



24° ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

**A PUBLICAÇÃO DE DADOS DE PESQUISA: UMA ANÁLISE A PARTIR DOS DATA PAPERS
PUBLICADOS POR AUTORES AFILIADOS À UNIVERSIDADE FEDERAL DE SANTA CATARINA**

***THE PUBLICATION OF RESEARCH DATA: AN ANALYSIS FROM THE DATA PAPERS PUBLISHED
BY AUTHORS AFFILIATED WITH THE FEDERAL UNIVERSITY OF SANTA CATARINA***

Crislaine Zurilda Silveira – Universidade Federal de Santa Catarina (UFSC)

Thiago Magela Rodrigues Dias – Universidade Federal de Santa Catarina (UFSC)

Patricia da Silva Neubert – Universidade Federal de Santa Catarina (UFSC)

Modalidade: Trabalho Completo

Resumo: A publicação de dados de pesquisa em *data papers* vem crescendo como forma de compartilhamento, elevando o status dos dados de pesquisa a uma publicação científica legítima e capaz de ser indexada por bases de dados. No cenário brasileiro, embora haja uma tímida presença de *data papers* nos periódicos nacionais, os pesquisadores vêm o adotando. O objetivo deste trabalho é identificar como está ocorrendo a publicação de dados de pesquisa pelos autores afiliados à Universidade Federal de Santa Catarina. Os objetivos específicos são: a) identificar a autoria por meio da afiliação institucional, países e fontes de financiamento; b) caracterizar os periódicos nos quais os *data papers* foram publicados e c) analisar os repositórios de dados de pesquisa. Foi realizada uma pesquisa de caráter exploratório e descritivo, cujos procedimentos adotados foram a pesquisa bibliográfica e documental. Os resultados permitiram identificar que os *data papers* publicados por coautores vinculados predominantemente a instituições de ensino superior internacionais, e que receberam financiamentos de agências públicas nacionais e internacionais. Os *data journals* em sua maioria, são das áreas de Tecnologia, Ciências da Vida e Biomedicina, publicados por *publishers* comerciais, concentrados no continente europeu. Os repositórios identificados são generalistas e mantidos por parcerias entre universidades e centros de pesquisa, cuja gratuidade é limitada. Com isso, a consolidação da publicação de *data papers* e do depósito de dados ainda carece de padronização e orientação quanto ao depósito nos repositórios de dados. Além disso, é preciso estar atento a inserção dos *publishers* comerciais publicação dos dados de pesquisa.

Palavras-chave: *data journals*; repositórios de dados de pesquisa; artigo de dados.

Abstract: the publication of research data in data papers has been growing as a form of sharing, elevating the status of research data to a legitimate scientific publication capable of being indexed by databases. In the Brazilian scenario, although there is a timid presence of data papers in national journals, researchers have been adopting them. The objective of this work is to identify how the publication of research data by authors affiliated with the Federal University of Santa Catarina is occurring. The specific objectives are: a) identify authorship through institutional affiliation, countries

and funding sources; b) characterize the journals in which the data papers were published and c) analyze the research data repositories. An exploratory and descriptive research was carried out, the procedures adopted were bibliographic and documentary research. The results made it possible to identify that the data papers published by co-authors were predominantly linked to international higher education institutions, and that received funding from national and international public agencies. The majority of data journals are in the areas of Technology, Life Sciences and Biomedicine, published by commercial publishers, concentrated on the European continent. The identified repositories are generalist and maintained by partnerships between universities and research centers, which are limited free of charge. As a result, the consolidation of the publication of data papers and the deposit of data still lacks standardization and guidance regarding the deposit in data repositories. Furthermore, it is necessary to pay attention to the inclusion of commercial publishers publishing research data.

Keywords: data journals; research data repositories; data article.

1 INTRODUÇÃO

O compartilhamento dos dados de pesquisa vem emergindo, devido a uma pressão crescente sobre os autores por meio de mandatos de financiadores e políticas de periódicos (Thelwall, 2020). No entanto, esses mandatos raramente oferecem incentivos substanciais para que os autores disponibilizem seus dados abertamente, com isso, além da falta de reconhecimentos, os autores relatam as dificuldades relacionadas ao trabalho para coletar e tratar os dados, restrições de tempo e financiamento, o esforço pendido compartilhar os dados de forma compreensível e adequada, como resistências ao não compartilhamento dos dados de pesquisa (Blanc; Ngo, 2019; Thelwall, 2020; Walters, 2020).

Visando diminuir essas resistências, a publicação de dados surge como um novo modelo de compartilhamento que garante a qualidade dos dados por meio de alguns procedimentos básicos consolidados na comunicação científica, que inclui a submissão de dados, revisão por pares, compartilhamento e armazenamento permanente de dados, citação e avaliação de impacto dos dados, o que pode maximizar o seu reuso (Kong *et al.*, 2019).

