados em repositórios. Dentre as características dos repositórios identificados, houve uma concentração de conjuntos de dados depositados naqueles mantidos por *publishers* comerciais, possivelmente pelas facilidades de utilização, atribuição de DOI e cobranças parciais.

Além disso, são repositórios localizados em países do Norte Global. Cerca de 2% dos conjuntos de dados associados aos *data papers* foram depositados em repositórios nacionais, embora o Brasil esteja iniciando a criação de infraestruturas para o depósito dos dados. Os resultados evidenciam a necessidade de políticas e ações que dêem visibilidade e povoamento aos repositórios de dados nacionais.

Ao se associar a publicação de dados de pesquisa aos *data papers*, *data journals* ou periódicos mistos e repositórios de dados, pôde-se observar como as políticas de financiamento vem modificando a comunicação científica, inserindo os dados no seu cerne e, assim, os *publishers*, em especial os comerciais, rapidamente incorporam esses mandatos às suas práticas, com a exigência de documentos que atestem a disponibilidade de dados, criação de repositórios de dados, criação de *data journals* e de *data papers*, para recompensar os autores pela publicação dos dados de pesquisa.

Portanto, considerando as características apresentadas pode-se afirmar que a tese cumpriu com o seu objetivo geral ao utilizar os *data papers* para caracterizar o

ecossistema associado à publicação dos dados de pesquisa no contexto brasileiro. No entanto, se reconhece como limitação principal desta tese que foram utilizados *data papers* indexados nas bases de dados Scopus e Web of Science, bases comerciais que representam apenas parte do conhecimento produzido no mundo.

5.1 CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS TEMPORAIS E TEMÁTICAS DOS *DATA PAPERS*

A análise das características temporais e temáticas dos *data papers* publicados por autores filiados às instituições brasileiras evidenciou tendências importantes para compreender o cenário brasileiro de publicação de dados de pesquisa. Do ponto de vista temporal, observou-se que a publicação ainda é jovem, tendo iniciado em 2014, crescendo de maneira gradual em todas as grandes áreas do conhecimento, com um aumento de 58% entre 2021 e 2024. Isso demonstra uma tendência de crescimento por parte dos autores filiados às instituições brasileiras na maioria das áreas do conhecimento.

Esse cenário de crescimento possivelmente vem sendo influenciado pelas políticas mandatórias das agências de financiamento e dos periódicos, o que pode indicar que há, também influência de políticas externas, uma vez que no Brasil ainda são incipientes aquelas relacionados às agências de financiamento nacionais e aos periódicos.

No aspecto temático, verificou-se que, embora desigual, o crescimento na produção dos *data papers* tem ocorrido em todas as áreas do conhecimento. Os temas são variados, mas os achados na pesquisa revelaram que os temas de interesse global, como mudanças climáticas, conservação, genética e ecologia, sequenciamento de RNA e os temas que envolvem doenças arbovirais como Chagas e Leishmaniose, têm sido uma tendência nas publicações dos autores filiados às instituições brasileiras.

Essa distribuição temática e a tendência na produção dos *data papers* reforça a relevância desses documentos como instrumento de valorização de temas de interesse mundial e nacional, o que pode contribuir para a elaboração de políticas de financiamento voltadas para essas áreas e temas.

5.2 CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DAS FILIAÇÕES DOS AUTORES E DOS FINANCIADORES

A análise das características de filiação dos autores dos *data papers* revelou elementos importantes para compreender o papel das colaborações entre instituições e países na sua produção. Os dados mostraram uma ampla participação institucional, com 2530 instituições diferentes, distribuídas entre 113 países, podendo indicar que os *data papers* possuem um potencial de ampliar a colaboração internacional. A presença predominante dos países localizados no Norte Global na coautoria dos *data papers* indica uma tendência de colaboração entre o Brasil e esses países.

A tipologia predominante dessas instituições são universidades que têm em comum a presença de políticas institucionais voltadas ao acesso aberto e o compartilhamento dos dados de pesquisa. Esses dados apontam que as políticas institucionais voltadas para incentivo e reconhecimento das práticas individuais dos seus pesquisadores podem ter resultados positivos e, além disso, reforça o papel que as universidades podem ter na disseminação das práticas voltadas à Ciência Aberta.

O financiamento público das pesquisas foi mencionado em mais de 70% dos *data papers* e em quase 50% deles houve o financiamento de pelo menos uma agência nacional. Esse dado indica que as agências nacionais, têm um papel importante no incentivo ao depósito dos dados de pesquisa.

A influência das agências estrangeiras nos financiamentos das pesquisas que originam os *data papers* pode indicar que os mandatos voltados ao compartilhamento de dados de pesquisa podem ter influenciado a publicação dos *data papers* e destacam a necessidade de o Brasil criar suas políticas mandatórias que incentivem o depósito dos dados obtidos a partir de pesquisas que receberam recursos públicos, em repositórios do país.

5.3 CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DOS PERIÓDICOS

A análise das características dos periódicos que publicaram os *data papers*, evidenciou que os *data journals*, tiveram preferência pelos autores filiados às instituições brasileiras, em detrimento dos periódicos mistos, sugerindo uma contribuição desses canais para a publicação dos dados de pesquisa. Porém sete dos dez *data journals* identificados não possuíam classificação no sistema Qualis,

indicando uma falta de consolidação desses canais no sistema de avaliação da pós-graduação brasileira.

Os *data journals* identificados são publicados por *publishers* comerciais, cujo modelo de acesso aberto predominante entre os periódicos identificados foi a via ouro, com cobrança de taxas APC que, ao longo dos anos, tem se aproximando dos valores cobrados pelos periódicos tradicionais.

Esses dados revelam uma lacuna nacional na oferta de periódicos que aceitam *data papers*, levando os autores filiados às instituições brasileiras a buscarem alternativas internacionais, especialmente em periódicos comerciais de cobertura multidisciplinar. Além disso, demonstram a ausência da consolidação dos *data journals* e da publicação dos dados nas discussões da pós-graduação brasileira.

Os modelos tradicionais de revisão por pares, como o simples cego e o duplo cego, continuam sendo os mais utilizados. Embora todos os periódicos afirmem realizar revisão por pares, apenas nove detalham suas políticas específicas para *data papers* e *datasets* e, desses, apenas sete *data journals* informam os critérios utilizados. Esses dados indicam um cenário de falta de práticas padronizadas de critérios de avaliação dos *data papers* por parte dos periódicos, o que repercute para uma baixa transparência nas políticas de revisão, podendo acarretar em desconfiças quanto à revisão dos *data papers* e dos dados.

5.4 CONSIDERAÇÕES SOBRE AS CARACTERÍSTICAS DOS REPOSITÓRIOS DE DADOS

As análises das características dos repositórios de dados revelaram as múltiplas estratégias utilizadas para disponibilizar os dados, embora tenha prevalecido o uso exclusivo de repositórios de dados em aproximadamente 55% dos *data papers*. Essa diversidade de mecanismos pode estar relacionada à adesão gradual dos *publishers* às práticas da Ciência Aberta, bem como às especificidades disciplinares dos periódicos envolvidos.

No que se refere às características dos repositórios utilizados, entre os 66 repositórios identificou-se que são majoritariamente mantidos por instituições sem fins lucrativos. Apesar dessa característica, os quatro repositórios comerciais concentram cerca de 30% (199) dos conjuntos de dados depositados, evidenciando a penetração dos *publishers* comerciais no ecossistema de dados de pesquisa.

Geograficamente, apenas seis repositórios estão sediados no Brasil, enquanto os demais se concentram no Norte Global, demonstrando o domínio dessa região na gestão dos dados de pesquisa, fenômeno também observado na distribuição dos periódicos analisados.

A gratuidade oferecida pelos repositórios, embora limitada em termos de tamanho dos conjuntos de dados e revisão dos conteúdos, parece ser um fator determinante na escolha dos autores filiados às instituições brasileiras. A capacidade de armazenamento, aliada à ausência de custos, pode ter influenciado diretamente na decisão de seleção de depósito dos dados.

A presença significativa dos repositórios disciplinares nos resultados revela a necessidade de se avaliar a criação de repositórios disciplinares ou multidisciplinares no país ou, ainda, a necessidade de se pensar na criação de repositórios disciplinares mantidos em parcerias entre os países para garantir o povoamento, a visibilidade e a sua manutenção. A pesquisa demonstrou que, apesar de existirem variações entre as áreas do conhecimento, os autores tendem a depositar seus dados em repositórios especializados nas respectivas disciplinas.

Por fim, foi verificado o grau de conformidade dos autores filiados às instituições brasileiras em relação às orientações dos periódicos quanto ao depósito dos dados e à tipologia recomendada. As diretrizes variaram entre o incentivo e a obrigatoriedade, sendo que os dados indicam uma tendência dos autores a seguir as exigências dos periódicos. No entanto, quando não há especificação clara sobre os repositórios a serem utilizados, observa-se uma maior diversidade na tipologia escolhida pelos autores. Com essa análise, evidencia-se que periódicos com políticas claras e obrigatórias apresentam maior aderência, enquanto os que adotam posturas mais flexíveis enfrentam desafios para garantir o cumprimento.

5.5 RECOMENDAÇÕES E ESTUDOS FUTUROS

Essa pesquisa não se esgota com as análises apresentadas, muito pelo contrário, ela demonstra que o cenário voltado a publicação de dados de pesquisa é amplo e traz implicações para as instituições de filiação dos autores, sobretudo para as universidades, agências de financiamento, periódicos e repositórios de dados, abrindo possibilidades para recomendações e estudos futuros. As sugestões estão arroladas a partir dos estudos e recomendações de cada elemento explorado na tese.

Em relação aos autores e práticas voltadas à abertura da ciência, as pesquisas futuras podem explorar as motivações dos autores para publicar *data papers*, considerando tempo de carreira, áreas do conhecimento e percepção de legitimidade desse formato na sua área de atuação. Recomenda-se investigar como o compromisso com práticas abertas pode alterar o fazer científico e pesquisar o conhecimento dos autores sobre a curadoria, descrição e disponibilização dos dados em repositórios reconhecidos. Para tanto, questionários e entrevistas podem ser utilizados para compreender barreiras e incentivos percebidos pelos pesquisadores.

Em relação ao papel das agências de financiamento, sugere-se investigar, a partir de pesquisas documentais, se os financiadores possuem mandatos que influenciam a publicação de *data papers* e o compartilhamento dos dados e instruções voltadas para o depósito dos dados em repositórios disciplinares ou multidisciplinares. Recomenda-se a realização de estudos que proponham o desenvolvimento de ferramentas de acompanhamento de adesão a esses mandatos por parte dos pesquisadores, permitindo identificar se as políticas de gestão de dados podem evoluir para incluir mecanismos de validação dos dados e, assim, contribuir para a efetividade na promoção da Ciência Aberta.

Em relação aos editores e periódicos, sugere-se analisar o impacto das APCs e dos custos de depósito de dados como barreiras à publicação de *data papers* e depósito dos *datasets*, bem como investigar possíveis falhas nos processos de revisão, considerando que poucos periódicos descreveram critérios claros de avaliação dos *data papers* e dos *datasets*. Recomenda-se estudos que analisem como os periódicos podem aprimorar a revisão dos conjuntos de dados, exigindo pareceres específicos e capacitação dos revisores para avaliar a descrição e a qualidade dos dados. A partir dos dados coletados nessa pesquisa, pode-se analisar se os *datasets* coletados estão aderentes aos princípios FAIR.

Em relação aos repositórios de dados, sugere-se investigar junto às universidades a capacidade de criação e manutenção dessas infraestruturas, bem como, analisar a possibilidade de criação de repositórios em parcerias entre as instituições com menor capacidade econômica e infraestrutura. Recomenda-se verificar estratégias de povoamento e políticas que estimulem o uso dos repositórios de dados nacionais já criados, como o Deposita Dados, de modo a garantir a soberania sobre os dados produzidos e financiados por instituições brasileiras.

Nessa pesquisa foi possível verificar que o Norte Global destacou-se em todos os aspectos que envolveram a publicação dos dados de pesquisa, tanto nas colaborações entre autores, a partir das instituições, quanto nos periódicos e repositórios onde os dados foram coletados, o que denota uma maior infraestrutura da região em publicar e armazenar os dados. Porém acende um alerta, sobre o poder que essas nações podem acumular no futuro em relação aos dados de pesquisa, aliado ao advento da inteligência artificial.

Por fim, salvo as limitações já apresentadas, esta pesquisa demonstrou para a área da Ciência da Informação que os *data papers* podem servir como fontes de informação para os estudos voltados ao comportamento das áreas quanto à publicação dos dados de pesquisa, o que pode contribuir para o desenvolvimento da Ciência Aberta. Além disso, trouxe como contribuições para a área trabalhos apresentados em eventos nacionais e internacionais, destacados nas subseções 3.4.1 a 3.4.4, além de disponibilizar os dados coletados, assim como os dados utilizados para gerar as figuras.

REFERÊNCIAS

- AGUIRRE-LIGÜERA, N. *et al.* Cenário da produção científica latino-americana sobre ciência aberta. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 23., Aracaju, 2023. **Anais [...]**. Aracaju, 2023. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxxiiienancib/paper/viewFile/1794/1198>. Acesso em: 23 jul. 2025.
- ALVAREZ, G. R.; CAREGNATO, S. E. Financiamento nas publicações brasileiras com agradecimentos indexadas na Web of Science. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 26, n. Especial, p. 1–21, 2021. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/78358>. Acesso em: 29 jul. 2025.
- ANGLADA, L.M.; ABADAL, E. Open access: a journey from impossible to probable, but still uncertain. **El Profesional de la Información**, Madrid, v. 32, n. 1, 2023. Disponível em: <https://doi.org/10.3145/epi.2023.ene.13>. Acesso em: 13 nov. 2025.
- ANSELMO, A. C. B.; RODRIGUES, R. S.; MUGNAINI, R. Periódicos científicos: Acesso aos artigos brasileiros. **Informação & Informação**, Londrina, v. 27, n. 4, p. 32-59, 2022. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/47179>. Acesso em: 28 out. 2025.
- ANNOUNCEMENT: Launch of an online data journal. **Nature**, v. 502, n. 142, 2013. Disponível em: <https://doi.org/10.1038/502142>. Acesso em: 23 nov. 2025.
- ARAÚJO, C. A. A. Correntes teóricas da ciência da informação. **Ciência da Informação, Brasília**, v. 38, n. 3, p.192-204, set./dez. 2009. Disponível em: <https://www.scielo.br/j/ci/a/qhsrgPL7T6RbKKVbMwrPMNb/?format=html&stop=next&lang=pt#>. Acesso em: 30 ago. 2023.
- ARAÚJO, C. A. A. **O que é Ciência da Informação**. Belo Horizonte: KMA, 2018.
- ARAUJO, L.; PRINCIPE, E. Diretrizes editoriais para a publicação de artigos de dados em Ciências da Saúde. **BiblioCanto**, Natal, v. 9, n. 2, p. 114–122, 2023. Disponível em: <https://periodicos.ufrn.br/bibliocanto/article/view/33586>. Acesso em: 7 out. 2025.
- ARAUJO, L.; PRINCIPE, E. ALENCAR, M.S.M. Data papers brasileiros: análise da publicação de artigos de dados indexados na Web of Science e Scopus. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 21., Rio de Janeiro, 2021. **Anais [...]**. Aracaju, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/177>. Acesso em: 24 out. 2025.
- ARAUJO, L.; PRINCIPE, E. ALENCAR, M.S.M. Panorama da publicação de artigos de dados em periódicos científicos: uma revisão sistemática. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*,

22., Porto Alegre, 2022. **Anais [...]**. Porto Alegre, 2022. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxiienancib/paper/view/1045>. Acesso em: 24 out. 2025.

ARITA, M. *et al.* Publishing in the Open Access and Open Science era. **Genes to cells**, v. 29, n. 4, p. 275-281, 2024. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/gtc.13100>. Acesso em: 21 ago. 2024.

ASSANTE, M. *et al.* Are scientific data repositories coping with research data publishing? **Data Science Journal**, v. 15, n. 6, p. 79–83, 2016. Disponível em: <http://dx.doi.org/10.5334/dsj-2016-006>. Acesso em: 27 maio 2025.

AVENTURIER, P.; ALENCAR, M. C. F. Os desafios dos dados de pesquisas abertos. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 10, n. 3, p. 1-19, jul./set., 2016. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069>. Acesso em: 14 nov. 2025.

BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 9.ed. Florianópolis: Ed. da UFSC, 2015.

BARATA, R. B. Estratégias de avaliação da produção científica: reflexões sobre suas limitações e complementaridades. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 28, p.1-22, 2023. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/42242/39205>. Acesso em: 4 de out. 2024.

BATTHYÁNY, K.; VOMMARO, P. Introducció. *In*: BATTHYÁNY, K.; VOMMARO, P.; ROVELLI, L. (ed.). **Iniciativas y regulaciones multinivel para la Ciencia Abierta**: Infraestructuras abiertas y sistemas de evaluación em Iberoamérica. Madrid: Fundación Carolina, 2023. p. 5-6. Disponível em: <https://biblioteca-repositorio.clacso.edu.ar/bitstream/CLACSO/249213/1/Iniciativas-regulaciones.pdf>. Acesso em: 1 dez. 2023.

BERTIN, P. R. B. **Governo Aberto na Prática**: Resultados dos Compromissos - Ciência Aberta. Entrevista. [Brasília], 07 jun. 2021. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2021/6/governo-aberto-na-pratica-resultados-dos-compromissos-compromisso-ciencia-aberta>. Acesso em: 31 jul. 2025.