A publicação de *data papers*¹ vêm sendo considerada uma saída para a falta de incentivos e recompensas aos esforços dos pesquisadores para tornar os dados abertamente disponíveis, elevando o status dos dados de pesquisa a uma publicação científica legítima e capaz de ser indexada por bases de dados, possibilitam a descrição minuciosa dos dados, facilitando a verificação, replicação e a reprodutibilidade da pesquisa e aumentam o acesso a diferentes produções associadas ao conjunto de dados, possibilitando mais citações e

¹ A tradução literal é artigo de dados, no entanto, não há uma tradução consolidada em Língua Portuguesa, portanto, neste trabalho optou-se pelo uso do termo em inglês.

oportunidades de colaboração entre pesquisadores com interesses comuns (Curty; Aventurier, 2017; Pampel; Dallmeier-Tiessen, 2014; Walters, 2020).

No Brasil, a publicação de dados ainda é tímida, se dá no âmbito da Ciência da Informação, pela *Revista Encontros Bibli*, que, a partir de 2020, começou a publicar *data papers*. No entanto, apesar da quase ausência de revistas nacionais que publicam esse tipo de documento, há publicações de *data papers*, indexados nas bases de dados Scopus e Web of Science por autores brasileiros.

Com isso, o objetivo deste trabalho é identificar, a partir dos *data papers* indexados em base de dados, como está ocorrendo a publicação de dados de pesquisa pelos autores afiliados à Universidade Federal de Santa Catarina (UFSC). Os objetivos específicos são: a) identificar a autoria por meio da afiliação institucional, países e fontes de financiamento; b) caracterizar os periódicos nos quais os *data papers* foram publicados e c) analisar os repositórios de dados de pesquisa, nos quais os dados dos *data papers* foram publicados.

Por fim concorda-se com Schiessl e Shintaku (2022) que essa temática é abrangente e requer estudos e aplicações, mas apresenta certa inovação na comunicação científica, uma vez que ela causa uma mudança no ecossistema de publicações científicas.

2 DATA PAPERS

Chavan e Penev (2011, p. 1, tradução nossa) definem o *data paper* “[...] como uma publicação acadêmica de um documento de metadados pesquisável que descreve um conjunto de dados acessível *online* específico, ou um grupo de conjuntos de dados, publicado de acordo com as práticas acadêmicas padrão”. Pode-se complementar ainda, que os *data papers*, são revisados por pares, implicando algum tipo de controle de qualidade e seleção, por isso, contribuem para a confiabilidade dos dados de pesquisa (Schöpfel *et al.*, 2019).

Em relação ao conteúdo, os *data papers* são uma publicação cujos objetivos são conter fatos sobre dados e descrevê-los, com isso, destinam-se exclusivamente 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 (Chavan; Penev, 2011; Santos; Nassi-Caló, 2020).

A redação de um *data paper* deve incluir a descrição de vários elementos importantes, por exemplo, informações sobre autor(es), quem coletou e quem possui os dados, natureza e o formato do conjunto de dados, objetivos, contexto, metodologia, processos, data de

produção e condições de uso da pesquisa, os intervalos espaciais e temporais e a cobertura regional dos registros de dados, detalhes de qual software foi usado para criar os dados ou usado para visualizá-los, além discutir sua utilidade potencial para outros pesquisadores ou outros atores (sociedade civil, tomadores de decisão, organizações não governamentais, dentre outros) e como os dados de pesquisa podem ser reutilizados (Chavan; Penev, 2011; Schöpfel *et al.*, 2019; Silva; Stueber; Witt, 2022).

Para Candela *et al.* (2015) e Curty e Aventurier (2017) os *data papers* devem estar sempre atrelados a um conjunto de dados depositados num repositório certificado, pois os repositórios de dados são parte essencial do ecossistema da publicação de dados, e constituem por si só como uma abordagem de *data publishing*, uma vez que tornam públicas coleções de dados acompanhadas por recursos que otimizem seu potencial de reuso.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa tem um caráter exploratório e descritivo, cujo objetivo é conhecer o universo da publicação de dados de pesquisa em periódicos por parte de autores afiliados a UFSC e descrever suas características. O recorte utilizado considera a *expertise* dessa instituição na publicação de *data papers*, uma vez que a revista Encontros Bibli é mantida pelo Programa de Pós-Graduação em Ciência da Informação da UFSC.

Os procedimentos, considerados os passos necessários para levantamento dos dados para a elaboração da pesquisa foram pesquisa bibliográfica e documental (Prodanov; Freitas, 2013). A partir deles as coletas dos dados foram realizadas em regime de complementaridade, na qual a partir da busca em bases de dados, foram localizados os *data papers* e na pesquisa documental foram realizados levantamentos nas políticas editoriais dos periódicos, nos metadados dos repositórios indexados no *Registry of Research Data Repository (R3data)*² e em páginas institucionais.

A coleta dos dados bibliográficos foi realizada em 27 de setembro de 2023 e compreendeu a recuperação de *data papers* publicados por autores afiliados à UFSC, nas bases Scopus e Web of Science. Nessas bases foi aplicado o filtro de tipo de documento que restringiu a busca apenas para “*data papers*”, em seguida foram aplicados os filtros de país “Brazil” e de afiliação institucional “UFSC”, sem limites de data.

² Disponível em: <https://www.re3data.org/>

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Em seguida os metadados dos *data papers* foram exportados formato comma-separated values (.csv) para o software Microsoft Excel 2016, onde foi realizada limpeza dos registros duplicados, que resultou em 35 *data papers*. Em seguida, metadados foram exportados para o software Open Refine³, no qual foi realizada a uniformização de letras maiúsculas e minúsculas, a padronização dos nomes das instituições para a forma completa e no caso de instituições brasileiras, foram realizadas as traduções para o idioma português, quando necessário.

A análise dos dados se deu com o uso de métodos mistos, ou seja, o uso do método quantitativo, cujos passos são sequenciais e comprobatórios e perpassam pela definição dos objetivos e pela medição das variáveis para estabelecer uma série de conclusões (Creswell; Clark, 2013; Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013). E pela análise qualitativa, na qual buscou-se entender o significado dessas variáveis (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013). O tratamento dos dados foi realizado no Microsoft Excel 2016 e as variáveis foram categorizadas conforme a aderência aos objetivos propostos. Por fim, a apresentação dos dados foi realizada através tabelas e figuras elaboradas na versão gratuita do software Visme⁴.

4 RESULTADOS

Os *data papers* foram publicados em periódicos das áreas da Tecnologia, com 51,43% (18), seguidos da área de Ciências da Vida e Biomedicina, com 31,17% (13) (coluna total DP), conforme Tabela 1.

Tabela 1 – Características gerais da autoria

Áreas Web of Science	Quantidade de autores*		Coautoria Nacional ****		Coautoria Internacional *****		Média de autores	Total DP	
	F**	%	F	%	F	%	F	F	%
Ciências da Vida e Biomedicina	2667	80,14%	1	2,86%	12	34,29%	205,15	13	37,14%
Ciências da Vida e Biomedicina; Tecnologia	10	0,30%	-	-	1	2,86%	10	1	2,86%
Ciências Físicas	19	0,57%	1	2,86%	2	5,71%	6,33	3	8,57%
Tecnologia	632	18,99%	6	17,14%	12	34,29%	35,11	18	51,43%
Artes e Humanidades	-***	-	-	-	-	-	-	-	-
Ciências Sociais	-	-	-	-	-	-	-	-	-
	3328	100%	8	22,86%	27	77,14%		35	100%

Fonte: dados da pesquisa.

³ Disponível em: <https://openrefine.org/>

⁴ Disponível em: <https://www.visme.co/pt-br/>

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Nota: *Trata-se do somatório de autores dos *data papers*, ou seja, estão incluídos autores que publicaram um ou mais *data papers*

**F significa a frequência ou o número indicado na coluna

***Os números zeros foram substituídos por um traço (-).

****Foi considerada coautoria nacional, a menção de instituições localizadas no Brasil.

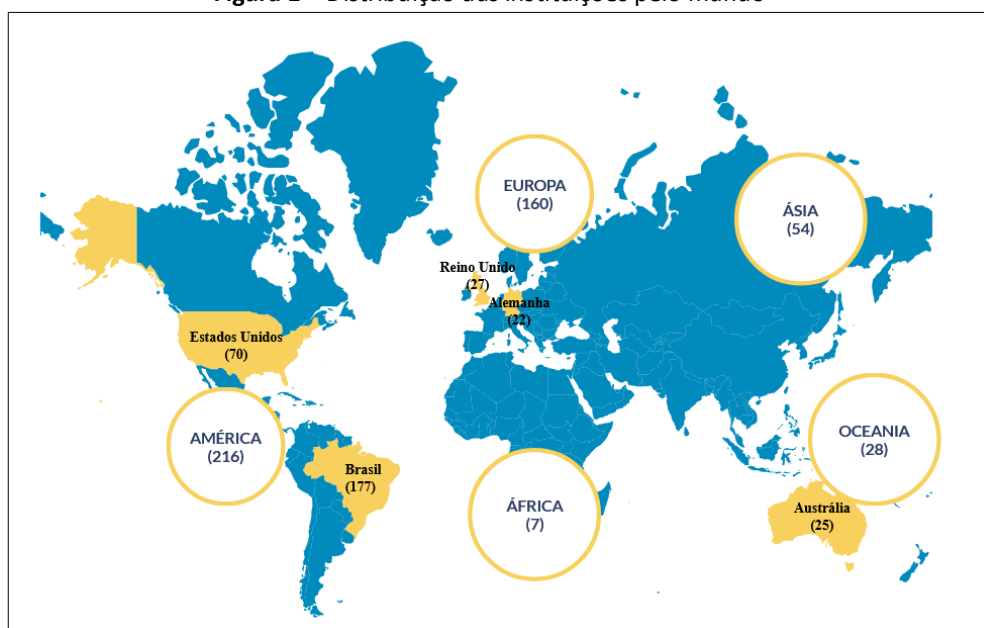
***** Foi considerada coautoria internacional a menções a vinculação de uma ou mais instituições internacionais em parceria com uma instituição nacional.