BOGADO, A. C.; HILÁRIO, C. M.; GRÁCIO, M. C. C. Acesso aberto e financiamento de pesquisa: a pesquisa em hanseníase do Instituto Lauro de Souza Lima (2016-2022). *In*: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 9., 2024, Brasília. **Anais [...]**. Brasília, 2024. Disponível em: <https://ebbc.inf.br/ojs/index.php/ebbc/article/view/252/209>. Acesso em: 13 ago. 2025.

BJÖRK, B. C.; SOLOMON, D. Article processing charges in OA journals: relationship between price and quality. **Scientometrics**, v. 103, p. 373-385, 2015.

Disponível em: <https://doi.org/10.1007/s11192-015-1556-z>. Acesso em: 13 nov. 2025.

BONETTI, L. G; ARAKAKI, A. C. S. Princípios FAIR e a avaliação de datasets no repositório institucional da UFSCAR. **Informação & Informação**, Londrina, v. 27, n. 1, p. 485-510, 2022. Disponível: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/44328>. Acesso em: 10 jan. 2024.

BORREGO, A. OpenAlex: breve guia de consulta a través de R. **Infonomy**, v. 2, n. 1, e24011, 2024. Disponível em: <https://infonomy.scimagoepe.com/index.php/infonomy/article/view/26/51>. Acesso em: 4 abr. 2024.

BRASIL. **Balanco**: plano de ação do Brasil na parceria para Governo Aberto. [Brasília], 2013. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/1o-plano-de-acao/balanco-primeiro-plano.pdf>. Acesso em: 15 dez. 2023.

BRASIL. Ministério da Transparência e Controladoria-Geral da União. **4º plano de ação nacional em Governo Aberto**. Brasília, 2018a. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/4-plan-acao-nacional-atualizado2020.pdf>. Acesso em: 15 dez. 2023.

BRASIL. Controladoria Geral da União. **Inovação e governo aberto na ciência - 1º oficina de cocriação**. Brasília, 2018b. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/inovacao-e-governo-aberto-na-ciencia-1o-oficina-de-cocriacao>. Acesso em: 15 dez. 2023.

BRASIL. Controladoria-Geral da União. **5º plano de ação nacional em Governo Aberto**. Brasília, 2021a. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/5o-plano-de-acao-brasileiro/5-plano-acao-nacional-30-03-2022.pdf>. Acesso em: 15 dez. 2023.

BRASIL. Controladoria Geral da União. **Transparência em Ciência: novos mecanismos de avaliação para o avanço da Ciência Aberta - Primeira etapa das Oficinas de Cocriação**. Brasília, 2021b. Disponível em: https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/5o-plano-de-acao-brasileiro/compromisso-8-transparencia-em-ciencia-novos-mecanismos-de-avaliacao-para-o-avanco-da-ciencia-aberta/primeira-etapa/copy_of_combate-a-corrupcao-no-setor-publico-primeira-etapa-das-oficinas-de-cocriacao-1. Acesso em: 15 dez. 2023.

BRASIL. Controladoria Geral da União. **Relatório de autoavaliação final**. Brasília, [2021?]. Disponível em: https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/5o-plano-de-acao-brasileiro/relatorio-final-de-autoavaliacao_5pan-ultimaversao.pdf. Acesso em: 31 jul. 2025.

BRASIL. Controladoria Geral da União. **6º Plano de Ação Nacional em Governo Aberto**. Brasília, 2023a. Disponível em: <https://www.gov.br/cgu/pt-br/governo->

[aberto/a-ogp/planos-de-acao/6deg-plano-de-acao-brasileiro/brazil_action-plan_2023-2027_december_pt.pdf](#). Acesso em: 9 jul. 2025.

BRASIL, A. Rethinking a national classification of research and graduate education. **Profesional de la información**, v. 32, n. 2, e320224, 2023b. Disponível em: <https://doi.org/10.3145/epi.2023.mar.24>. Acesso em: 22 set. 2023.

BREURE, L; DOORN, P.; VOORBIJ, H. Data Showcases: the Data Journal in a Multimodal World. **International Journal of Digital Curation**, v. 17, n. 1, p. 1-24, 2022. Disponível em: <https://ijdc.net/ijdc/article/view/789>. Acesso em: 23 nov. 2025.

BUTLER, L. *et al.* The oligopoly's shift to open access: how the big five academic publishers profit from article processing charges. **Quantitative Science Studies**, [S. l.], v. 4 n. 4, p. 778-799, 2023. Disponível em: <https://direct.mit.edu/qss/article/4/4/778/118070/The-oligopoly-s-shift-to-open-access-How-the-big>. Acesso em: 2 set. 2024.

CABALLERO-RIVERO, A.; SÁNCHEZ-TARRAGÓ, N.; SANTOS, R. N. M. dos. Práticas de Ciência Aberta da comunidade acadêmica brasileira: estudo a partir da produção científica. **Transinformação**, Campinas, v.31, e190029, 2019. Disponível em: <http://www.redalyc.org/articulo.oa?id=384365067026>. Acesso em: 26 dez. 2023.

CALLAGHAN, S. *et al.* Making data a first class scientific output: data citation and publication by NERC's Environmental Data Centres. **International Journal of Digital Curation**, v. 7, n. 1, p. 107–113, 2012. Disponível em: <https://ijdc.net/index.php/ijdc/article/view/208/277>. Acesso em: 26 dez. 2023.

CANDELA, L. *et al.* Data journals: A survey. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1747-1762, 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.23358>. Acesso em: 28 jul. 2023.

CAREGNATO, S.E. *et al.* Práticas e percepções dos pesquisadores brasileiros sobre serviços de acesso aberto a dados de pesquisa. **Liinc em Revista**, Brasília, v. 15, n. 2, 2019. Disponível em: <https://revista.ibict.br/liinc/article/view/4771>. Acesso em: 10 jan. 2024.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da União Europeia. **BIBLOS** - Revista do Instituto e Ciências Humanas e da Informação, Rio Grande, v. 31, n. 1, p. 73-98, 2017. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789>. Acesso em: 10 jan. 2024.

CELLARD, A. Análise documental. In: POUPART, J. *et al.* **Pesquisa qualitativa: enfoques epistemológicos e metodológicos**. 3. ed. Petrópolis, Vozes, 2012. p. 295-316.

CENTER FOR OPEN SCIENCE (FOR). **TOP 2025 provides a suite of tools to increase the verifiability of research claims**. Washington, [2025]. Disponível em: <https://www.cos.io/initiatives/top-guidelines>. Acesso em: 26 ago. 2025.

CHAVAN, V., PENEV, L.. The *data paper*: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, v. 12, supl. 15, p. 1-12, 2011. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S15-S2#citeas>. Acesso em: 19 jun. 2023.

CHEN, Y. An analysis of characteristics and structures embedded in data papers: a preliminary study. **Libellarium**, v.9, n.2, p.145-156, 2017. Disponível em: <https://morepress.unizd.hr/journals/index.php/libellarium/article/view/3820>. Acesso em: 7 nov. 2025.

CHEN, P.-Y.; LI, K.; JIAO, C. A Preliminary Analysis of Geography of Collaboration in *Data papers* by S&T Capacity Index. In: PROCEEDINGS OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 85, Pittsburgh, 2022. **Proceedings** [...]. Pittsburgh, 2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pr2.676>. Acesso em: 19 jun. 2023.

CHO, J. Study of Asian RDR based on re3data. **The Electronic Library**, v. 37, n. 2, p. 302-313, 2019. Disponível em: <https://doi-org.ez46.periodicos.capes.gov.br/10.1108/EL-01-2019-0016>. Acesso em: 16 de jul. 2025.

CLARIVATE. **Journal Citation Report Help**. [S./], c2021. Disponível em: <https://jcr.help.clarivate.com/Content/jcr3-glossary.htm>. Acesso em: 28 dez. 2023.

CLARIVATE. **Research Areas (Categories / Classification)**. [S./], c2023. Disponível em: <http://webofscience.help.clarivate.com/en-us/Content/research-areas.html?Highlight=research%20areas>. Acesso em: 6 fev. 2023.

CLARIVATE. **[Base de dados] Web of Science**. [S./], c2024. Disponível em: <https://www-webofscience.ez46.periodicos.capes.gov.br/wos/woscc/basic-search>. Acesso em: 15 dez. 2023.

CLINIO, A. Ciência Aberta na América Latina: duas perspectivas em disputa. **Transinformação**, Campinas, v. 31, p. 1-12, 2019. Disponível em: <https://www.scielo.br/j/tinf/a/pPH6wwwxN6rGhyVJM83pGSnp/?format=pdf&lang=pt>. Acesso em: 28 jul. 2023.

CODINA, L. OpenAlex: ¿una alternativa a Scopus y Web of Science? **Lluís Codina**, 2024. Disponível em: <https://www.lluiscodina.com/openalex-scopus/>. Acesso em: 4 abr. 2024.

COMMITTEE FOR OPEN SCIENCE. **Second French Plan for Open Science: Generalising open science in France**. [S./, 2025?]. Disponível em: <https://www.ouvrirelascience.fr/second-national-plan-for-open-science/>. Acesso em: 22 jul. 2025.

CONNAWAY, L. S.; POWELL, R. R. **Basic research methods for librarians**. 5 th. Califórnia: Libraries Unlimited, 2010.

COPERNICUS PUBLICATION. **Data policy**. [Alemanha], [2024]. Disponível em: https://www.earth-system-science-data.net/policies/data_policy.html. Acesso em: 8 jan. 2024.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Sobre a avaliação**. Brasília, 2021. Disponível em: <https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/sobre-aavaliacao/avaliacao-o-que-e/sobre-a-avaliacao-conceitos-processos-e-normas/conceito-avaliacao>. Acesso em: 23 set. 2023.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Relatório de avaliação**: Comunicação e Informação: Avaliação Quadrienal 2021. Brasília, 2022. Disponível em: https://www.gov.br/capes/ptbr/centrais-deconteudo/documentos/avaliacao/19122022_RELATORIO_AVALIACAO_QUADRIENAL_comnotaComunicacao_Informao.pdf. Acesso em: 21 set. 2023.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Documento técnico do Qualis periódicos**. Brasília, 2023. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/avaliacao-quadrienal-2017/DocumentotcnicoQualisPeridicosfinal.pdf>. Acesso em: 19 jun. 2023.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Documento referencial**: diretrizes comuns da avaliação de permanência dos programas de pós-graduação stricto sensu. Ciclo avaliativo 2025-2028. Avaliação Quadrienal 2029. Brasília, 2025. Disponível em: https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19052025_20250502_DocumentoReferencial_FICHA.pdf. Acesso em: 31 out. 2025.

CUNHA, J. S.; VEIGA, V. Atuação de agências de fomento brasileiras e a percepção de pesquisadores quanto ao compartilhamento de dados de pesquisa sobre Covid-19: um estudo exploratório. **BiblioCanto**, Natal, v. 9, n. 2, p. 123–132, 2023. Disponível em: <https://periodicos.ufrn.br/bibliocanto/article/view/33808>. Acesso em: 28 jul. 2025.

CUNHA, M. B. **Manual de fontes de informação**. Brasília: Briquet de Lemos/Livros, 2010.

CUNHA, M.B.; CAVALCANTI, C. R. O. Comunicação científica. *In*: CUNHA, M.B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos/ Livros, 2008a. p. 97

CUNHA, M.B.; CAVALCANTI, C. R. O. Comunidade científica. *In*: CUNHA, M.B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos/ Livros, 2008b. p. 98

CURTY, R.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA

DA INFORMAÇÃO - ENANCIB, 18., 2017, Marília, **Anais** [...]. Marília, SP, 2017. Disponível em: <https://hal.archives-ouvertes.fr/hal-01637987>. Acesso em: 1 jul. 2023.

DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG). DFG supports plans for legislation on research data. **Press Release** n. 10, 11 abril, 2023. Disponível em: <https://www.dfg.de/en/service/press/press-releases/2023/press-release-no-10>. Acesso em: 28 jul. 2025.

DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG). **What is the DFG?** Última atualização, July 3, 2025. Disponível em: <https://www.dfg.de/en/about-us/about-the-dfg/what-is-the-dfg>. Acesso em: 28 jul. 2025.

DIAS, G. A. *et al.* Publicando dados de pesquisa: Contextualizando as principais etapas e elementos envolvidos no processo. **Ciência da Informação**, Brasília, v. 49, n. 3, p. 115–127, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5501>. Acesso em: 7 jun. 2023.

DIAS, C. G. S. Panorama das políticas federais de ciência aberta no Brasil no período de 2020 a 2023. **RICI: Revista Ibero-americana da Ciência da Informação**, Brasília, v. 18, n. 1, p.131-153, jan./ abril 2025. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/55276/42262>. Acesso em: 17 jun. 2025.

DIMITROVA, M. *et al.* A streamlined workflow for conversion, peer review, and publication of genomics metadata as omics *data papers*. **GigaScience**, v. 10, n. 5, p. 1-13, 2021. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33983435/>. Acesso em: 14 nov. 2025.

DOAJ. **About DOAJ**. [S. l.], c2024. Disponível em: <https://doaj.org/about/>. Acesso em: 10 out. 2024.

ELSEVIER. **Scopus**. [S. l.], c2024a. Disponível em: <https://www-scopus.ez46.periodicos.capes.gov.br/search/form.uri?display=basic#basic>. Acesso em: 25 maio 2024.

ELSEVIER. **Guide for authors**. [S. l.], c2024b. Disponível em: <https://www.sciencedirect.com/journal/chemical-data-collections/publish/guide-for-authors>. Acesso em: 8 jan. 2024.

ELSEVIER. **About the journal**. [S. l.], c2024c. Disponível em: <https://www.sciencedirect.com/journal/data-in-brief>. Acesso em: 9 out. 2024.

EL-TAWIL, S.; AGRAWAL, A. K. *Data papers*: a new submission category. **Journal of Structural Engineering (United States)**, 145, n. 12, p.1-2, 2019. Disponível em: <https://ascelibrary.org/doi/10.1061/%28ASCE%29ST.1943-541X.0002574>. Acesso em: 24 jul. 2023.

ENWALD, H. *et al.* Data sharing practices in open access mode: a study of the willingness to share data in different disciplines. **Information Research**, v. 27, n .2,

2022. Disponível em: <http://InformationR.net/ir/27-2/paper932.html>. Acesso em: 24 abr. 2023.

EUROPEAN COMMISSION. **Horizon Europe, Open Science**: early knowledge and data sharing, and open collaboration. Publications Office of the European Union, 2021. Disponível em: <https://data.europa.eu/doi/10.2777/18252>. Acesso em: 21 ago. 2024.

EUROPEAN SCIENCE FOUNDATION. **Open access is foundational to the scientific enterprise**. c2024. Disponível em: <https://www.coalition-s.org/why-plan-s/>. Acesso em: 21 ago. 2024.

FAIR OPEN ACCESS ALLIANCE. **The Fair Open Access Principles**. [S.l.], 2021. Disponível em: <https://www.fairopenaccess.org/the-fair-open-access-principles/>. Acesso em: 28 out. 2025.

FARINA, A.; MARONGIU, P.; RODDA, M. A. Editorial: Representing the Ancient World through Data. **Journal of Open Humanities Data**, v.10, n. 57, p. 1-6, 2024. Disponível: <https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.245>. Acesso em: 23 nov. 2025.

FARINA, A. *et al.* When data meets the past: data collection, sharing, and reuse in ancient world studies. **Open Information Science**, v. 9, 20250014, 2025. Disponível em: <https://www.degruyterbrill.com/document/doi/10.1515/opis-2025-0014/html>. Acesso em: 19 nov. 2025.

FEDERER, L. M. *et al.* Data sharing in Plos One: an analysis of Data Availability Statements. **Plos One**, v. 13, n. 5, p. 1-12, 2018. Disponível em: <https://doi.org/10.1371/journal.pone.0194768>. Acesso em: 8 abr. 2025.

FEDERER, L. M. Long-term availability of data associated with articles in Plos One. **Plos One**, v. 17, n. 8, p. e0272845, 2022. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9401135/>. Acesso em: 30 abr. 2025.

FERNANDES, D. R. R.; PITTA, R. M. When the small is beautiful: five years publishing short communications. **Entomological Communications**, [S. l.], v. 6, p. ec06001, 2024. Disponível em: <https://www.entomologicalcommunications.org/index.php/entcom/article/view/ec06001>. Acesso em: 31 jul. 2025.

FERREIRA, L.S.; FERRO, R.R. **Técnicas de pesquisa bibliográfica e de elaboração de monografia**. São Luís, ABPEM, 1983.

FORTALEZA, J. M.; BERTIN, P. R. B.; DRUCKER, D. P. O compromisso pela ciência aberta na Open Government Partnership: avanços na governança de dados científicos no Brasil. In: PRÍNCIPE, E.; RODE, S. M. (org.). **Comunicação científica aberta**. Rio de Janeiro: Ibict, 2022. p. 189-206. Disponível em: <https://www.abecbrasil.org.br/painel/uploads/www/geral/E-book-Comunicacao-cientifica-aberta.pdf>. Acesso em: 15 dez. 2023.

FUNDAÇÃO OSWALDO CRUZ (Fiocruz). **Arca dados**. Sobre. Rio de Janeiro, [2023?]. Disponível em: <https://arcadados.fiocruz.br/www/sobre.html>. Acesso em: 3 maio 2023.