Em relação as características gerais da autoria, a coluna quantidade de autores refere-se ao número total de autores mencionados nos 35 *data papers*, sendo autores nacionais ou internacionais a depender da afiliação indicada, sendo assim, os *data papers* foram publicados por 3.328 autores, desses 77,14% (27) são autores afiliados a instituições internacionais. Logo pode-se verificar que há uma predominância de coautorias internacionais na publicação de *data papers* por pesquisadores afiliados a UFSC e em periódicos da área da Ciências da Vida e Biomedicina, com a concentração de 80,14% (2667) autores.

A média dos autores foi realizada a partir da quantidade de autores identificadas em cada área dividida pelo total de *data papers* publicados naquela área. Logo, a maior média de autores foi na área de Ciências da Vida e Biomedicina, com 205,15 autores por *data paper*.

Em relação a afiliação, que se refere às instituições informadas pelos autores e pelo tratamento dado pelas bases de dados. Os autores identificados estavam afiliados a 579 instituições diferentes, localizadas em 76 países, distribuídas em 5 continentes, conforme Figura 1.

Figura 1 – Distribuição das instituições pelo mundo



Fonte: elaborado pelos autores

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Os cinco países, que mais as concentraram foram: Brasil (177), Estados Unidos (70), Reino Unido (27), Austrália (25) e Alemanha (22). Com exceção da Austrália, o destaque desses quatro países na coautoria científica com o Brasil, coaduna com os resultados descritos no Relatório da Unesco, publicado em 2021, o qual considera os artigos científicos do período entre 2017 a 2019 (Schneegans; Lewis; Straza, 2021).

Em relação ao tipo de instituições 75,47% (437) são instituições de ensino superior, 10,02% (58) são institutos de pesquisa, 5,01% (29) outras instituições, 4,84% (28) são as organizações não-governamentais (ONGs) e 4,66% (27) são órgãos do governo.

As instituições internacionais, cujos autores publicaram em coautoria com autores afiliados a UFSC, foram: a Universidad Nacional Autónoma de México, 0,89% (13), o Consejo Nacional de Investigaciones Científicas y Técnicas, da Argentina, 0,83% (12), a University of London 0,62% (9), a University of California System, dos Estados Unidos, com 0,48% (7) e o Centre National de la Recherche Scientifique, da França com 0,48% (7)

As instituições brasileiras, as quais representam 30,57% (177) do total de afiliações identificadas, 38,60% (66) estava localizada na região Sudeste, 21,64% (37) na região Sul, 18,13% (31) na região Nordeste, 12,87% (22) na região Centro Oeste e 8,77% (15) na região norte⁵. Essas são predominantemente instituições de ensino superior, 66,08% (113), sendo que o mesmo se refletiu a todas as regiões do país. Os institutos de pesquisa representam 9,94% (17) das afiliações no país, 11,70% (20) outros, 6,43% (11) órgãos do governo e 5,85% (10) ONGs.

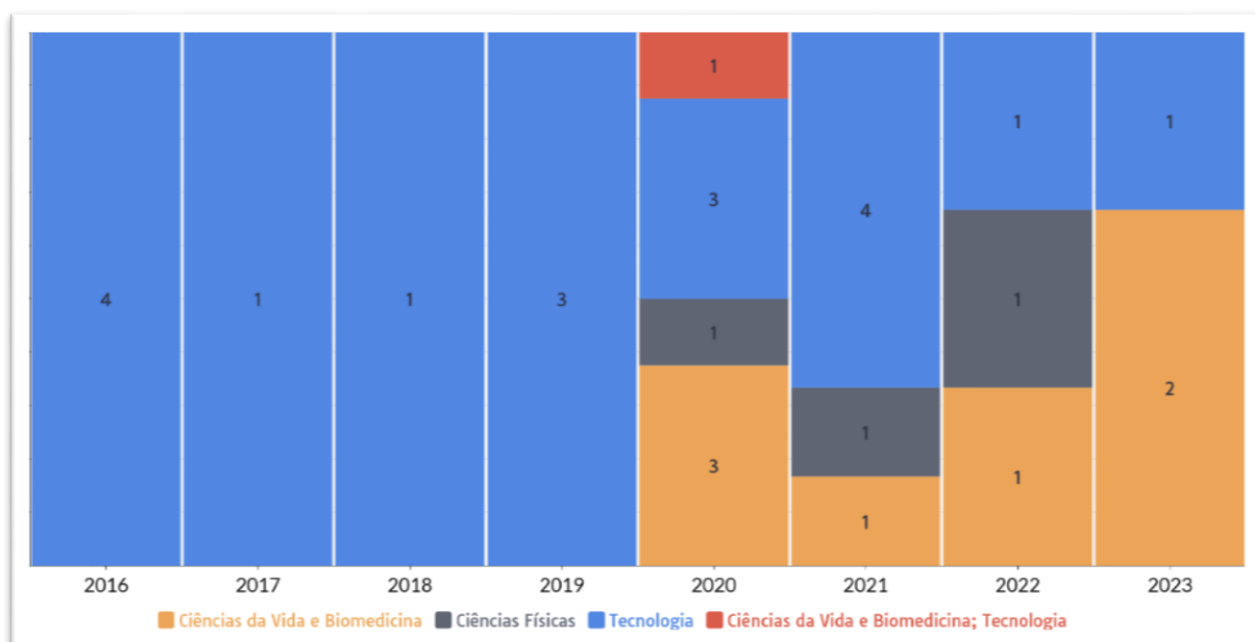
As cinco instituições brasileiras que mais publicaram em coautoria com autores a afiliados a UFSC são autores afiliados a Universidade de São Paulo, 3,58% (52), autores da própria UFSC, 3,30% (48), da Universidade Estadual Paulista Júlio de Mesquita Filho, 3,16% (46), da Universidade Federal do Paraná, 1,79% (26) e da Universidade Federal de São Carlos 1,65% (24).

4.1 Características dos financiamentos

Dos 35 *data papers* da amostra, 82,86% (29) receberam algum tipo de financiamento e 17,14% (6) não receberam. Desses, houve uma concentração de financiamentos no ano de 2020, na área de Tecnologia, conforme Figura 2.

⁵ Não foi possível identificar o estado de seis instituições, por isso o somatório de instituições brasileiras dispersas pelo Brasil, soma 171.

Figura 2 – Comparação dos financiamentos por área



Fonte: elaborado pelos autores

Foram identificadas 118 agências de financiamentos diferentes, distribuídas entre 27 países, sendo que aqueles que mais concentraram agências foram: Brasil (38), Estados Unidos (24), Alemanha, Austrália e Reino Unido com 7 cada um. Em relação do tipo de agência, 68,64% (81) são públicas, seguidas pelas ONGs com 16,95% (20) e 14,41% (17) são agências privadas de financiamento. O Brasil, segue a tendência mundial, na qual os financiamentos são concentrados em agências públicas, sendo 83,78% (31), seguido das privadas com 13,51% (5), das ONGs com 2,70% (1) e uma instituição não foi identificada sua tipologia e nem localização.

Em suma, os financiamentos foram realizados, em sua maioria, 51,74% (104), por agências internacionais, com destaque para a National Science Foundation (NSF) com 3,48% (7) e 48,26% (97) dos financiamentos foram realizados por agências nacionais, com destaque para o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com 8,96% (18) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), com 5,47% (11).

4.2 Características dos periódicos

As publicações de *data papers* pelos autores afiliados a UFSC representam 0,28% (35) de todos os *data papers* indexados e recuperados nas bases Scopus e Web of Science, estando concentrados em sete periódicos, conforme tabela 2

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

Tabela 2 – Características dos periódicos

Nome do periódico	Área WOS	Publisher (País)	APC (USD)**	Estimativa de gastos (USD)	Ano Fund.	Ano publ. dp	F	%
Data in Brief	Tecnologia	Elsevier (NLD)	840,00	10.920,00	2014	2015	13	37,14%
Ecology	Ciências da Vida e Biomedicina	Wiley (USA)	3.283,12	36.114,32	1920	2017	11	31,43%
Scientific Data	Tecnologia	Nature (ENG)	2.068,36	10.341,80	2014	2014	5	14,29%
Earth System Science Data	Ciências Físicas	Copernicus Gesellschaft MBH (DEU)	1.532,12	3.064,24	2009	2023	2	5,71%
Biodiversity Data Journal	Ciências da Vida e Biomedicina	Pensoft Publishers (BGR)	711,34	1.422,68	2013	2013	2	5,71%
Chemical Data Collections	Ciências Físicas	Elsevier (NLD)	1.030,00	1.030,00	2016	2016	1	2,86%
BMC Research Notes	Ciências da Vida e Biomedicina; Tecnologia	Springer Nature (ENG)	1.521,18	1.521,18	2008	2018	1	2,86%
Total geral				64.414,22			35	100%