FUNDAÇÃO OSWALDO CRUZ (Fiocruz). **Arca Dados se torna a primeira plataforma no Brasil a receber selo CoreTrustSeal**. Rio de Janeiro, 12 de setembro de 2024. Disponível em: <https://fiocruz.br/noticia/2024/09/arca-dados-se-torna-primeira-plataforma-no-brasil-receber-selo-coretrustseal>. Acesso em: 4 ago. 2025.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Open Science @ FAPESP**. São Paulo, [2023?]. Disponível em: <https://fapesp.br/openscience/>. Acesso em: 28 fev. 2023.

GAMA, I. O.; CIANCONI, R. B.; GÓMEZ, M. N. G. A abertura científica: o processo de resignificação a partir dos movimentos open access e open science. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.27, n. 4, p. 28 -53, out./dez. 2022. Disponível em: <https://www.scielo.br/j/pci/a/4fqh8qH6WLGf9B6w75kSZDd/>. Acesso em: 3 ago. 2023.

GARCÍA-GARCÍA, A.; LOPEZ-BORRULL, A.; PESET, F. Data journals: emergence of new journals specializing in data. **Profesional de la informacion**, v. 24, n. 6, p. 845-854, 2015. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2015.nov.17>. Acesso em: 24 jul. 2023.

GIBBON, C. *et al.* Ciência aberta no Brasil e as ações das fundações de amparo à pesquisa. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 24., 2024, Vitória. **Anais [...]**. Vitória, 2024. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/viewFile/2276/1782>. Acesso em: 28 jul. 2025.

GINER-MIGUELEZ, J.; GÓMEZ, A.; CABOT, J. On the readiness of scientific data papers for a FAIR and transparent use in machine learning. **Scientific Data**, v. 12, n. 61, p. 1-16, 2025. Disponível em: <https://www.nature.com/articles/s41597-025-04402-4>. Acesso em: 6 jun. 2025.

GINGRAS, Y. The transformation of the scientific paper: from knowledge to accounting unit. *In: BIAGIOLI, M.; LIPPMAN, A.. Gaming the metrics: misconduct and manipulation in academic research*. Cambridge: MIT Press, 2020. p. 43-55. Disponível em: <https://direct.mit.edu/books/book/4598/chapter/211130/The-Transformation-of-the-Scientific-PaperFrom>. Acesso em: 6 jun. 2025.

GONZÁLEZ-PEREIRA, B.; GUERRERO-BOTE, V. P.; MOYA-ANEGÓN, F. A new approach to the metric of journals' scientific prestige: The SJR indicator. **Journal of Informetrics**, v. 4, n. 3, p. 379–391, 2010. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157710000246?via%3Dihub>. Acesso em: 18 set. 2024.

GONZÁLEZ-TERUEL, A. Drivers and barriers in the transition to open science: the perspective of stakeholders in the Spanish scientific community. **Profesional de la información**, v. 31, n. 3, e310305, 2022. Disponível em:

<https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86857>.

Acesso em: 13 jul. 2023.

GORGOLEWSKI, K. J.; MARGULIES, D. S.; MILHAM, M. P. Making data sharing count: a publication-based solution. **Frontiers in Neuroscience**, v. 7, p. 1-7, 2013.

Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3565154/>. Acesso em: 24 jul. 2023.

GRAF, C. *et al.* The open data challenge: an analysis of 124,000 data availability statements and an ironic lesson about data management plans. **Data Intelligence**, v.2, n. 4, p. 554-568, 2020. Disponível em: https://doi.org/10.1162/dint_a_00061.

Acesso em: 28 ago. 2025.

GRAY, D. E. **Pesquisa no mundo real**. 2.ed. Porto Alegre: Penso, 2012.

GREGORY, K. M. *et al.* Understanding data search as a socio-technical practice. **Journal of Information Science**, v. 46, n. 4, p. 459-475, 2020. Disponível em:

<https://journals.sagepub.com/doi/10.1177/0165551519837182>. Acesso em: 20 dez. 2023.

GUÉDON, J-C. Oldenburg's long shadow: librarians, research scientists, publishers and the control of scientific publishing. *In*: CREATING THE DIGITAL FUTURE: ASSOCIATION OF RESEARCH LIBRARIES, 138., Toronto, Ontario (Canada), may 23-25, 2001. **Proceedings Annual Meeting**. Toronto, 2001. Disponível em:

<http://eprints.rclis.org/6375/>. Acesso em: 20 ago. 2023.

GUÉDON, J-C. El acceso abierto y la división entre ciencia “principal” y “periférica”. **Crítica y Emancipación**, [S. l.], ano 3, n. 6, p. 135-180, 2011. Disponível em:

https://bibliotecavirtual.clacso.org.ar/ar/libros/secret/CyE/6/CyE-6_Guedon-CLACSO.pdf. Acesso em: 12 nov. 2025.

GUÉDON, J-C *et al.* **Future of scholarly publishing and scholarly communication: report of the expert group to the european commission**.

Luxembourg: Publications Office of the European Union, 2019. 57 p. Disponível em:

https://www.eosc-portal.eu/sites/default/files/KI0518070ENN.en_.pdf. Acesso em: 20 dez. 2023.

GUIFENG, L. *et al.* Analysis of multi-level data reuse topic differences based on highly cited individual data paper. **Information studies: theory & application**, v. 47, n. 5, 2024.

GUIFENG, L.; CHEN, S. HAN, M. Research on open sharing and association model of data journals based on FAIR principle. **Journal of Modern Information**, v. 44, n. 2, p.30-42, 2024. Disponível em:

<https://www.sciengine.com/JMI/articleIndex/crossref?doi=10.3969/j.issn.1008-0821.2024.02.003>. Acesso em: 19 nov. 2025.

HE, L.; NAHAR, V. Reuse of scientific data in academic publications: an investigation of Dryad Digital Repository. **Aslib Journal of Information Management**, v. 68, n. 4, p. 1-23, 2016. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/AJIM-01-2016-0008/full/html>. Acesso em: 10 jan. 2024.

HENG Y.; YONGLIN, Y; FENGHONG, F. How do data authors perform in data-intensive research activities? Evidence from author contribution statement in data papers. **Information Research**, v. 30, n. iConf, p. 219-239, 2025. Disponível em: <https://publicera.kb.se/ir/article/view/47524>. Acesso em: 30 jun. 2025

HERNÁNDEZ SAMPIERI, R.; FERNÁNDEZ COLLADO C., BAPTISTA LUCIO. P. **Metodologia de pesquisa**. 5.ed. São Paulo: McGraw-Hill, 2013.

HRYNASZKIEWICZ, I., *et al.* Publishing descriptions of non-public clinical datasets: proposed guidance for researchers, repositories, editors and funding organisations. **Research Integrity and Peer Review**, v. 1, n. 6, p. 1-11, 2016. Disponível em: <https://doi.org/10.1186/s41073-016-0015-6>. Acesso em: 7 out. 2025.

HUANG, P. P.; JENG, W. Data Paper as a Reward? Motivation, consideration, and perspective behind data paper submission. *In*: PROCEEDINGS OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 85., Pittsburgh, 2022. **Proceedings** [...]. Pittsburgh, 2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.648>. Acesso em: 24 jul. 2023.

HUANG, P. P.; JENG, W. Data Paper's Functions in Scholarly Communication Ecosystem as Perceived by Natural Scientists. *In*: ANNUAL MEETING OF THE ASSOCIATION FOR INFORMATION SCIENCE & TECHNOLOGY, 86., Londres, 2023. **Proceedings** [...]. Londres, 2023. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pra2.920>. Acesso em: 2 out. 2025.

HUANG, X.; HAWKINS, B. A.; QIAO, G. Biodiversity Data Sharing: will peer-reviewed *data papers* work? **Bioscience**, v. 63, n. 1, p. 5-6, 2013. Disponível em: <https://academic.oup.com/bioscience/article/63/1/5/241043>. Acesso em: 24 jul. 2023.

HUANG, X. Analysis on the Sustainable Development of Data Journals Based on Article Processing Charge Policy. **Journal of Library and Information Sciences in Agriculture**, Beijing, v. 30, n. 7, p. 133-136, 2018. Disponível em: <https://www.proquest.com/docview/2861359883?fromopenview=true&pq-origsite=gscholar&sourcetype=Scholarly%20Journals>. Acesso em: 11 nov. 2025.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Ibict, CNPq e RNP lançam edital para a criação de repositórios de dados de pesquisa**. Brasília, 2021. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2021/julho2021/ibict-cnpq-e-rnp-lancam-edital-para-a-criacao-de-repositorios-de-dados-de-pesquisa>. Acesso em: 28 fev. 2023.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Repositório Aleia conquista certificação internacional CoreTrustSeal, que reconhece repositórios digitais confiáveis**. Brasília, 06 de junho de 2025. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2025/junho/repositorio-aleia-conquista-certificacao-internacional-coretrustseal-que-reconhece-repositorios-digitais-confiaveis>. Acesso em: 4 ago. 2025.

INTERNATIONAL ASSOCIATION OF UNIVERSITIES (IAU). **Open Science: the challenge for universities**. Paris: Unesco, 2024. Disponível em: https://www.iau-aiu.net/IMG/pdf/open_science_the_challenge_for_universities.pdf.pdf. Acesso em: 22 jul. 2025.

JIAO, C; DARCH, P. T. The role of the data paper in scholarly communication. *In: PROCEEDINGS OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*, 83., 2020. **Proceedings** [...]. [S. l.], 2020. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.316>. Acesso em: 19 jun. 2023.

JIAO, C; DARCH, P. T. Peer review of data papers: does it achieve expectations for facilitating data sharing and reuse? **Journal of Information Science**, 2025. Disponível em: <https://doi.org/10.1177/01655515251379048>. Acesso em: 31 out. 2025.

JIAO, C.; LI, K.; FANG, Z. Data sharing practices across knowledge domains: a dynamic examination of data availability statements in Plos One publications. **Journal of Information Science**, v. 50, n. 3, p. 673-689, 2022. Disponível em: <https://doi-org.ez46.periodicos.capes.gov.br/10.1177/01655515221101830>. Acesso em: 14 abr. 2025.

JIAO, C.; LI, K.; FANG, Z. How are exclusively data journals indexed in major scholarly databases? An examination of four databases. **Scientific Data**, v.10, n. 737, p. 1-9, 2023. Disponível em: <https://www.nature.com/articles/s41597-023-02625-x>. Acesso em: 05 jun. 2025.

JIAO, C. *et al.* Dissemination effect of data papers on scientific datasets. **Journal of the Association for Information Science and Technology**, p. 1-17, 2023. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24843>. Acesso em: 19 nov. 2025.

KAR, S.; RATH, D. S. Open data in Social Sciences: growth, impact, and equity in data paper publishing. **DESIDOC Journal of Library & Information Technology**, v. 45, n. 4, p. 350-366, 2025. Disponível em: <https://publicationsdrdo.in/index.php/djlit/article/view/21026>. Acesso em: 24 out. 2025.

KAUSHIK, V.; WALSH, C.A. Pragmatism as a research paradigm and its implications for social work research. **Social. Science**, v. 8, 255, p. 1-17, 2019. Disponível em: <https://www.mdpi.com/2076-0760/8/9/255>. Acesso em: 17 nov. 2025.

KEMBELLE, G.; LE DEUFF, O. **Poétique et ingénierie des data papers**. Revue française des sciences de l'information et de la communication, [S.l.], 2022. Disponível em: <https://journals.openedition.org/rfsic/12938>. Acesso em: 7 nov. 2025.

KIM, J. An analysis of data paper templates and guidelines: types of contextual information described by data journals. **Science Editing**, v. 7, n. 1, p. 16-23, 2020. Disponível em: <http://escienceediting.org/journal/view.php?doi=10.6087/kcse.185>. Acesso em: 24 jul. 2023.

KRATZ, J.; STRASSER, C. Data publication consensus and controversies. **F1000Research**, v. 3, n. 94, p. 1-21, 2014. Disponível em: <https://doi.org/10.12688/f1000research.3979.3>. Acesso em: 1 jul. 2023.

KONG, L. *et al.* A data quality evaluation index for data journals. In: LI, J. *et al.* (eds) **Big Scientific Data Management**. Switzerland: Springer, 2019. p. 291-300.

KOSMOPOULOS, C; SCHÖPFEL, J. Introduction. In: KOSMOPOULOS, C; SCHÖPFEL, J. **Publier, partager, réutiliser les données de la recherche: les data papers et leurs enjeux**. Lille: Presses universitaires du Septentrion, 2025. p. 5-12. Disponível em: https://www.septentrion.com/html/WYSIWYGfiles/files/HNSO2_Publier-Partager-R%C3%A9utiliser.pdf. Acesso em: 18 set. 2025.

KOTTI, Z. *et al.* Standing on shoulders or feet? An extended study on the usage of the MSR data papers. **Empirical Software Engineering**, v.25, p. 3288–3322, 2020. Disponível em: <https://link.springer.com/article/10.1007/s10664-020-09834-7>. Acesso em: 24 jul. 2023.

LE DEUFF, O. **Une nouvelle rubrique pour la RFSIC: Le Data Paper**. Revue française des sciences de l'information et de la communication, [S.l.], 2018. Disponível em: <https://journals.openedition.org/rfsic/5275>. Acesso em: 7 nov. 2025.

LE FOURNER, V.; SCHÖPFEL, J. Le paysage des data papers. In: KOSMOPOULOS, C; SCHÖPFEL, J. **Publier, partager, réutiliser les données de la recherche: les data papers et leurs enjeux**. Lille: Presses universitaires du Septentrion, 2025. p. 2-17, parte 1. Disponível em: https://www.septentrion.com/html/WYSIWYGfiles/files/HNSO2_Publier-Partager-R%C3%A9utiliser.pdf. Acesso em: 18 set. 2025.

LEE, J.; KIM, J. Korean researchers' motivations for publishing in data journals and the usefulness of their data: a qualitative study. **Science Editing**, v. 8, n. 2, p. 145-152, 2021. Disponível em: <https://www.escienceediting.org/upload/kcse-246.pdf>. Acesso em: 24 jul. 2023.

LEE, S., KIM, J. Factors influencing authors' intention to continue publishing in data journals: a cross-sectional survey. **Science Editing**, v.12, 2, p. 183-189, 2025. Disponível em: <https://escienceediting.org/journal/view.php?doi=10.6087/kcse.383>. Acesso em: 5 nov. 2025.

LERMA GONZÁLEZ, H. D. **Metodología de la investigación**: propuesta, anteproyecto y proyecto. 4. ed. Bogotá: Ecoe Ediciones, 2009.

LI, K.; CHEN, P.-Y. The narrative structure as a citation context in *data papers*: A preliminary analysis of Scientific Data. In: PROCEEDINGS OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 81., Vancouver, 2018.

Proceedings [...]. Vancouver, 2018. Disponível em:

<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pras2.2018.14505501147?af=R>.

Acesso em: 24 jul. 2023.

LI, K.; GREENBERG, J.; DUNIC, J. Data objects and documenting scientific processes: an analysis of data events in biodiversity *data papers*. **Journal of the Association for Information Science and Technology**, v. 71, p. 172-182, 2020.

Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24226>.

Acesso em: 21 jun. 2023.

LI, K.; HUANG, P.; JENG, W. Are data papers cited as research data? Preliminary analysis on interdisciplinary data paper citations. **Information Research**, v. 30, n. iConf, p. 1225-1233, 2025. Disponível em: <https://doi.org/10.47989/ir30iConf46918>.

Acesso em: 8 abr. 2025.

LI, K.; LU, C.; JIAO, C. A survey of exclusively data journals and how they are indexed by scientific databases. In: PROCEEDINGS OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 84., Salt Lake City, 2021.

Proceedings [...]. Salt Lake City, 2021. Disponível em:

<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pras2.557>. Acesso em: 23 nov.

2025.

LI, K.; JIAO, C. How are data paper abstracts constructed? Preliminary analysis of rhetorical moves in data paper abstracts from Scientific Data and Data in Brief. In: INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS, 18., 2021. **Proceedings** [...], 2021. Disponível em:

<https://chenyuejiao.github.io/paper/ISSI.pdf>. Acesso em: 21 jun. 2023.

LI, K.; JIAO, C. The data paper as a sociolinguistic epistemic object: A content analysis on the rhetorical moves used in data paper abstracts. **Journal of the Association for Information Science and Technology**, v. 73, n. 6, p. 834-846,

2022. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24585>.

Acesso em: 21 jun. 2023.

LICHTFOUSE, E. Publishing science without results and recycling research.

Environmental Chemistry Letters, v. 16, n. 1, p. 1-4, 2018. Disponível em:

<https://link.springer.com/article/10.1007/s10311-017-0690-7>. Acesso em: 24 jul.

2023.

LIMA, T.C.S.; MIOTO, R.C.T. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. **Revista Katálysis**, Florianópolis, v.10, n. especial, p. 37-45, 2007. Disponível em:

<https://www.scielo.br/rk/a/HSF5Ns7dkTNjQVpRyvvhc8RR/?lang=pt#>. Acesso em: 18

jul. 2023.

LI-SCHOLZ, A. Data journal articles in the electronic publishing environment. **Atomic Data and Nuclear Data Tables**, v. 79, n. 2, p. 185- 186, 2001.

LIU, X. Discussion on the structural model and constituent elements of data papers. **Resources Data Journal**, v. 1, p. 2-9, 2022. Disponível em: https://www.jstage.jst.go.jp/article/rdj/1/0/1_2/pdf. Acesso em: 24 jul. 2023

MACHUCA-MARTÍNEZ, F. Importance of scientific data and its publication as data paper. **Ingeniería y Competitividad**, v. 22, n. 1, p. 1-3, 2020. Disponível em: <https://www.redalyc.org/journal/2913/291365764009/html/>. Acesso em: 24 jul. 2023.