Fonte: dados da pesquisa

Notas: *Abreviação de países conforme ISO 3166-2:2020

**Nesta amostra os valores se referem tanto para artigos completos quanto para data papers

O periódico *Data in Brief* concentrou 37,14% (13) do total de *data papers* publicados, é também o periódico com mais *data papers* publicados em âmbito mundial, com 7.107. Este destaque no número de publicações pode estar relacionado ao caráter multidisciplinar da sua cobertura de assuntos e/ou ao valor da APC, uma vez que a taxa cobrada é a mais baixa dentre os periódicos identificados, US\$ 840.

Candela *et al.* (2015) criou uma classificação dos periódicos para dividir entre puros e mistos, sendo que os primeiros seriam 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. Atualmente, pode-se considerar que 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*, conforme pôde ser observado nos periódicos identificados nesta amostra. Logo, o termo “puro” pode ser substituído por *data journals*⁶, cujo enfoque é publicar dados e não exclusivamente *data*

⁶ A tradução literal é periódico de dados, no entanto, não há uma tradução consolidada em Língua Portuguesa, portanto, neste trabalho optou-se pelo uso do termo em inglês.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

papers. Com isso, nesta amostra, apenas os periódicos Ecology e o BMC Research Notes são periódicos mistos, todos os demais são *data journals*.

A adesão a publicação de *data papers* pelos periódicos, ocorre anos após a sua, como ocorre com Ecology, fundado em 1920 e publicou o primeiro *data paper* em 2017 e o BMC Research Notes fundado em 2008 e publicou seu primeiro *data paper* em 2018. Enquanto que, os *data journals*, começaram a publicar *data papers* em geral um ano após sua fundação.

Em relação aos *publishers* (editores dos periódicos) identificados, todos são editores comerciais, cuja taxa de APC varia de US\$ 840,00 a US\$ 3.283,12, em sua maioria localizados em países do continente europeu, tais como: Holanda (NLD), Inglaterra (ENG), Alemanha (DEU) e Bulgária (BGR).

A partir do valor da APC multiplicado pela quantidade de *data papers* publicados por cada periódico, estima-se que houve um investimento de US\$ 64.414,22 na publicação desse tipo de documento, desse total, estima-se que US\$ 37.537,00, um pouco mais da metade, foram investidos em periódicos da área das Ciências da Vida e Biomedicina, US\$ 21.261,80 da Tecnologia, US\$ 4.094,24 das Ciências Física e US\$ 1521,18 das áreas comuns de Ciências da Vida e Biomedicina e Tecnologia.

Quase todos os periódicos passam pela revisão por pares simples cego, que significa os editores e pareceristas tem informações sobre a identidade e afiliação do autor, mas o autor não (Santos; Nassi-Caló, 2020). As exceções são o *Scientific Data* que não deixa claro o tipo de revisão, o *Biodiversity Data Journal* no qual o autor pode escolher a abertura ou não dos pareceres e no caso do periódico *Earth System Science Data* a revisão por pares é chamada de *Interactive review process*, o qual envolve uma etapa discussão com a participação dos revisores e membros da comunidade científica, ficando sob responsabilidade dos autores a condução das discussões e a resposta os questionamentos dos revisores e da comunidade, e, por fim, a decisão final pela publicação ou não do manuscrito fica sob responsabilidade do editor, que considera as discussões (Copernicus Publications, [2024a]).

Dentre outros aspectos, a revisão por pares, tem como objetivo verificar o depósito dos dados de pesquisa, sendo assim, todos os periódicos identificados na amostra descreveram em seus sites orientações para a publicação dos conjuntos de dados de pesquisa em repositórios de dados preferencialmente temáticos.

No entanto, a efetivação da publicação dos dados em repositórios só ocorreu a partir de 2018, anteriormente todos os dados de pesquisa foram inseridos material suplementar, o

que corresponde a 11,43% (4). A partir de 2020, percebe-se uma tendência a adesão a publicação em repositórios, com isso no período de 2018 a 2023, a publicação dos dados em repositórios, representou 54,3% (19) das publicações, enquanto que 45,7% (16) foi publicado como material suplementar.

Candela *et al.* (2015) observou que essa característica de anexar os dados com material suplementar vem sendo progressivamente descontinuada em favor da publicação dos conjuntos de dados num repositório apropriado. A desvantagem relacionada a publicação como material suplementar, é que os dados de pesquisa requerem curadoria e preservação dos arquivos, além disso, não permite que os leitores para encontrem o link dos dados independentemente da publicação principal (Candela *et al.*, 2015). Portanto, o depósito em repositório de dados é mais vantajoso, uma vez que a atribuição de identificadores persistentes e a descrição dos metadados torna os conjuntos de dados únicos e exclusivos e sem barreiras para a identificação por parte de possíveis reutilizadores.

4.3 Caracterização dos repositórios

Visando mapear a efetividade das recomendações fornecidas pelos periódicos em relação ao depósito de dados em repositórios, na Tabela 3, foram identificadas algumas de suas características.

Tabela 3 – Características dos repositórios

Repositório	F	%	Mantenedor (País)	Cobrança
Zenodo	8	22,9%	CERN Data Centre (NI*)	Gratuito até 50GB
Figshare	3	8,6%	Pessoas físicas (ROU; USA; GBR)	Gratuito até 20GB
Mendeley Data	3	8,6%	Elsevier (GBR)	Gratuito até 10GB
Australian Ocean Data Network (AODN)	1	2,9%	National Collaborative Research Infrastructure Strategy (NCRIS); University of Tasmania (AUS)	Não
Dryad	1	2,9%	Centro Nacional de Análise e Síntese Ecológica (NCEAS) e Centro Nacional de Síntese Evolucionária (NESCent) (USA)	US\$ 150
Git-Hub	1	2,9%	Microsoft Corporation (USA)	Gratuito até 2GB
Global Biodiversity Information Facility (GBIF)	1	2,9%	Universidade de Aarhus (DNK)	NI**
Open Science Framework (OSF)	1	2,9%	Center for Open Science (GBR)	Gratuito
TOTAL	19	54,3%		

Fonte: dados da pesquisa

Notas: * Informação retirada do R3data

**NI: não identificado

*** Os números zeros foram substituídos por um traço

Foram identificados oito repositórios diferentes. Desses, apenas o *Australian Ocean Data Network* (AODN), com 5,26% (1), especializado em dados marítimos e climáticos e o

Global Biodiversity Information Facility (GBIF), com 5,26% (1) especializado em dados sobre a biodiversidade, são repositórios temáticos. A grande maioria dos repositórios identificados 89,48% (17) são repositórios generalistas.

Com isso, ressalta-se que os periódicos, apesar de trazerem nas suas diretrizes orientações que priorizam o depósito dos dados em repositórios aderentes às áreas do conhecimento, não parecem verificar de fato, no momento da avaliação por pares se o depósito está acontecendo conforme suas diretrizes, visto que apenas dois conjuntos dos 19, foram publicados em repositório temático, sendo um publicado na revista *Scientific Data* (SCD) e um publicado na revista *Biodiversity Data Journal* (BDJ).

O depósito prioritariamente vem ocorrendo em repositórios vinculados a centros de pesquisa e/ou mantidos em parceria com universidades, sendo que cinco repositórios que juntos concentram 12 conjuntos de dados depositados. Dois repositórios são mantidos, por empresas privadas, sendo o Mendeley Data mantido pela Elsevier e o Git-Hub mantido pela Microsoft Corporation, respectivamente eles concentram 3 e 1 conjunto de dados. E três conjuntos de dados foram depositados no Figshare, repositório mantido por pessoas físicas. Esses mantenedores estão localizados em países do continente europeu tais como: Romênia (ROU), Reino Unido (GBR) e Dinamarca (DNK); no continente americano, os Estados Unidos (USA) e na Oceania, na Austrália (AUS).

Em relação a cobrança, em geral, não há cobrança para o depósito dos dados, mas os repositórios limitam o tamanho, isto é, quando esse limite é excedido, é realizada cobrança. Logo, é preciso frisar que a publicação do conjunto de dados, pode vir a ter dois custos, o custo pelo depósito num repositório e o custo pela publicação do *data paper* no periódico, portanto, é preciso que o autor considere esses custos ao decidir publicar (Candela *et al.*, 2015).

Embora o repositório Figshare, mantido por um grupo de pessoas físicas e o Mendeley Data, mantido pela Elsevier, uma editora comercial, oferecem maiores espaços para o depósito de dados, o depósito dos dados nesses repositórios é uma decisão que precisa cautela, pois futuramente, esses repositórios podem vir a cobrar pelos acessos aos dados, uma vez que eles já garantiram uma dependência dos dados depositados.