MANCO, A. Landscape of open science policies research. **Sage Open**, v.12, n. 4, p. 1-17, 2022. Disponível: <https://journals-sagepub-com.ez46.periodicos.capes.gov.br/doi/10.1177/21582440221140358>. Acesso em: 21 jul. 2025.

MATERIALS CLOUD MODERATION POLICIES. v.1.3. [S./], 2020 Disponível em: https://archive.materialscloud.org/deposit/information#moderation_policies. Acesso em: 4 dez. 2024.

MAYERNIK, M. S., *et al.* Peer review of datasets: when, why, and how. **Bulletin of the American Meteorological Society**, v. 96, p. 191–201, 2015. Disponível em: <https://doi.org/10.1175/BAMS-D-13-00083.1>. Acesso em 3 jul. 2025.

MCGILLIVRAY, B. *et al.* Deep Impact: a study on the impact of *data papers* and datasets in the humanities and social sciences. **Publications**, v. 10, n. 4, 2022. Disponível em: <https://www.mdpi.com/2304-6775/10/4/39>. Acesso em: 28 jul. 2023.

MCGUINNESS, L. A.; SHEPPARD A. L. A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts. **Plos One**, v.16, n. 5, p.e0250887, p. 2021. Disponível em: <https://doi.org/10.1371/journal.pone.0250887>. Acesso em: 28 ago. 2025.

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

MORRISON, H. *et al.* Change and growth in open access journal publishing and charging trends 2011-2021. **Journal of the Association for Information Science and Technology**, v. 73, n. 12, pp. 1793-1805, 2022. Disponível em: <https://doi.org/10.1002/asi.24717>. Acesso em: 13 nov. 2025.

MUGNAINI, R.; IGAMI, M. P. Z.; KRZYZANOWSKI, R. F. Acesso aberto e financiamento da pesquisa no Brasil: características e tendências da produção científica. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 27, p. 1-26, 2022. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/78818>. Acesso em: 15 jul. 2023.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, maio/ago.

2006. Disponível em:

<https://www.scielo.br/j/ci/a/nGD3MkKfNxtjnnWshf3YVjP/?format=pdf&lang=pt>.

Acesso em: 11 nov. 2025.

MUELLER, S. P. M. Literatura científica, comunicação científica. *In*: Toutain, L.M. B. B. (org.). **Para entender a ciência da informação**. Salvador: EDUFBA, 2007. pp. 125-144

NEUBERT, P. *et al.* Custo de apc em periódicos Qualis: análise por estrato e área de avaliação. *In*: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 9., 2024, Brasília. **Anais [...]**. Unb, Ibict, 2024. Disponível em: <https://cip.brapci.inf.br/download/306018>. Acesso em: 4 out. 2024.

NEWMAN, P; CORKE P. Editorial: Data Papers - Peer Reviewed Publication of High Quality Data Sets: **The International Journal of Robotics Research**, v. 28, n.5, 2009. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0278364909104283>. Acesso em: 23 nov. 2025.

OPEN REFINE'S. **OpenRefine**. [S./], [2023]. Disponível em: <https://openrefine.org/>. Acesso em: 28 dez. 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Principles and guidelines for access to research data from public funding**. France: OECD publications, 2007. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 28 fev. 2023.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO). **Recomendação da Unesco sobre Ciência Aberta**. Paris, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 9 jul. 2025.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **British Medical Journal**, v. 372, n. 71, 2021. Disponível em: <https://www.bmj.com/content/372/bmj.n71>. Acesso em: 28 jul. 2023.

PAMPEL, H.; DALLMEIER-TIESSEN, S. Open research data: from vision to practice. *In*: BARTLING, S., FRIESIKE, S. (ed.). **Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. London, SpringerLink, 2014. p. 213-224. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-00026-8_14. Acesso em: 7 dez. 2023.

PANDO, D. A.; ALMEIDA, C. C. Análise sobre a epistemologia e sua aplicação à ciência da informação. **Informação & Informação**, Londrina, v. 26, n. 2, p. 680-705, 2021. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41635/pdf_1. Acesso em: 29 ago. 2023.

PENEV, L. *et al.* Strategies and guidelines for scholarly publishing of biodiversity data. **Research Ideas and Outcomes**, v. 3, e12431, 2017. Disponível em: <https://riojournal.com/article/12431/list/9/>. Acesso em: 13 jul. 2023.

PENSOFT PUBLISHERS. **About**. Sofia, c2024. Disponível em: <https://bdj.pensoft.net/about>. Acesso em: 9 out. 2024.

PÉREZ-LUQUE, A. J.; ROS-CANDEIRA, A. Sharing data in Ecology: how to add more value to data. **Ecosistemas**, v. 28, n. 3, p. 150-159, 2019.

PINTO, F. M. A. G. **Publicação de dados de pesquisa em Enfermagem: novas práticas em Ciência Aberta**. 2022. Tese. (Doutorado em Ciência da Informação). – Escola de Comunicação e o Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2022. Disponível em: <https://ridi.ibict.br/handle/123456789/1354>. Acesso em: 16 jun. 2025.

PINTO, F. M. A. G. Publicação de dados de pesquisa: panorama terminológico-conceitual sobre a prática no âmbito da comunicação científica. *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA (WIDaT), 6., 2023a, Brasília. **Anais [...]**. Brasília, 2023a. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/80>. Acesso em: 27 set. 2023.

PINTO, F. M. A. G. Publicação de dados de pesquisa em enfermagem: novas práticas em Ciência Aberta. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis/SC, Brasil, v. 28, p. 1-11, 2023b. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/93296>. Acesso em: 19 nov. 2025.

PINTO, F. M. A. G.; PINHEIRO, L. V. R. O Periódico de dados e as transformações na comunicação científica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis, 2025. Disponível em: <https://www.brapci.inf.br/v/123199>. Acesso em: 19 nov. 2025.

PIZZANI, L. *et al.*, A arte da pesquisa bibliográfica na busca do conhecimento. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 10, n. 1, p. 53-66, 2012. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1896>. Acesso em: 17 jul. 2023.

PIWOWAR, H. *et al.* The state of OA: a large-scale analysis of the prevalence and impact of open access articles. **PeerJ**, n. 6, e4375, 2018. Disponível em: <https://peerj.com/articles/4375/#>. Acesso em: 3 set. 2024.

PORTENOY, J. Topics. *In*: PRIEM, J., PIWOWAR, H., & ORR, R. **OpenAlex**: a fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv, 2022. Atualizado em 24 July 2024. Disponível em: <https://help.openalex.org/hc/en-us/articles/24736129405719-Topics>. Acesso em: 23 ago. 2024.

PRIEM, J.; PIWOWAR, H.; ORR, R. OpenAlex: A fully-open index of scholarly

works, authors, venues, institutions, and concepts. **ArXiv**, 2022. Disponível em: <https://arxiv.org/abs/2205.01833>. Acesso em: 27 maio 2024.

PRITYCHENKO, B. New editor discusses plans to maintain and improve ADNDT as the leading atomic and nuclear data journal. **Atomic Data and Nuclear Data Tables**, v. 100, p. 833, 2014. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0092640X14000084>. Acesso em: 23 nov. 2025.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

PUERTA-PIÑERO, C.; PÉREZ-LUQUE, A. J.; RODRÍGUEZ-ECHEVERRÍA, S. Ecosistemas is committed to the publication of *data papers*. **Ecosistemas**, v. 29, n. 3, p.1-2, 2020. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099622212&doi=10.7818%2FECOS.2118&partnerID=40&md5=fe10a336677df22873eebbe41df5951f>. Acesso em: 24 jul. 2023.

R3DATA COREF. **Reviewing the subject classification in re3data**. [S.l.: s.n.], 2021. Disponível em: <https://coref.project.re3data.org/blog/reviewing-the-subject-classification-in-re3data>. Acesso em: 13 ago. 2025.

RAMÍREZ, P. A.; SAMOILOVICH, D. **Ciencia Abierta en América Latina**. Montevideo: Unesco, 2021. Disponível em: <http://forocilac.org/wp-content/uploads/2022/03/PolicyPapers-CienciaAbierta-ES-v2.pdf>. Acesso em: 8 dez. 2023.

RAUB, F. *et al.* Data paper publication in Arachnologische Mitteilungen – Goals, review criteria, editorial procedures, format, data management and mobilization. **Arachnologische Mitteilungen**: Arachnology Letters, v.66, n.1, p. 1-4, 2023. Disponível em: <https://doi.org/10.30963/aramit6612>. Acesso em: 23 nov. 2025.

RIBEIRO, N. C.; OLIVEIRA, D. A.; SANTOS, S. R. O. Ações e estratégias voltadas para a ciência aberta em universidades estaduais paulistas: um estudo multicaso. **Em Questão**, Porto Alegre, v. 27, n. 2, p. 164-192, 2021. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/102078>. Acesso em: 23 jul. 2025.

RICHARDSON, J.; WOLSKI, M. The peer review process: an expanded role for repositories. **Journal of Systemics, Cybernetics and Informatics**, v. 12, n.1, p. 58-62, 2014. Disponível em: <https://www.iiisci.org/journal/sci/FullText.asp?var=&id=GA195DM13>. Acesso em: 30 maio 2025.

ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B.; SANTANA, R. C. Proposed structure of a data paper structure as scientific publication. **Revista Espanola de Documentacion Cientifica**, v. 40, n. 1, p.1-13, 2017. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/971>. Acesso em: 24 jul. 2023.

RESEARCH ORGANIZATION REGISTRY (ROR). **What is ROR?** [S.l.: s.n.], [2025?]. Disponível em: <https://ror.org/about/>. Acesso em: 27 maio 2024.

ROVELLI, L.; GONZÁLEZ, A. L. Lineamientos para la Ciencia Abierta e incentivos desde la evaluación de la investigación en Iberoamérica. In: BATTHYÁNY, K.; VOMMARO, P.; ROVELLI, L. (ed.). **Iniciativas y regulaciones multinivel para la Ciencia Abierta: Infraestructuras abiertas y sistemas de evaluación** em Iberoamérica. Madrid: Fundación Carolina, 2023. p. 7-19. Disponível em: <https://www.fundacioncarolina.es/iniciativas-y-regulaciones-multinivel-para-la-ciencia-abierta-infraestructuras-abiertas-y-sistemas-de-evaluacion-en-iberoamerica/>. Acesso em: 1 dez. 2023.

RUSHBY, N. Editorial: *Data papers*. **British Journal of Educational Technology**, v. 46, n. 5, p. 899-903, 2015. Disponível em: <https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.12337>. Acesso em: 24 jul. 2023.

SANTOS, A. C. G. Ciência Aberta nas instituições de ensino: abordagens e estratégias. **Revista Brasileira de Biblioteconomia e Documentação**, v. 20, p. 1–17, 2024. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/2042>. Acesso em: 31 jul. 2025.

SANTOS-D'AMORIM, K. A comunicação científica em movimento: das origens aos debates atuais. **Brazilian Journal of Information Science: Research trends**, v. 15, publicação continua, 2021, e02103. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/11468/7041>. Acesso em: 12 nov. 2025.

SANTOS, S. M.; NASSI-CALÓ, L. Gestão editorial: tendências e desafios na transição para a ciência aberta. In: SILVEIRA, L.; SILVA, F. C. C. (Orgs.). **Gestão editorial de periódicos científicos: tendências e boas práticas**. Florianópolis: BU Publicações; Edições do Bosque/UFSC, 2020. p. 17-55. Disponível em: <https://repositorio.ufsc.br/handle/123456789/208680>. Acesso em: 07 dez. 2023.

SAYÃO, L. F.; SALES, L. F. Periódicos de resultados negativos: revelando uma parte invisível da ciência. In: SHITAKU, M.; SALES, L. (org.). **Ciência aberta para editores científicos**. Botucatu, SP: ABEC, 2019. p. 97-102.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa?. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 34, n. 2, p. 32-51, jul./dez. 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 10 jan. 2023.

SAYÃO, L. F.; SALES, L. F. Plataformas de gestão de dados de pesquisa: expandindo o conceito de repositórios de dados. **Palavra chave**, v. 12, n. 1, e171, 2022. Disponível em: https://www.scielo.org/ar/scielo.php?script=sci_arttext&pid=S1853-99122022000200171&lng=es&nrm=iso. Acesso em: 14 nov. 2025.

SCARTASSINI, V. B.; MOURA, A. M. M. O financiamento público de pesquisas brasileiras: uma análise da produção científica indexada na Web of Science. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, Brasil, v. 11, n. 1, p. 33–51, 2020. Disponível em: <https://revistas.usp.br/incid/article/view/153894>. Acesso em: 29 jul. 2025.

SCHIESSL, I. T.; SHINTAKU, M. Artigos de dados: conceitos e tendências no cenário editorial. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 13, n. 2, p. 88-106, 2022. Disponível em: <https://www.revistas.usp.br/incid/article/view/194536>. Acesso em: 8 jun. 2023.

SCHNEIDER, A.; HAGENER, M. Data Papers: an introduction. **Necsus: European Journal of Media Studies**, n. 12, p. 359-362, 2023. Disponível em: <https://doi.org/10.25969/mediarep/19751>. Acesso em: 23 nov. 2025.

SCHÖPFEL, J. *et al.*. *Data papers* as a new form of knowledge organization in the field of research data. In: COLLOQUE INTERNATIONAL D'ISKO-FRANCE: Données et mégadonnées ouvertes en SHS: de nouveaux enjeux pour l'état et l'organisation des connaissances? 12., 2019, Montpellier. **Proceedings** [...]. Montpellier, France, 2019. Disponível em: <https://shs.hal.science/halshs-02284548/document>. Acesso em: 22 jun. 2023.

SCIMAGO. **About us**. [S.l.], c2017-2025. Disponível em: <https://www.scimagojr.com/aboutus.php>. Acesso em: 28 dez. 2025.

SEO, S.; KIM, J. Data journals: types of peer review, review criteria, and editorial committee members' positions. **Science Editing**, v.7, n.2, p. 130-135, 2020. Disponível em: <https://www.escienceediting.org/journal/view.php?number=219>. Acesso em: 10 jul. 2025.

SEVKUSIC, M. *et al.* Building National Open Science Cloud Initiatives (NOSCI) in Southeast Europe: supporting research and scholarly communication. **Publications**, v. 10, n. 42, p. 1-19, 2022. Disponível em: <https://www.mdpi.com/2304-6775/10/4/42>. Acesso em: 16 jul. 2025.

SHIN, N. *et al.* Toward more data publication of long-term ecological observations. **Ecological Research**, v. 35, n. 5, p. 700-707, 2020. Disponível em: <https://esj-journals.onlinelibrary.wiley.com/doi/10.1111/1440-1703.12115>. Acesso em: 10 jul. 2023.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. **Transinformação**, Campinas, v. 28, n. 1, p. 15–32, jan. 2016. Disponível em: <https://www.scielo.br/j/tinf/a/tvBDyptMBFSxRSt3VngySRC/>. Acesso em: 23 jul. 2025.

SILVA, F. C. C.; STUEBER, K.; WITT, A. S. *Data papers* e data journals sobre Biodiversidade: considerações acerca da preservação digital. **Revista Brasileira de Preservação Digital**, Campinas, SP, v. 3, e022012, p. 1-16, 2022. Disponível em:

<https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/16556>. Acesso em: 15 jun. 2023.

SILVA, F. C. C. *et al.* Contribuições brasileiras para a biodiversidade global: análise de data papers utilizando datasets do GBIF. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 24., 2024, Vitória, ES. **Anais [...]**. Vitória, 2024. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2796>. Acesso em: 20 ago. 2025.

SILVEIRA, C. Z.; DIAS, T. M. R.; NEUBERT, P. S. Compartilhamento de dados de pesquisa em periódicos científicos: uma análise dos data papers indexados na Scopus. *In: CONFERÊNCIA LUSÓFONA DE CIÊNCIA ABERTA*, 14., 2023a, Natal. **Anais [...]**. Natal, 2023. Disponível em: <https://zenodo.org/records/8399384>. Acesso em: 18 ago. 2025.

SILVEIRA, C. Z.; DIAS, T. M. R.; NEUBERT, P. S. Data papers publicados por pesquisadores brasileiros: uma análise da área da biodiversidade. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 23., 2023b, Aracaju. **Anais [...]**. Aracaju, 2023. Disponível em: <https://ancib.org/enancib/index.php/enancib/xxxiiienancib/paper/view/1464>. Acesso em: 13 ago. 2025.

SILVEIRA, C. Z.; DIAS, T. M. R.; NEUBERT, P. S. A publicação de dados de pesquisa: uma análise a partir dos data papers publicados por autores afiliados à Universidade Federal de Santa Catarina. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 24., 2024, Vitória, ES. **Anais [...]**. Vitória, 2024. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2410>. Acesso em: 13 ago. 2025.

SILVEIRA, C.Z.; NEUBERT, P. S; DIAS, T. M. R. Características dos dados abertos de pesquisa uma análise a partir dos data papers brasileiros sobre a COVID-19 indexados na Scopus. *In: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA*, 4., Brasília, 2023. **Anais [...]**. Brasília, 2023. v. 6. p. 1-12. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/issue/view/1>. Acesso em: 23 nov. 2025.