Em relação a revisão dos dados, com exceção dos repositórios AODN, do Dryad, do Mendeley Data, todos os demais não realizam a revisão dos dados de pesquisa. No entanto, essa revisão está muito mais relacionada ao formato dos dados e com a presença dos documentos que os acompanham do que com a qualidade da sua curadoria.

5 CONSIDERAÇÕES FINAIS

A publicação de dados de pesquisa por pesquisadores afiliados a UFSC, está ocorrendo em coautoria de pesquisadores vinculados a instituições de ensino superior predominantemente internacionais, cujas pesquisas estão sendo financiados por agências públicas e a publicação dos *data papers* vem se dando em *data journals* comerciais e o depósito dos dados realizado em repositórios generalistas cuja gratuidade é limitada.

Com isso, pode-se verificar que a publicação de *data papers* vem sendo uma saída adotada pelos autores para o cumprimento dos mandatos das agências de financiamento e o recebimento de recompensas, pois estes uma vez considerados como parte do ecossistema de publicação científica, são indexados em bases de dados podendo ser recuperados e citados, fornecendo visibilidade para os dados originais depositados nos repositórios.

No entanto, a consolidação da publicação de *data papers* e do depósito de dados, como integrantes desse ecossistema ainda carece de padronização e orientação quanto ao depósito nos repositórios de dados, no momento da avaliação por pares, principalmente. Além disso, é preciso estar atento a inserção dos *publishers* comerciais publicação dos dados de pesquisa.

Para pesquisas futuras sugere-se a ampliação dessa coleta de dados para todos os pesquisadores brasileiros de maneira a ter um panorama mais preciso, para que se possa pensar em estratégias de financiamento, políticas e estruturas para a publicação de dados, políticas para a avaliação desses canais e a criação de periódicos brasileiros especializados, e por parte dos autores, pensar em maneiras de realizar a curadoria dos dados bem como a reserva de recursos para a sua publicação.

REFERÊNCIAS

BLANC, A. K.; NGO, T. D. Data papers. **Studies in Family Planning**, Hoboken, p. 1-2, 2019.

CANDELA, L.; CASTELLI, D.; MANGHI, P.; TANI, A. Data journals: A survey. **Journal of the Association for Information Science and Technology**, Nager, 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.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

CHAVAN, V.; PENEV, L. The *data paper*: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, London, 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.

COPERNICUS PUBLICATION. **Process of peer review and publication**. [Alemanha], [2024a]. Disponível em: https://www.earth-system-science-data.net/peer_review/interactive_review_process.html. Acesso em: 06 jan. 2024.

CRESWELL, J. W.; CLARK, V. L. P. **Pesquisa de métodos mistos**. 2. ed. Porto Alegre: Penso, 2013.

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: 01 jul. 2023.

HERNÁNDEZ SAMPIERI, R.; FERNÁNDEZ COLLADO C., BAPTISTA LUCIO. P. **Metodologia de pesquisa**. 5.ed. São Paulo: McGraw-Hill, 2013.

KONG, L.; XI, Y.; LANG, Y.; WANG, Y.; ZHANG, Q. A data quality evaluation index for data journals. *In*: LI, J.; MENG, X.; ZHANG, Y.; CUI, W.; DU, Z. (ed.) **Big Scientific Data Management**. Switzerland: Springer, 2019. p. 291-300.

PAMPEL, H.; DALLMEIER-TIESSSEN, 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: 07 dez. 2023.

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.

SANTOS, S. M.; NASSI-CALÓ, L. Gestão editorial: tendências e desafios na transição para a ciência aberta. *In*: SILVEIRA, L.; SILVA, F. C. C. (Org.). **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.

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.

SCHNEEGANS, S.; LEWIS, J.; T. STRAZA (ed.). **Relatório de Ciências da UNESCO**: A corrida contra o tempo por um desenvolvimento mais inteligente – Resumo executivo. Paris: UNESCO Publishing, 2021

SCHÖPFEL, J.; FARACE, D. J.; PROST, H.; ZANE, A. 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.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Disponível em: <https://shs.hal.science/halshs-02284548/document>. Acesso em: 22 jun. 2023.

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.

THELWALL, M. Data in Brief: Can a mega-journal for data be useful? **Scientometrics**, Massachusetts, v. 124, n. 1, p. 697-709, 2020. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03437-1>. Acesso em: 30 maio de 2024.

WALTERS, W. H. Data journals: incentivizing data access and documentation within the scholarly communication system. **Insights: the UKSG journal**, Massachusetts, v.33, n.1, p. 1-18, 2020. Disponível em: <https://insights.uksg.org/articles/10.1629/uksg.510>. Acesso em: 11 jul. 2022.