SILVEIRA, C.Z.; NEUBERT, P. S; DIAS, T. M. R. A publicação dos dados de pesquisa: data paper. *In: RONALDO, F. A., R. F; FACHIN, J. (org.). Ciência aberta, divulgação científica e marketing na ciência: questões teóricas e práticas*. Maceió: Edufal, 2025. cap. 2, p. 23-37.

SILVEIRA, L. *et al.* Taxonomia da Ciência Aberta: revisada e ampliada. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis/SC, Brasil, v. 28, p. 1-22, 2023. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/91712>. Acesso em: 17 nov. 2025.

SILVEIRA, L.; SILVA, F. C. C.; BARBOSA, A. D. Políticas editoriais de dados científicos em periódicos da área de comunicação e informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 16, n. 1, p. 286–306, 2023.

Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/42055>. Acesso em: 26 ago. 2025.

SOLIMAN, L. R.; TRIPIER-MONDANCIN, O.; BOURG, A. Contribution à la méthodologie du Data paper par la valorisation de données sur l'enseignement musical. **Journal de recherche en éducation musicale**, v. 15, p. 1-17, 2024. Disponível em: <http://journals.openedition.org/jrem/547>. Acesso em: 19 nov. 2025.

SOUZA, M. I. F., *et al.*. Catalogação de datasets ômicos no Repositório de Dados de Pesquisa da Embrapa. **Revista Brasileira de Biblioteconomia e Documentação**, v. 18, p. 1–28, 2022. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1706>. Acesso em: 24 ago. 2025.

STRECKER, D. *et al.* **Metadata Schema for the Description of Research Data Repositories**: version 4.0. [S.l.: s.n.], 2023. Disponível em: <https://doi.org/10.48440/re3.014>. Acesso em: 13 ago. 2025.

STUART, D. Data bibliometrics: metrics before norms. **Online Information Review**, v. 41, n. 3, p. 428-435, 2017. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/OIR-01-2017-0008/full/html>. Acesso em: 24 jul. 2023.

STUMPF, I. R. C. Pesquisa bibliográfica. *In*: DUARTE, J.; BARROS, A. **Métodos e técnicas de pesquisa em Comunicação**. São Paulo: Atlas, 2012.

TENOPIR, C. *et al.* Changes in data sharing and data reuse practices and perceptions among scientists worldwide. **Plos One**, v.10, n. 8, e0134826, 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>. Acesso em: 13 jul. 2023.

TENOPIR, C. *et al.* Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide. **Plos One**, v. 15, n.3, e0229003, 2020. Disponível em: <https://journals.plos.org/plosone/article%3Fid=10.1371/journal.pone.0229003>. Acesso em: 10 jan. 2023.

TERUI, T.; MINAMIYAMA, Y.; YAMAJI, K. Possibility and Prevention of Inappropriate Data Manipulation in Polar Data Journal. *In*: INTERNATIONAL CONGRESS ON ADVANCED APPLIED INFORMATICS, 8., Tayoma, 2019. **Proceedings** [...], Institute of Electrical and Electronics Engineers Inc., 2019. Disponível em: <https://ieeexplore.ieee.org/document/8992707>. Acesso em: 28 jul. 2023.

THELWALL, M. Data in Brief: Can a mega-journal for data be useful? **Scientometrics**, v. 124, n. 1, p. 697-709, 2020. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03437-1>. Acesso em: 28 jul. 2023.

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. 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:

https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 7 dez. 2023.

ÜNAL, Y.; AKBULUT, M. Who funds what: an assessment of research funding networks in data papers. **Information Development**, 2025. Disponível em:

<https://doi.org/10.1177/02666669251352185>. Acesso em: 24 jul. 2025.

UNIVERSIDADE DE SÃO PAULO (USP). **Ibict lança Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**. São Paulo, 03 de outubro de 2016. Disponível em: https://www.acessoaberto.usp.br/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada/?doing_wp_cron=1750111127.4176180362701416015625. Acesso em: 4 ago. 2025.

UNIVERSIDADE DE SÃO PAULO (USP). **Ciência Aberta USP**. São Paulo, [2025a?]. Disponível em: <https://cienciaaberta.usp.br/>. Acesso em: 22 jul. 2025.

UNIVERSIDADE DE SÃO PAULO (USP). **Anuário Estatístico**. São Paulo, [2025b?]. Disponível em: <https://uspdigital.usp.br/anuario/AnuarioControle>. Acesso em: 22 jul. 2025.

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC). Centro de Ciências da Educação. Programa de Pós-Graduação em Ciência da Informação. Linhas de pesquisa. Florianópolis, 2023. Disponível em: <https://pgcin.ufsc.br/areas-de-pesquisa/>. Acesso em: 20 dez. 2023.

UNIVERSITY OF CALIFORNIA (UC). Office of Scholarly Communication. **Campus Library Open Access Publishing Funds**. June 18, 2025. Disponível em: <https://osc.universityofcalifornia.edu/for-authors/open-access-publishing-at-uc/oa-publishing-funds/>. Acesso em: 22 jul. 2025.

VAN OORT, T; NOORDEGRAAF, J. Structured data for performing arts history: na introduction to a special issue of data papers. **Research Data Journal for the Humanities and Social Sciences**, v. 5, n. 2, p. 1-12, 2020. Disponível em: <https://www.sciencedirect.com/org/science/article/pii/S2452366623000385>. Acesso em: 23 nov. 2025.

VÁZQUEZ, I. *et al*. Improvements for research data repositories: the case of text spam. **Journal of Information Science**, v. 49, n. 2, 2021. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0165551521998636>. Acesso em: 10 jan. 2024.

VITÓN-CASTILLO, A. A.; GARCÍA-ESPINOSA, E.; ARENCIBIA-PAREDES, N. M. Bases para la implementación de la ciencia abierta. **Revista Información Científica**, Guantánamo, v. 99, n. 2., p.168-177, 2020. Disponível em:

<http://www.revinfcientifica.sld.cu/index.php/ric/article/view/2890>. Acesso em: 20 dez. 2023.

VOLPATO, G. **Ciência**: da filosofia à publicação. 7.ed. Botucatu: Best Writing, 2019.

WALTERS, W. H. Data journals: incentivizing data access and documentation within the scholarly communication system. **Insights: the UKSG journal**, v.33, n.1, p. 1-18, 2020. Disponível em: <https://insights.uksg.org/articles/10.1629/uksg.510>. Acesso em: 11 jul. 2022.

WANG, X.; XU, L. What Are journals and reviewers concerned about in data papers? Evidence from journal guidelines and review reports. **Learned Publishing**, v. 38, e2001, p. 1-14, 2025. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/leap.2001?af=R>. Acesso em: 26 jun. 2025.

WEITZEL, S. R. As publicações científicas como fonte de renda econômica de editores comerciais: o acesso aberto solapado. *In*: PRÍNCIPE, E.; RODE, S. M. (org.). **Comunicação científica aberta**. Rio de Janeiro: Ibict, 2022. p. 173-188.

WIGDOROWITZ, M. *et al.* It takes a village! Editorship, advocacy, and research in running an open access data journal. **Publications**, v. 12, n. 3, 2024. Disponível em: <https://doi.org/10.3390/publications12030024>. Acesso em: 15 set. 2025.

WILKINSON, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 11 jan. 2024.

WOLFRAM, D. *et al.* Open peer review: promoting transparency in open science. **Scientometrics**, n. 125, p. 1033-1051, 2020. Disponível em: <https://doi.org.ez46.periodicos.capes.gov.br/10.1007/s11192-020-03488-4>. Acesso em: 23 set. 2024.

YANG, W.; HUANG, J. The trend and impact of Open Science: three sunshines, three transformations, and four challenges. **Chinese Science Bulletin**, v.70, n. 10, p. 1434 -1439, 2025. Disponível em: <https://www.sciengine.com/CSB/doi/10.1360/TB-2024-1254?trans=true>. Acesso em: 16 jul. 2025.

YI, H. J. *et al.* Korean scholarly journal editors' and publishers' attitudes towards journal data sharing policies and data papers (2023): a survey-based descriptive study. **Science Editing**, v.10, n. 2, p. 141-148, 2023. Disponível em: <https://www.escienceediting.org/journal/view.php?number=322>. Acesso em: 26 ago. 2025.

ZHANG, L. *et al.* Should open access lead to closed research? The trends towards paying to perform research. **Scientometrics**, v.127, p. 7653-7679, 2022. Disponível em: <https://doi.org/10.1007/s11192-022-04407-5>. Acesso em: 1 nov. 2025.

APÊNDICE A – LISTA DE DOCUMENTOS CONSULTADOS NAS PÁGINAS DOS PERIÓDICOS E NOS *DATA PAPERS*

Quadro 16 - Lista de documentos consultados nas páginas dos periódicos e nos *data papers*

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
Annals of Forest Science	Articles https://annforsci.biomedcentral.com/articles	Valor Article processing charges (APC) https://annforsci.biomedcentral.com/submission-guidelines/fees-and-funding	Política de revisão por pares. https://annforsci.biomedcentral.com/submission-guidelines/peer-review-policy	Availability of data and materials https://www.biomedcentral.com/getpublished/editorial-policies#availability+of+data+and+materials	Availability of data and material
Big Earth Data	Lista de números https://www.tandfonline.com/loi/tbed20	Article Publishing Charges (APC) https://www.tandfonline.com/journals/tbed20/about-this-journal#open-access	Peer review https://authorservices.taylorandfrancis.com/editorial-policies/?_ga=2.245443588.1225536593.1739825393-1307913494.1739825393	Data sharing policies https://authorservices.taylorandfrancis.com/editorial-policies/?_ga=2.68625297.144688497.1739992272-1358742802.1739992272	Data availability statement
Biodiversity Data Journal	Artigos https://bdj.pensoft.net/articles	Article Processing Charges https://bdj.pensoft.net/about#ArticleProcessingCharges	Peer review https://bdj.pensoft.net/about#PeerReview	Data Publishing Guidelines https://bdj.pensoft.net/about#Datapublication	Data Resources ou Supplementary Material
BMC Bioinformatics	Busca em artigos https://bmcbioinformatics.biomedcentral.com/articles	Fees and funding https://bmcbioinformatics.biomedcentral.com/submission-guidelines/fees-and-funding	Peer-review policy https://bmcbioinformatics.biomedcentral.com/submission-guidelines/peer-review-policy	Availability of data and materials https://www.biomedcentral.com/getpublished/editorial-policies#availability+of+data+and+materials	Availability of data and materials ou additional information
BMC Research Notes	Articles	Fees and funding	Peer-review policy	Availability of data and materials	Availability of data materials

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
	https://bmresnotes.biomedcentral.com/articles	https://bmresnotes.biomedcentral.com/submission-guidelines/fees-and-funding	https://bmresnotes.biomedcentral.com/submission-guidelines/peer-review-policy	https://www.biomedcentral.com/getpublished/editorial-policies#availability+of+data+and+materials	
Cahiers Agricultures	Lista de números https://www.cahiersagricultures.fr/component/issues/?task=all&Itemid=121	APCs Price List https://www.edpsciences.org/en/publication-charges#cagri	Reviewer Tutorial https://www.ariessys.com/wp-content/uploads/EM-Reviewer-English.pdf	Não identificado	Description data
Chemical Data Collections	Issues https://www.sciencedirect.com/journal/chemical-data-collections/issues	Article publishing options https://www.sciencedirect.com/journal/chemical-data-collections	Peer review https://www.sciencedirect.com/journal/chemical-data-collections/publish/guide-for-authors	Research data https://www.elsevier.com/about/policies-and-standards/research-data	Tabela de especificações
Comparative Cytogenetics	Articles https://compcytogen.pensoft.net/articles	Article Charges https://compcytogen.pensoft.net/about#ArticleProcessingCharges	Guidelines for Reviewers https://compcytogen.pensoft.net/about#GuidelinesforReviewers	Data Publishing Guidelines https://compcytogen.pensoft.net/about#DataPublishingGuidelines	Supplementary materials
Data	Search for articles https://www.mdpi.com/journal/data	Open Access and Article Processing Charge (APC) https://www.mdpi.com/journal/data/apc	Data — Guidelines for Reviewers https://www.mdpi.com/journal/data/guidelines	Supplementary Materials, Data Deposit and Software Source Code https://www.mdpi.com/journal/data/instructions	Possui uma seção específica para inserir o link dos data papers, localizado após o resumo
Data in Brief	All issues https://www.sciencedirect.com/journal/data-in-brief/issues	Article publishing option https://www.sciencedirect.com/journal/data-in-brief	How Data in Brief manuscripts are reviewed https://www.sciencedirect.com/journal/data-in-brief/about/policies-and-	Sharing research data https://www.elsevier.com/researcher/author/tools-and-resources/research-data	Data acessibility

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
			guidelines/how-data-in-brief-manuscripts-are-reviewed		
Earth System Science Data	Data description paper https://essd.copernicus.org/data_description_paper.html	Article processing charges https://www.earth-system-science-data.net/about/article_processing_charges.html	Process of peer review and publication https://www.earth-system-science-data.net/peer_review/interactive_review_process.html	DataPolicy https://www.earth-system-science-data.net/policies/data_policy.html	Code and data availability
Ecology	Issue Archive https://esajournals.onlinelibrary.wiley.com/loi/19399170	Open Access https://esa.org/publications/open-access/	Peer Review https://esa.org/publications/be-an-author/peer-review/	Open Research Policy https://esa.org/publications/data-policy/	Supporting Information ou Open Research
Ecosistemas	Números anteriores https://www.revistaecosistemas.net/index.php/ecosistemas/issue/archive	Normas para envios https://www.revistaecosistemas.net/index.php/ecosistemas/about/submissions			Registro e disponibilidade dos dados
Frontiers in Bioinformatics	Articles https://www.frontiersin.org/journals/bioinformatics/articles	Publishing fees https://www.frontiersin.org/journals/bioinformatics/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/bioinformatics/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/bioinformatics/for-authors/author-guidelines	Data availability statement
Frontiers in Earth Science	Articles https://www.frontiersin.org/journals/earth-science/articles	Publishing fees https://www.frontiersin.org/journals/earth-science/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/earth-science/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/earth-science/for-authors/author-guidelines	Data availability
Frontiers in Ecology And Evolution	Articles	Publishing fees	Editor guidelines	Author guidelines	Data availability statement

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
	https://www.frontiersin.org/journals/ecology-and-evolution/articles	https://www.frontiersin.org/journals/ecology-and-evolution/for-authors/publishing-fees	https://www.frontiersin.org/journals/ecology-and-evolution/for-authors/editor-guidelines	https://www.frontiersin.org/journals/ecology-and-evolution/for-authors/author-guidelines	
Frontiers in Genetics	Articles https://www.frontiersin.org/journals/genetics/articles	Publishing fees https://www.frontiersin.org/journals/genetics/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/genetics/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/genetics/for-authors/author-guidelines	Data availability statement
Frontiers in Marine Science	Articles https://www.frontiersin.org/journals/marine-science/articles	Publishing fees https://www.frontiersin.org/journals/marine-science/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/marine-science/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/marine-science/for-authors/author-guidelines	Data availability
Frontiers in Microbiology	Articles https://www.frontiersin.org/journals/microbiology/articles	Publishing fees https://www.frontiersin.org/journals/microbiology/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/microbiology/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/microbiology/for-authors/author-guidelines	Não informou ao longo do texto
Frontiers in Molecular Biosciences	Articles https://www.frontiersin.org/journals/molecular-biosciences/articles	Publishing fees https://www.frontiersin.org/journals/molecular-biosciences/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/molecular-biosciences/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/molecular-biosciences/for-authors/author-guidelines	Não informou ao longo do texto
Frontiers in Neuroscience	Articles https://www.frontiersin.org/journals/neuroscience/articles	Publishing fees https://www.frontiersin.org/journals/neuroscience/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/neuroscience/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/neuroscience/for-authors/author-guidelines	Data availability
Frontiers in Psychology	Articles	Publishing fees	Editor guidelines	Author guidelines	Data availability

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
	https://www.frontiersin.org/journals/psychology/articles	https://www.frontiersin.org/journals/psychology/for-authors/publishing-fees	https://www.frontiersin.org/journals/psychology/for-authors/editor-guidelines	https://www.frontiersin.org/journals/psychology/for-authors/author-guidelines	
Frontiers in Public Health	Articles https://www.frontiersin.org/journals/public-health/articles	Publishing fees https://www.frontiersin.org/journals/public-health/for-authors/publishing-fees	Editor guidelines https://www.frontiersin.org/journals/public-health/for-authors/editor-guidelines	Author guidelines https://www.frontiersin.org/journals/public-health/for-authors/author-guidelines	Data availability
Geoscience Data Journal	Issue archive https://rmets.onlinelibrary.wiley.com/loi/20496060	Article Publication Charges https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/article-publication-charge	Editorial policies https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/about/author-guidelines#Editorialpolicies	Appropriate Repositories and Data Centres https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/about/author-guidelines#ardc	No início do data paper arrola todos os dados utilizados
Gigascience	Articles https://academic.oup.com/gigascience/search-results?page=1&_ContentType=Journal+Article&fl_SiteID=5589&sort=Date+%E2%80%93+Newest+First	Open Access and APCs https://academic.oup.com/gigascience/pages/charges_licensing_and_self_archiving	Editorial Policies & Reporting Standards https://academic.oup.com/gigascience/pages/editorial_policies_and_reporting_standards	Availability of Data and Materials https://academic.oup.com/gigascience/pages/editorial_policies_and_reporting_standards#Availability%20of%20Data%20and%20Materials	Availability of Supporting data
Global Ecology and Biogeography	Issue archive https://onlinelibrary.wiley.com/loi/14668238	Article Publication Charges (APCs) https://onlinelibrary.wiley.com/page/journal/14668238/homepage/fundedaccess.html	Editorial policies and ethical considerations https://onlinelibrary.wiley.com/page/journal/14668238/homepage/forauthors.html#policies	Author Guidelines https://onlinelibrary.wiley.com/page/journal/14668238/homepage/forauthors.html	Open Research ou Supporting Information

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
Human Mutation	Articles https://onlinelibrary.wiley.com/index/humu	Open Access Charges and Licensing https://onlinelibrary.wiley.com/page/journal/humu/homepage/open-access	Peer review https://onlinelibrary.wiley.com/page/journal/humu/homepage/author-guidelines	Wiley's Data Sharing Policies https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html	Não informou
Journal of Biogeography	Issue archive https://onlinelibrary.wiley.com/loi/13652699	Article Publication Charges (APCs) https://onlinelibrary.wiley.com/page/journal/13652699/homepage/fundedaccess.html	Editorial policies and ethical considerations https://onlinelibrary.wiley.com/page/journal/13652699/homepage/forauthors.html#PreviousReviews	Author Guidelines https://onlinelibrary.wiley.com/page/journal/13652699/homepage/forauthors.html#preparingsubmissions	Open research and Supporting Information
Journal of Open Humanities Data	Articles https://openhumanitiesdata.metajnl.com/articles	Publication fees https://openhumanitiesdata.metajnl.com/about/submissions#publication-fees	Editorial Policies https://openhumanitiesdata.metajnl.com/about/editorialpolicies	Data policy https://openhumanitiesdata.metajnl.com/about/editorialpolicies	Repository name
Marine Genomics	All Issues https://www.sciencedirect.com/journal/marine-genomics/issues	Article publishing options https://www.sciencedirect.com/journal/marine-genomics	Peer review https://www.sciencedirect.com/journal/marine-genomics/publish/guide-for-authors	Research data https://www.sciencedirect.com/journal/marine-genomics/publish/guide-for-authors	Nucleotide Sequence Accession
Microbiology Resource Announcements	Currents https://journals.asm.org/toc/mra/current	Publication fees https://journals.asm.org/publication-fees	Peer Review Process https://journals.asm.org/peer-review-process?journal=mra	Open Data Policy https://journals.asm.org/open-data-policy	Data Availability
Nature Conservation	Issues	Article Processing Charges	Peer Review	Data Publishing Guidelines	Supplementary material

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
	https://natureconservation.pensoft.net/issues	https://natureconservation.pensoft.net/about#ArticleProcessingCharges	https://natureconservation.pensoft.net/about#PeerReview	https://natureconservation.pensoft.net/about#DataPublishingGuidelines	
Proteomics	All Issues https://analyticalsciencejournals.onlinelibrary.wiley.com/loi/16159861	Article Publication Charges (APCs) https://analyticalsciencejournals.onlinelibrary.wiley.com/hub/journal/16159861/homepage/2120_onlineopen.html	Author Guidelines https://analyticalsciencejournals.onlinelibrary.wiley.com/hub/journal/16159861/homepage/2120_authorguidelines.html#nih		Supporting Information
Science of Nature (Naturwissenschaften)	Articles https://link.springer.com/journal/114/articles	Fees and funding https://link.springer.com/journal/114/how-to-publish-with-us#Fees%20and%20funding	Peer-Review Policy https://www.springer.com/gp/editorial-policies/peer-review-policy-process	Data Availability Statement https://www.springer.com/gp/editorial-policies/data-availability-statement	Electronic supplementary material
Scientific Data	Research articles https://www.nature.com/sdata/research-articles	Open Access Fees and Funding https://www.nature.com/sdata/open-access	Peer-review and publication criteria https://www.nature.com/sdata/policies/editorial-and-publishing-policies#publication-criteria	Data Policy https://www.nature.com/sdata/policies/data-policies	Machine Accessible metadata file
Scientific Reports	Research articles https://www.nature.com/srep/research-articles	Open Access Fees and Funding https://www.nature.com/srep/open-access	Peer-review policies https://www.nature.com/srep/journal-policies/peer-review	Editorial and publishing policies https://www.nature.com/srep/journal-policies/editorial-policies	Data records
Taxon	Issue archive	Article Publication Charges (APCs)	Editorial policies and ethical considerations	Data sharing and Data Accessibility	Não informou

Nome do periódico	Ano e publicação de <i>data papers</i>	Valor APC	Revisão por pares	Orientação para os dados	Seção do <i>data paper</i>
	https://onlinelibrary.wiley.com/loi/19968175	https://onlinelibrary.wiley.com/page/journal/19968175/homepage/fundedaccess.html	https://onlinelibrary.wiley.com/page/journal/19968175/homepage/author-guidelines#edpoliciesandconsid	https://onlinelibrary.wiley.com/page/journal/19968175/homepage/author-guidelines#prepsub	
ZooKeys	Articles https://zookeys.pensoft.net/articles	Article Processing Charges https://zookeys.pensoft.net/about#ArticleProcessingCharges	Peer Review https://zookeys.pensoft.net/about#PeerReview	Data Publishing Guidelines https://zookeys.pensoft.net/about#DataPublishingGuidelines	Dataset description

Fonte: elaborado pela autora

APÊNDICE B – LISTA DOS DOCUMENTOS CONSULTADOS NAS PÁGINAS DOS REPOSITÓRIOS

Quadro 17 - Repositórios e documentos sobre o depósito, cobrança e revisão dos dados

N	Repositório	Depósito/ Cobrança	Revisão
1	Mendeley Data	What are the requirements for data sets? https://data.mendeley.com/faq	Are data set moderated? https://data.mendeley.com/faq
2	Zenodo	What are the size limitations of Zenodo? https://support.zenodo.org/help/en-gb/1-upload-deposit/80-what-are-the-size-limitations-of-zenodo	Why is my record in review? https://support.zenodo.org/help/en-gb/6-collaborate-and-share/3-why-is-my-record-in-review
3	Figshare	File size limits and storage https://info.figshare.com/user-guide/file-size-limits-and-storage/	Step 5: Submit Your Items for Review https://info.figshare.com/user-guide/figshare-plus-user-guide/#step_5
4	Sequence Read Archive	SRA End User Cloud Access Costs https://www.ncbi.nlm.nih.gov/sra/docs/sra-cloud-access-costs/	SRA Submission Quick Start https://www.ncbi.nlm.nih.gov/sra/docs/submit/
5	Genbank	NLM GenBank and SRA Data Processing https://www.ncbi.nlm.nih.gov/sra/docs/sequence-data-processing/	Validation https://www.ncbi.nlm.nih.gov/genbank/validation/
6	Pangaea	Benefits and Details https://www.pangaea.de/submit/	Submit Data to PANGAEA https://www.pangaea.de/submit/
7	Lancaster University Research Directory	Pure Support Pages https://www.lancaster.ac.uk/research/research-services/research-information--systems/pure/	Não informa
8	Open Science Framework	How can I get started on using OSF? https://help.osf.io/article/546-getting-started-faq-s	Content Plagiarism, Review, and Appeals Procedures https://help.osf.io/article/546-getting-started-faq-s
9	Global Biodiversity Information Facility	Quick guide to publishing data through GBIF.org https://www.gbif.org/publishing-data	Data quality requirements https://www.gbif.org/data-quality-requirements
10	Github	Pricing https://github.com/pricing	
11	BioProject	https://www.ncbi.nlm.nih.gov/bioproject/	Quais tipos de validação meu BioProject deve passar? https://www.ncbi.nlm.nih.gov/bioproject/docs/faq/

N	Repositório	Depósito/ Cobrança	Revisão
12	Proteomics Identifications Database	What is PRIDE Archive https://www.ebi.ac.uk/pride/markdownpage/citationpage	What happens to my data after upload https://www.ebi.ac.uk/pride/markdownpage/submitdatapage
13	Gene Expression Omnibus	What is GEO? https://www.ncbi.nlm.nih.gov/geo/info/faq.html	Submitting data https://www.ncbi.nlm.nih.gov/geo/info/submission.html
14	Dryad	Fees https://datadryad.org/join_us#fees	Our curation and publication process https://datadryad.org/mission#our-curation-and-publication-process
15	European Nucleotide Archive	Data submission https://www.ebi.ac.uk/submission/	
16	Mass Spectrometry Interactive Virtual Environment (Massive)	MassIVE/GNPS Fair Use Policy https://ccms-ucsd.github.io/MassIVEDocumentation/#fair_use_policy/	
17	Cambridge Crystallographic Data Center	Donations https://www.ccdc.cam.ac.uk/about-us/legal-status/donations/	Scientific Data Preservation https://www.ccdc.cam.ac.uk/community/access-deposit-structures/deposit-a-structure/scientific-data-preservation/
18	Science Data Bank	How much do you charge for data publishing ? https://www.scidb.cn/en/help?p=faq	Publishing Processes https://www.scidb.cn/en/help?p=publishing_process
19	GigaScience Database	What are the charges for submitting data? https://gigadb.org/faq	What curation do you carry out? https://gigadb.org/faq
20	BioSample	Scope https://www.ncbi.nlm.nih.gov/books/NBK169436/	Will NCBI apply further curation to my BioSample records? https://www.ncbi.nlm.nih.gov/biosample/docs/submission/faq/
21	Synapse	Plans https://www.synapse.org/Plans:default	Plans https://www.synapse.org/Plans:default
22	Dados Abertos do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	Quick guide to publishing data through GBIF.org https://www.gbif.org/publishing-data	Não informa
23	Plasmodium Genomics Resource	Donate https://plasmodb.org/plasmo/app#support	Frequently Asked Questions https://plasmodb.org/plasmo/app/static-content/faq.html
24	Atmospheric Science Data Center (ASDC)	Data Publication Guidelines https://pub.earthdata.nasa.gov/data_publication_guidelines#publication	Data Publication Guidelines https://pub.earthdata.nasa.gov/data_publication_guidelines#publication
25	Data repository	Página fora do ar	
26	National Centers for Environmental Information	Why should I archive my data? https://www.ncei.noaa.gov/archive	How does NCEI decide if my data will be archived? https://www.ncei.noaa.gov/archive

N	Repositório	Depósito/ Cobrança	Revisão
27	ArrayExpress	Submission overview https://www.ebi.ac.uk/fg/annotate/help/what_to_submit.html	Validation https://www.ebi.ac.uk/fg/annotate/help/validate_exp.html#validate
28	Harvard DataVerse	About https://support.dataverse.harvard.edu/	
29	Map of Life (MOL)	Who we are https://mol.org/about/who-we-are	Não informa
30	RCSB Protein Data Bank	Starting a deposition session https://www.wwpdb.org/deposition/faq#start	Validation https://www.wwpdb.org/validation/validation-reports
31	Sea Scientific Open Data Publication	About Seanoe https://www.seanoe.org/html/about.htm	
32	MetaboLights	Data Submission Updates https://www.ebi.ac.uk/metabolights/newWorkflow	Data Submission Updates https://www.ebi.ac.uk/metabolights/newWorkflow
33	Environmental Data Initiative Repository	How much do EDI services cost? https://edirepository.org/support/frequently-asked-questions#how-much-do-edi-services-cost	
34	Herbarium of Federal University of Mato Grosso do Sul	Quick guide to publishing data through GBIF.org https://www.gbif.org/publishing-data	Data quality requirements https://www.gbif.org/data-quality-requirements
35	NCBI DataSets	Informações não localizadas posteriormente	
36	Immunology Database and Analysis Portal	Data Submission Introduction https://docs.immport.org/datasubmission/general/	
37	Ag Data Commons	Ag Data Commons Collection Policies https://www.nal.usda.gov/services/agdatacommons/policies	
38	Archive of Data on Disability to Enable Policy	https://www.disabilitystatistics.org/dataset-directory/repository/17	Não informa
39	British Antarctic Survey	Data deposit https://www.bas.ac.uk/data/uk-pdc/data-deposit/	
40	British Oceanographic Data Centre	General Guide https://www.bodc.ac.uk/submit_data/submission_guidelines/general_guide/	Data processing steps https://www.bodc.ac.uk/submit_data/what_do_we_do_with_your_data/data_processing_steps/
41	Cirad dataverse	Home https://dataverse.cirad.fr/	
42	Dagshub	Pricing https://dagshub.com/pricing/	
43	DataSuds	Datasuds https://dataverse.ird.fr/	

N	Repositório	Depósito/ Cobrança	Revisão
44	DataverseNL	Fee https://dans.knaw.nl/en/data-services/dataversenl/	Disclaimer https://dans.knaw.nl/en/disclaimer/
45	Digital CSIC	Formulario Digital.CSIC: envío de documentos para su archivo en el repositorio https://digital.csic.es/dc/peticiones/?locale=es	
46	Digital Rocks Portal	Digital porous media portal policy, data and software usage agreement https://digital-porous-media.github.io/dpm_docs/user_agreement/	
47	Fairdata IDA Research Data Storage Service	Safe storage for FAIR research data https://www.fairdata.fi/en/ida/	Não informa
48	Freshwater Research and Environmental Database	General Data Protection Regulation (GDPR) https://fred.igb-berlin.de/datapolicy/data_protection_declaration	Não informa
49	GFZ Data Service	Publication Instructions https://dataservices.gfz-potsdam.de/web/publish-data/publication-instructions	
50	Global freshwater biodiversity research	Site fora do ar	
51	Global Health Data Exchange	https://ghdx.healthdata.org/about-ghdx/our-information-sources	Não informa
52	ICOS Carbon Portal	Data Flow https://www.icos-cp.eu/data-services/data-collection/data-flow	
53	International Neuroimaging Data-Sharing Initiative	Data & Resources https://fcon_1000.projects.nitrc.org/	Não informa
54	IsoArch	Who can use the Isoarch database? https://isoarch.org/faq	How is the data curated? https://isoarch.org/faq
55	Materials Cloud repository	What is the maximum file size I can upload? https://archive.materialscloud.org/help	Why do I have to register to the Materials Cloud Archive to upload a record? https://archive.materialscloud.org/help
56	Metagenomics analysis server	Terms of Service https://www.mg-rast.org/legal.html	
57	NoMaD Repository & Archive	NOMAD & NOMAD Oasis https://nomad-lab.eu/nomad-lab/	
58	OTELo Research Data Repository	About ORDaR https://ordar.otelo.univ-lorraine.fr/about	
59	ProteomeXchange	What happens in the case proteomics datasets do not fulfil all the PX requirements? https://www.proteomexchange.org/docs/guidelines_px.pdf	

N	Repositório	Depósito/ Cobrança	Revisão
60	Repositório da UNESP	O que é um repositório institucional? https://unesp.libanswers.com/faq/291314	Concluí a submissão. Quando receberei o e-mail informando sobre o aceite? https://unesp.libanswers.com/faq/291353
61	Repositório da Universidade de São Paulo	Repositório de Dados Científicos USP e outras iniciativas na universidade https://www.abcd.usp.br/apoio-pesquisador/dados-pesquisa/iniciativas-universidade-sao-paulo-gestao-dados/	
62	Repositório de Dados de Pesquisa da Fundação Getulio Vargas	Repositório de Dados de Pesquisa da Fundação Getulio Vargas https://dataverse.fgv.br/	Não informa
63	Repositório de Dados de Pesquisa da Unicamp (REDU)	O que é o REDU? https://www.sbu.unicamp.br/sbu/duvidas-frequentes-redu/	Quem deve depositar os dados no REDU? https://www.sbu.unicamp.br/sbu/duvidas-frequentes-redu/
64	Repository of Open Access Datasets	Página fora do ar	
65	Saint Mary's University Dataverse	SMU Dataverse https://libguides.smu.ca/rdm/repositories	Não informa
66	TRY	Intellectual Property Guidelines for the TRY Plant Trait Database https://www.try-db.org/TryWeb/TRY_Intellectual_Property_Guidelines.pdf	

Fonte: elaborado pela autora.

APÊNDICE C – LISTA DE GRANDES ÁREAS, ÁREAS E TEMAS DOS DATA PAPERS

Quadro 18 - Grandes áreas, áreas e temas dos *data papers*

Grande área	Área	Tema	Quantidade	%
Ciências da Saúde			77	12,15
	Enfermagem		2	0,32
		Nutrição e Dietética	2	0,32
	Medicina		66	10,41
		Cardiologia e Medicina Cardiovascular	1	0,16
		Cirurgia	3	0,47
		Doenças Infecciosas	5	0,79
		Epidemiologia	9	1,42
		Farmacologia	4	0,63
		Fisiologia	4	0,63
		Genética	1	0,16
		Medicina complementar e alternativa	1	0,16
		Medicina de emergência	1	0,16
		Medicina interna	1	0,16
		Medicina Pulmonar e Respiratória	4	0,63
		Medicina Reprodutiva	1	0,16
		Nefrologia	1	0,16
		Neurologia	1	0,16
		Oftalmologia	1	0,16
		Oncologia	5	0,79
		Ortopedia e Medicina Esportiva	2	0,32
		Patologia e Medicina Legal	1	0,16
		Pediatria, Perinatologia e Saúde Infantil	1	0,16
		Psiquiatria e Saúde Mental	1	0,16
		Radiologia, Medicina Nuclear e Imagem	6	0,95
		Saúde Pública, Saúde Ambiental e Ocupacional	12	1,89
	Profissões de saúde		7	1,10
		Fisioterapia, Terapia Desportiva e Reabilitação	1	0,16
		Gestão de Informação em Saúde	1	0,16
		Profissões de Saúde Geral	2	0,32

Grande área	Área	Tema	Quantidade	%
		Serviços médicos de emergência	1	0,16
		Terapia complementar e manual	2	0,32
	Veterinária		2	0,32
		Pequenos animais	2	0,32
Ciências da Vida			207	32,65
	Bioquímica, Genética e Biologia Molecular		85	13,41
		Biofísica	1	0,16
		Biologia Celular	3	0,47
		Biologia do Desenvolvimento	1	0,16
		Biologia Molecular	45	7,10
		Bioquímica	1	0,16
		Bioquímica Clínica	2	0,32
		Biotecnologia	4	0,63
		Endocrinologia	3	0,47
		Envelhecimento	1	0,16
		Genética	19	3,00
		Medicina molecular	1	0,16
		Pesquisa sobre o câncer	4	0,63
	Ciências Agrárias e Biológicas		93	14,67
		Ciência Animal e Zoologia	3	0,47
		Ciência das Plantas	25	3,94
		Ciência do solo	4	0,63
		Ciência dos alimentos	8	1,26
		Ciência dos insetos	4	0,63
		Ecologia, Evolução, Comportamento e Sistemática	46	7,26
		Horticultura	1	0,16
		Silvicultura	2	0,32
	Farmacologia, Toxicologia e Farmacêutica		6	0,95
		Ciência Farmacêutica	1	0,16
		Farmacologia	3	0,47
		Toxicologia	2	0,32
	Imunologia e Microbiologia		18	2,84
		Imunologia	5	0,79

Grande área	Área	Tema	Quantidade	%
		Microbiologia	1	0,16
		Parasitologia	9	1,42
		Virologia	3	0,47
	Neurociência		5	0,79
		Neurociência Celular e Molecular	1	0,16
		Neurociência Cognitiva	3	0,47
		Neurologia	1	0,16
Ciências Físicas			316	49,84
	Ciência Ambiental		154	24,29
		Ciência e Tecnologia da Água	6	0,95
		Conservação da Natureza e da Paisagem	27	4,26
		Ecologia	42	6,62
		Engenharia Ambiental	9	1,42
		Engenharia Industrial e de Manufatura	5	0,79
		Gestão, Monitoramento, Política e Direito	5	0,79
		Modelagem Ecológica	18	2,84
		Mudança Global e Planetária	30	4,73
		Poluição	2	0,32
		Química Ambiental	4	0,63
		Saúde, Toxicologia e Mutagênese	6	0,95
	Ciência da Computação		17	2,68
		Inteligência artificial	6	0,95
		Processamento de sinais	1	0,16
		Redes de Computadores e Comunicações	2	0,32
		Sistemas de Informação	3	0,47
		Teoria Computacional e Matemática	3	0,47
		Visão Computacional e Reconhecimento de Padrões	2	0,32
	Ciência dos Materiais		13	2,05
		Biomateriais	2	0,32
		Polímeros e Plásticos	2	0,32
		Química de Materiais	9	1,42
	Ciências da Terra e Planetárias		45	7,10
		Ciência atmosférica	16	2,52

Grande área	Área	Tema	Quantidade	%
		Geofísica	3	0,47
		Geoquímica e Petrologia	1	0,16
		Oceanografia	16	2,52
		Paleontologia	5	0,79
		Processos da superfície da Terra	4	0,63
	Energia		1	0,16
		Energias Renováveis, Sustentabilidade e Meio Ambiente	1	0,16
	Engenharia		50	7,89
		Construção e Edificação	6	0,95
		Engenharia Aeroespacial	1	0,16
		Engenharia Automotiva	3	0,47
		Engenharia Biomédica	8	1,26
		Engenharia Civil e Estrutural	7	1,10
		Engenharia de Controle e Sistemas	2	0,32
		Engenharia Elétrica e Eletrônica	5	0,79
		Engenharia Industrial e de Manufatura	2	0,32
		Engenharia Mecânica	6	0,95
		Engenharia Oceânica	3	0,47
		Mecânica dos Materiais	4	0,63
		Segurança, Risco, Confiabilidade e Qualidade	1	0,16
		Tecnologia de mídia	2	0,32
	Engenharia Química		3	0,47
		Bioengenharia	1	0,16
		Filtração e Separação	1	0,16
		Fluxo de fluidos e processos de transferência	1	0,16
	Física e Astronomia		3	0,47
		Astronomia e Astrofísica	1	0,16
		Física Nuclear e de Altas Energias	1	0,16
		Radiação	1	0,16
	Matemática		5	0,79
		Modelagem e Simulação	5	0,79
	Química		25	3,94
		Espectroscopia	5	0,79

Grande área	Área	Tema	Quantidade	%
		Química Analítica	5	0,79
		Química Inorgânica	3	0,47
		Química orgânica	12	1,89
Ciências Sociais			34	5,36
	Artes e Humanidades		2	0,32
		Arqueologia	1	0,16
		História	1	0,16
	Ciências da Decisão		3	0,47
		Ciência da Administração e Pesquisa Operacional	2	0,32
		Sistemas de Informação e Gestão	1	0,16
	Ciências Sociais		11	1,74
		Educação	1	0,16
		Geografia, Planejamento e Desenvolvimento	1	0,16
		Pesquisa de Segurança	2	0,32
		Saúde	1	0,16
		Sociologia e Ciência Política	4	0,63
		Transporte	2	0,32
	Economia, Econometria e Finanças		5	0,79
		Economia e Econometria	3	0,47
		Finança	2	0,32
	Negócios, Gestão e Contabilidade		5	0,79
		Comportamento Organizacional e Gestão de Recursos Humanos	1	0,16
		Contabilidade	1	0,16
		Marketing	2	0,32
		Sistemas de Informação Gerencial	1	0,16
	Psicologia		8	1,26
		Psicologia Clínica	4	0,63
		Psicologia Experimental e Cognitiva	1	0,16
		Psicologia Social	3	0,47
Total Geral			634	100

Fonte: elaborado pela autora

Notas: Coluna F: refere-se à quantidade de *data papers* pertencente a área e tema.

Coluna %: refere-se à porcentagem de *data papers* concentradas no tema em relação ao total identificado

[illegible]

[illegible]

País		Educação		Instituto de pesquisa		Governo		ONG		Empresa		Saúde		Arquivo		Outros		Total	
		F	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%		
Europa		518	20,47	213	8,42	66	2,61	38	1,50	18	0,71	39	1,54	20	0,79	1	0,04	913	36,09
França	FR	66	2,61	54	2,13	12	0,47	5	0,20	1	0,04	6	0,24	1	0,04	1	0,04	146	5,77
Alemanha	DE	68	2,69	35	1,38	6	0,24	3	0,12	4	0,16	8	0,32	5	0,20	-	-	129	5,10
Grã-Bretanha	GB	75	2,96	13	0,51	7	0,28	6	0,24	3	0,12	5	0,20	5	0,20	-	-	114	4,51
Espanha	ES	54	2,13	25	0,99	7	0,28	7	0,28	1	0,04	5	0,20	2	0,08	-	-	101	3,99
Itália	IT	41	1,62	14	0,55	8	0,32	5	0,20	1	0,04	3	0,12	-	-	-	-	72	2,85
Portugal	PT	24	0,95	8	0,32	-	-	2	0,08	-	-	1	0,04	-	-	-	-	35	1,38
Suíça	CH	17	0,67	7	0,28	3	0,12	3	0,12	1	0,04	1	0,04	2	0,08	-	-	34	1,34
Holanda	NL	14	0,55	8	0,32	1	0,04	3	0,12	2	0,08	5	0,20	1	0,04	-	-	34	1,34
Polônia	PL	26	1,03	5	0,20	2	0,08	-	-	-	-	-	-	-	-	-	-	33	1,30
Bélgica	BE	12	0,47	3	0,12	1	0,04	1	0,04	3	0,12	-	-	2	0,08	-	-	22	0,87
Noruega	NO	9	0,36	11	0,43	-	-	-	-	1	0,04	-	-	-	-	-	-	21	0,83
Suécia	SE	13	0,51	5	0,20	-	-	1	0,04	-	-	1	0,04	-	-	-	-	20	0,79
Grécia	GR	13	0,51	2	0,08	2	0,08	-	-	-	-	1	0,04	-	-	-	-	18	0,71
Finlândia	FI	9	0,36	1	0,04	3	0,12	-	-	-	-	-	-	-	-	-	-	13	0,51
Eslováquia	SK	7	0,28	4	0,16	1	0,04	-	-	-	-	1	0,04	-	-	-	-	13	0,51
República Tcheca	CZ	9	0,36	2	0,08	1	0,04	-	-	-	-	-	-	-	-	-	-	12	0,47
Áustria	AT	8	0,32	1	0,04	-	-	2	0,08	-	-	-	-	-	-	-	-	11	0,43
Hungria	HU	4	0,16	5	0,20	2	0,08	-	-	-	-	-	-	-	-	-	-	11	0,43
Romênia	RO	9	0,36	-	-	1	0,04	-	-	-	-	-	-	-	-	-	-	10	0,40
Dinamarca	DK	6	0,24	1	0,04	-	-	-	-	1	0,04	-	-	1	0,04	-	-	9	0,36
Ucrânia	UA	3	0,12	2	0,08	3	0,12	-	-	-	-	-	-	-	-	-	-	8	0,32
Bulgária	BG	3	0,12	2	0,08	1	0,04	-	-	-	-	-	-	1	0,04	-	-	7	0,28
Irlanda	IE	6	0,24	-	-	1	0,04	-	-	-	-	-	-	-	-	-	-	7	0,28
Eslovênia	SI	2	0,08	4	0,16	-	-	-	-	-	-	1	0,04	-	-	-	-	7	0,28
Estônia	EE	5	0,20	-	-	1	0,04	-	-	-	-	-	-	-	-	-	-	6	0,24
Croácia	HR	5	0,20	-	-	-	-	-	-	-	-	1	0,04	-	-	-	-	6	0,24
Cyprus	CY	3	0,12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	0,12
Islândia	IS	-	-	1	0,04	1	0,04	-	-	-	-	-	-	-	-	-	-	2	0,08

País		Educação		Instituto de pesquisa		Governo		ONG		Empresa		Saúde		Arquivo		Outros		Total	
		F	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%
Lituânia	LT	1	0,04	-	-	1	0,04	-	-	-	-	-	-	-	-	-	-	2	0,08
Sérvia	RS	2	0,08	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0,08
Geórgia	GE	1	0,04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,04
Luxemburgo	LU	1	0,04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,04
Latvia	LV	1	0,04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,04
Moldávia	MD	-	-	-	-	1	0,04	-	-	-	-	-	-	-	-	-	-	1	0,04
Macedônia	MK	1	0,04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,04
Europa/Ásia		45	1,78	11	0,43	6	0,24	-	-	-	-	-	-	1	0,04	-	-	63	2,49
Turquia	TR	29	1,15	1	0,04	3	0,12	-	-	-	-	-	-	-	-	-	-	33	1,30
Rússia	RU	14	0,55	10	0,40	3	0,12	-	-	-	-	-	-	1	0,04	-	-	28	1,11
Azerbaijão	AZ	2	0,08	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0,08
Oceania		46	1,82	21	0,83	20	0,79	6	0,24	4	0,16	2	0,08	6	0,24	2	0,08	107	4,23
Austrália	AU	38	1,50	17	0,67	17	0,67	5	0,20	4	0,16	2	0,08	6	0,24	2	0,08	91	3,60
Nova Zelândia	NZ	8	0,32	4	0,16	3	0,12	-	-	-	-	-	-	-	-	-	-	15	0,59
Papua Nova Guiné	PG	-	-	-	-	-	-	1	0,04	-	-	-	-	-	-	-	-	1	0,04
Total Geral		1486	58,74	408	16,13	259	10,24	157	6,21	85	3,36	74	2,92	50	1,98	11	0,43	2530	100

Fonte: elaborado pela autora

Notas: Coluna F: refere-se à quantidade de instituições caracterizadas conforme classificação ROR.

Coluna %: refere-se à porcentagem de instituições.

APÊNDICE E – LISTA DE PAÍSES E TIPOLOGIAS DE INSTITUIÇÕES DE FINANCIAMENTO

Tabela 13 - Lista de países e tipologia das instituições de financiamento

Tipologia	Pública		Privada		ONG		Total	
Países	F	%	F	%	F	%	F	%
África	7	0,93	1	0,13	1	0,13	9	1,20
África do Sul	6	0,80	-	-	-	-	6	0,80
Burundi	1	0,13	-	-	-	-	1	0,13
Nigéria	-	-	1	0,13	-	-	1	0,13
Quênia	-	-	-	-	1	0,13	1	0,13
América Central	2	0,27	-	-	1	0,13	3	0,40
Costa Rica	1	0,13	-	-	-	-	1	0,13
Panamá	-	-	-	-	1	0,13	1	0,13
Porto Rico	1	0,13	-	-	-	-	1	0,13
América do Norte	99	13,15	33	4,38	43	5,71	175	23,24
Estados Unidos	69	9,16	30	3,98	40	5,31	139	18,46
Canadá	24	3,19	3	0,40	3	0,40	30	3,98
México	6	0,80	-	-	-	-	6	0,80
América do Sul	138	18,33	30	3,98	4	0,53	172	22,84
Brasil	102	13,55	27	3,59	3	0,40	132	17,53
Argentina	12	1,59	-	-	-	-	12	1,59
Colômbia	5	0,66	3	0,40	-	-	8	1,06
Chile	6	0,80	-	-	-	-	6	0,80
Peru	6	0,80	-	-	-	-	6	0,80
Uruguai	3	0,40	-	-	1	0,13	4	0,53
Equador	3	0,40	-	-	-	-	3	0,40
Venezuela	1	0,13	-	-	-	-	1	0,13
Ásia	54	7,17	4	0,53	5	0,66	63	8,37
China	22	2,92	2	0,27	1	0,13	25	3,32
Japão	5	0,66	-	-	1	0,13	6	0,80
Hong Kong	5	0,66	-	-	-	-	5	0,66
Índia	4	0,53	-	-	1	0,13	5	0,66
Israel	3	0,40	1	0,13	-	-	4	0,53
Arábia Saudita	3	0,40	-	-	-	-	3	0,40
Coréia do Sul	3	0,40	-	-	-	-	3	0,40
Singapura	2	0,27	-	-	1	0,13	3	0,40
Vietnã	3	0,40	-	-	-	-	3	0,40
Bangladesh	1	0,13	-	-	-	-	1	0,13
Emirados Árabes Unidos	-	-	-	-	1	0,13	1	0,13
Filipinas	-	-	1	0,13	-	-	1	0,13
Irã	1	0,13	-	-	-	-	1	0,13
Nepal	1	0,13	-	-	-	-	1	0,13
Paquistão	1	0,13	-	-	-	-	1	0,13
Europa	230	30,54	25	3,32	26	3,45	281	37,32
Grã-Bretanha	38	5,05	3	0,40	7	0,93	48	6,37
Espanha	39	5,18	2	0,27	2	0,27	43	5,71
Alemanha	31	4,12	4	0,53	7	0,93	42	5,58
França	27	3,59	4	0,53	3	0,40	34	4,52
Holanda	15	1,99	-	-	-	-	15	1,99
Suíça	5	0,66	4	0,53	1	0,13	10	1,33
Bélgica	9	1,20	-	-	-	-	9	1,20

Tipologia	Pública		Privada		ONG		Total	
Países	F	%	F	%	F	%	F	%
Noruega	7	0,93	-	-	2	0,27	9	1,20
Portugal	9	1,20	-	-	-	-	9	1,20
Áustria	6	0,80	1	0,13	1	0,13	8	1,06
Dinamarca	2	0,27	4	0,53	2	0,27	8	1,06
Itália	7	0,93	-	-	1	0,13	8	1,06
Polônia	8	1,06	-	-	-	-	8	1,06
Suécia	6	0,80	1	0,13	-	-	7	0,93
Irlanda	3	0,40	1	0,13	-	-	4	0,53
República Tcheca	4	0,53	-	-	-	-	4	0,53
Estônia	3	0,40	-	-	-	-	3	0,40
Finlândia	3	0,40	-	-	-	-	3	0,40
Chipre	1	0,13	1	0,13	-	-	2	0,27
Eslováquia	2	0,27	-	-	-	-	2	0,27
Hungria	2	0,27	-	-	-	-	2	0,27
Letônia	2	0,27	-	-	-	-	2	0,27
Eslovênia	1	0,13	-	-	-	-	1	0,13
Europa/Ásia	15	1,99	-	-	-	-	15	1,99
Rússia	10	1,33	-	-	-	-	10	1,33
Turquia	5	0,66	-	-	-	-	5	0,66
Oceania	27	3,59	3	0,40	5	0,66	35	4,65
Austrália	20	2,66	3	0,40	5	0,66	28	3,72
Nova Zelândia	7	0,93	-	-	-	-	7	0,93
Total Geral	572	75,96	96	12,75	85	11,29	753	100

Fonte: elaborado pela autora

Notas: Coluna F: refere-se à quantidade de instituições de financiamento localizadas no país

Coluna %: refere-se à porcentagem de instituições de financiamento

APÊNDICE F – INDICADORES DOS PERIÓDICOS

Quadro 19 - Indicadores e avaliações dos periódicos

Periódico	Scopus			Web of Science			F*	T**	APC (USD)***
	SJR (2023)	Quartil	Subárea	JIF (2023)	Quartil	Áreas			
Gigascience	4.621	Q1	Aplicações em Ciência da Computação	11.8	Q1	Ciências Multidisciplinares	5	D	2.638,00
			Informática aplicada à saúde						
Earth System Science Data	4.231	Q1	Ciências da Terra e Planetárias (miscelânea)	11.2	Q1	Geociências, Multidisciplinar	36	D	1.521,66
						Meteorologia e Ciências Atmosféricas			
Global ecology and biogeography	2.744	Q1	Ecologia, Evolução, Comportamento e Sistemática	6.3	Q1	Ecologia	9	M	5.000,00
			Ecologia			Geografia, Física			
			Mudança Global e Planetária						
Ecology	1.945	Q1	Ecologia, Evolução, Comportamento e Sistemática	4.4	Q1	Ecologia	51	M	3.970,00
Scientific data	1.937	Q1	Aplicações em Ciência da Computação	5.8	Q1	Ciências Multidisciplinares	103	D	2.490,00
			Sistemas de Informação						
			Estatística, Probabilidade e Incerteza						
			Estatística e Probabilidade						
			Educação						
			Biblioteconomia e Ciência da Informação						
Human mutation	1.686	Q1	Genética	3.3	Q2	Genética e Hereditariedade	1	M	2.490,00
			Genética (clínica)						
Journal of biogeography	1.460	Q1	Ecologia, Evolução, Comportamento e Sistemática	3.4	Q1	Genética e Hereditariedade	1	M	4.830,00
			Ecologia		Q2	Ecology			
Frontiers in molecular biosciences	1.232	Q1	Bioquímica	3.9	Q2	Bioquímica e Biologia Molecular	1	M	512,00
			Bioquímica, Genética e Biologia Molecular (miscelânea)						
		Q2	Biologia Molecular						
	1.065	Q1	Microbiologia	4.0	Q2	Microbiologia	1	M	512,00

Periódico	Scopus			Web of Science			F*	T**	APC (USD)***
	SJR (2023)	Quartil	Subárea	JIF (2023)	Quartil	Áreas			
Frontiers in microbiology			Microbiologia (médical)						
Frontiers in neuroscience	1.063	Q2	Neurociências (miscelânea)	3.2	Q2	Neurociências	2	M	512,00
Proteomics	1.011	Q2	Bioquímica	3.4	Q2	Métodos de Pesquisa em Bioquímica	4	M	4.650,00
			Biologia molecular			Bioquímica e Biologia Molecular			
BMC bioinformatics	1.005	Q1	Aplicações em Ciência da Computação	2.9	Q1	Biologia Matemática e Computacional	2	M	2.990,00
			Matemática aplicada						
		Q2	Bioquímica		Q2	Métodos de Pesquisa em Bioquímica			
			Biologia molecular			Biotecnologia e Microbiologia Aplicada			
			Estrutura biológica						
Big earth data	0.928	Q1	Computação aplicada às Ciências da Terra	4.2	Q1	Ciência da Computação; Sistemas de Informação	1	D	815,00
		Q2	Aplicações em Ciência da Computação		Q2	Sensoriamento Remoto			
Frontiers in marine science	0.907	Q1	Ciências Aquáticas	2.8	Q1	Biologia Marinha e de Água Doce	3	M	512,00
			Oceanografia						
			Engenharia Oceânica						
			Ciências Ambientais (miscelânea)						
			Ciência e Tecnologia da Água						
		Q2	Mudança Global e Planetária						
Scientific reports	0.900	Q1	Multidisciplinar	3.8	Q1	Ciências Multidisciplinares	1	M	2.690,00
Frontiers in public health	0.895	Q1	Saúde Pública, Saúde Ambiental e Ocupacional	3.0	Q2	Saúde Pública, Ambiental e Ocupacional	1	M	512,00
Taxon	0.891	Q1	Ecologia, Evolução, Comportamento e Sistemática	3.0	Q2	Biologia Evolutiva	1	M	2.630,00
			Botânica			Botânica			
Frontiers in Ecology and Evolution	0.889	Q1	Ecologia, Evolução, Comportamento e Sistemática	2.4	Q2	Ecologia	3	M	512,00

Periódico	Scopus			Web of Science			F*	T**	APC (USD)***
	SJR (2023)	Quartil	Subárea	JIF (2023)	Quartil	Áreas			
			Ecologia						
Frontiers in genetics	0.853	Q2	Genética	2.8	Q2	Genética e Hereditariedade	2	M	512,00
			Medicina molecular						
			Genética (clínica)						
Frontiers in psychology	0.800	Q2	Psicologia (miscelânea)	2.6	Q2	Psiologia, Multidisciplinas	2	M	512,00
Annals of forest science	0.784	Q1	Silvicultura; Ecologia	2.5	Q1	Silvicultura	1	M	2.390,00
Frontiers in earth science	0.691	Q1	Ciências da Terra e Planetárias (miscelânea)	2.0	Q3	Geociências, Multidisciplinar	2	M	512,00
ZooKeys	0.659	Q1	Ciências Agrárias e Biológicas (miscelânea)	1.3	Q2	Zoologia	2	M	978,21
			Ciências Animal e Zoologia						
			Entomologia						
			Paleontologia						
		Q2	Ecologia, Evolução, Comportamento e Sistemática						
Marine genomics	0.584	Q2	Ciências Aquáticas	1.3	Q4	Genética e Hereditariedade	2	M	2.960,00
		Q3	Genética						
Science of Nature	0.535	Q2	Ecologia, Evolução, Comportamento e Sistemática	2.1	Q2	Ciências Multidisciplinares	1	M	3.790,00
			Medicina (miscelânea)						
Data	0.503	Q2	Sistemas de Informação	2.2	Q2	Ciências Multidisciplinares	17	D	1.805,66
			Sistemas de Informação e Gestão						
		Q3	Aplicações em Ciência da Computação		Q3	Ciência da Computação; Sistemas de Informação			
Nature Conservation	0.501	Q2	Ecologia, Evolução, Comportamento e Sistemática	2.1	Q2	Conservação da Biodiversidade	1	M	1.086,00
			Conservação da Natureza e da Paisagem						
BMC research notes	0.486	Q2	Bioquímica, Genética e Biologia Molecular (miscelânea)	NP			5	M	1.690,00
		Q3	Medicina (miscelânea)						
	0.440	Q2		3.3	Q2	Geociências, Multidisciplinar	3	D	2.290,00

Periódico	Scopus			Web of Science			F*	T**	APC (USD)***
	SJR (2023)	Quartil	Subárea	JIF (2023)	Quartil	Áreas			
Geoscience data journal			Ciências da Terra e Planetárias (miscelânea)			Meteorologia e Ciências Atmosféricas			
Biodiversity data journal	0.367	Q2	Ciências Agrárias e Biológicas (miscelânea)	1.0	Q3	Conservação da Biodiversidade	26	D	865,28
			Ciência Animal e Zoologia						
			Botânica						
			Ecologia						
		Q3	Ecologia, Evolução, Comportamento e Sistemática						
			Entomologia						
Cahiers Agricultures	0.366	Q2	Agronomia e Ciência das Culturas	1.0	Q3	Agricultura, Multidisciplinar	1	M	-
			Ciência Animal e Zoologia			Agronomia			
		Q3	Gestão, Monitoramento, Política e Direito						
Comparative cytogenetics	0.296	Q3	Ciência Animal e Zoologia	1.0	Q3	Botânica	1	M	378,56
			Entomologia			Zoologia			
			Botânica						
			Biotecnologia						
		Q4	Genética		Q4	Genética e Hereditariedade			
Chemical data collections	0.270	Q3	Química (miscelânea)	NP			22	D	1.330,00
Microbiology resource announcements	0.266	Q3	Imunologia e Microbiologia (miscelânea)	0.7	Q4	Microbiologia	2	M	1.400,00
		Q4	Genética						
			Biologia molecular						
Ecosistemas	0.238	Q3	Ecologia	0.8	Q1	Ecologia	1	M	-
		Q4	Ecologia, Evolução, Comportamento e Sistemática						
Data in brief	0.208	Q3	Multidisciplinar	1.0	Q3	Ciências Multidisciplinares	315	D	1.010,00
			Educação						
Journal of open humanities data	0.112	Q4	Artes e Humanidades (miscelânea)	NP			1	D	646,15
			Sistemas de Informação						
			Biblioteconomia e Ciência da Informação						

Periódico	Scopus			Web of Science			F*	T**	APC (USD)***
	SJR (2023)	Quartil	Subárea	JIF (2023)	Quartil	Áreas			
Frontiers in bioinformatics			NP	2.8	Q2	Matemática & Biologia Computacional	1	M	512,00

Fonte: elaborado pela autora

Notas: F*: frequência refere-se à quantidade de *data papers* publicados nos periódicos por autores filiados às instituições brasileiras.

T** tipo: tipo de periódico

Valor da APC***: quando necessário os valores foram convertidos para dólares americanos (USD), em 07 de março de 2025, no Conversor de Moedas do Banco Central do Brasil, disponível em <https://www.bcb.gov.br/conversao>

APÊNDICE G – LISTA DOS MANTENEDORES DOS REPOSITÓRIOS IDENTIFICADOS

Quadro 20 - Mantenedores dos repositórios

N	Repositório	Mantenedor(es)	Tipo de Instituição
1	Mendeley Data	Elsevier Inc.; Mendeley Inc.	C
2	Zenodo	European Commission; European Commission, Research & Innovation; European Organization for Nuclear Research; OpenAIRE	SL
3	Figshare	Digital Science	C
4	Sequence Read Archive	National Institutes of Health	SL
5	Genbank	International Nucleotide Sequence Database Collaboration; National Center for Biotechnology Information; National Institutes of Health; U. S. National Library of Medicine; U.S. Department of Health & Human Services	SL
6	Pangaea	Alfred Wegener Institute for Polar and Marine Research; University of Bremen	SL
7	Lancaster University Research Directory	Lancaster University	SL
8	Open Science Framework	Center for Open Science	SL
9	Global Biodiversity Information Facility	Global Biodiversity Information Facility	SL
10	Github	GitHub, Inc.	C
11	BioProject	National Center for Biotechnology Information (NCBI)	SL
12	Proteomics Identifications Database	Biotechnology and Biological Sciences Research Council; European Commission, Research & Innovation; European Bioinformatics Institute	SL
13	Gene Expression Omnibus	National Center for Biotechnology Information; National Institutes of Health	SL
14	Dryad	California Digital Library; Dryad; National Science Foundation	SL
15	European Nucleotide Archive	Biotechnology and Biological Sciences Research Council; European Commission, Research & Innovation; The European Bioinformatics Institute; International Nucleotide Sequence Database Collaboration; Wellcome Trust Genome Campus	SL
16	Mass Spectrometry Interactive Virtual Environment (Massive)	National Institutes of Health; University of California at San Diego; University of California	SL
17	Cambridge Crystallographic Data Center	Cambridge Crystallographic Data Centre, Inc.	SL
18	Science Data Bank	Chinese Academy of Sciences	SL
19	GigaScience Database	Beijing Genomics Institute; China National Genebank; GigaScience Press	C; SL
20	BioSample	National Center for Biotechnology Information (NCBI)	SL
21	Synapse	Alfred P. Sloan Foundation; Children's Tumor Foundation; Sage Bionetworks; National Institutes of Health	SL
22	Dados Abertos do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	CKAN	SL
23	Plasmodium Genomics Resource	National Institutes of Health	SL
24	Atmospheric Science Data Center (ASDC)	NASA Langley Research Center; NASA's Earth Observing System Data and Information System	SL
25	Data repository	German Centre for Integrative Biodiversity Research (iDiv)	SL
26	National Centers for Environmental Information	DataONE; NOAA National Environmental Satellite, Data, and Information Service; National Oceanic and Atmospheric Administration; United States Department of Commerce	SL
27	ArrayExpress	European Molecular Biology Laboratory Heidelberg; European Commission; European Bioinformatics Institute; National Institutes of Health; National Science Foundation	SL
28	Harvard DataVerse	Harvard University	SL
29	Map of Life (MOL)	Yale University	SL
30	RCSB Protein Data Bank	National Institutes of Health; National Science Foundation; Research Collaboratory for Structural Bioinformatics; Rutgers; U.S. Department of Energy, Office of Science; UC San Diego	SL
31	Sea Scientific Open Data Publication	Ilfremer	SL

N	Repositório	Mantenedor(es)	Tipo de Instituição
32	MetaboLights	Biotechnology and Biological Sciences Research Council; The European Bioinformatics Institute; University of Cambridge	SL
33	Environmental Data Initiative Repository	DataONE; Environmental Data Initiative; Long Term Ecological Research Committees and Groups; National Science Foundation; University of New Mexico at Albuquerque; University of Wisconsin	SL
34	Herbarium of Federal University of Mato Grosso do Sul	Global Biodiversity Information Facility	SL
35	NCBI DataSets	National Center for Biotechnology Information; National Institutes of Health; U.S. Department of Health and Human Services; USAGov	SL
36	Immunology Database and Analysis Portal	National Institute of Allergy and Infectious Diseases; Northrop Grumman Health Solutions; United States Department of Health and Human Services; University of California San Francisco	C; SL
37	Ag Data Commons	United States Department of Agriculture, National Agricultural Library	SL
38	Archive of Data on Disability to Enable Policy	National Institutes of Health	SL
39	British Antarctic Survey	British Antarctic Survey; Natural Environment Research Council	SL
40	British Oceanographic Data Centre	British Oceanographic Data Centre; National Oceanography Centre, Liverpool Site	SL
41	Cirad dataverse	CIRAD; The Dataverse Project	SL
42	Dagshub	Empresa	C
43	DataSuds	Institut de Recherche pour le Développement	SL
44	DataverseNL	Data Archiving and Networked Services; Dutch Research Council; Royal Netherlands Academy of Arts and Sciences	SL
45	Digital CSIC	Ministry of Science, Innovation and Universities; Spanish National Research Council	SL
46	Digital Rocks Portal	National Science Foundation; University of Texas at Austin	SL
47	Fairdata IDA Research Data Storage Service	Academy of Finland; IT Center for Science Ltd; Ministry of Education and Culture	SL
48	Freshwater Research and Environmental Database	Deutsche Forschungsgemeinschaft; Leibniz-Institute for Freshwater Ecology and Inland Fisheries	SL
49	GFZ Data Service	GFZ German Research Centre for Geosciences	SL
50	Global freshwater biodiversity research	Belgian Biodiversity Platform; Biofresh Project; European Commission, Research & Innovation; Freshwater Information Platform; Global Biodiversity Information Facility; Leibniz Institute of Freshwater Ecology and Inland Fisheries; Royal Belgian Institute of Natural Sciences; University of Duisburg-Essen; University of Natural Resources and Life Sciences	SL
51	Global Health Data Exchange	Bill & Melinda Gates foundation; University of Washington	SL
52	ICOS Carbon Portal	Integrated Carbon Observation System; Lund University; Wageningen University & Research	SL
53	International Neuroimaging Data-Sharing Initiative	Child Mind Institute; Neuroimaging Informatics Tools and Resources Clearinghouse	C; SL
54	IsoArch	IsoArch Association; IsoArch Initiative	SL
55	Materials Cloud repository	National Centre of Competence in Research MARVEL; Swiss National Supercomputing Centre; École polytechnique fédérale de Lausanne;	SL
56	Metagenomics analysis server	Argonne National Laboratory; The University of Chicago	SL
57	NoMaD Repository & Archive	Humboldt-Universität zu Berlin; Max-Planck-Gesellschaft	SL
58	OTELo Research Data Repository	National Center for Scientific Research; Observatoire Terre Environnement de Lorraine; University of Lorraine	SL
59	ProteomeXchange	The European Bioinformatics Institute; European Commission, Community Research and Development Information Service; ProteomeXchange Consortium	SL
60	Repositório da UNESP	Universidade Estadual Paulista "Júlio de Mesquita Filho"	SL
61	Repositório da Universidade de São Paulo	Universidade de São Paulo	SL
62	Repositório de Dados de Pesquisa da Fundação Getúlio Vargas	Fundação Getúlio Vargas	SL
63	Repositório de Dados de Pesquisa da Unicamp (REDU)	Universidade de Campinas	SL

N	Repositório	Mantenedor(es)	Tipo de Instituição
64	Repository of Open Access Datasets	Flinders University	SL
65	Saint Mary's University Dataverse	Borealis; Saint Mary's University	SL
66	TRY	Climate-Environment-Society; Fondation pour la Recherche sur la Biodiversité; Future Earth; German Centre for Integrative Biodiversity Research; International Geosphere-Biosphere Programme; Max Planck Institute for Biogeochemistry;	SL

Fonte: elaborado pela autora

Notas: Tipo de instituição: SL: sem fins lucrativos; C: comercial; NI: Não foi identificado

ANEXO A – ESQUEMA DE CLASSIFICAÇÃO DA ALL SCIENCE JOURNAL CLASSIFICATION (ASJC)

Quadro 21 - Áreas e subáreas da ASJC

Ciências Físicas	Ciências da Saúde	Ciências Sociais	Ciências da Vida
Engenharia Química Química Ciência da Computação Ciências da Terra e Planetárias Energia Engenharia Ciências ambientais Ciências de materiais Matemática Física e Astronomia Multidisciplinar	Medicina Enfermagem Veterinária Odontologia Profissões da Saúde Multidisciplinar	Artes e Humanidades Negócios, Gestão e Contabilidade Ciências da Decisão Economia, Econometria e Finanças Psicologia Ciências Sociais Multidisciplinar	Ciências Agrárias e Biológicas Bioquímica, Genética e Biologia Molecular Imunologia e Microbiologia Neurociência Farmacologia, Toxicologia e Farmacêutica Multidisciplinar

Fonte: traduzido de Elsevier (c2024a)