

## XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

### GT 7 – PRODUÇÃO E COMUNICAÇÃO DA INFORMAÇÃO EM CIÊNCIA, TECNOLOGIA & INOVAÇÃO

#### A DISPONIBILIDADE DOS DADOS DE PESQUISA EM PERIÓDICOS CIENTÍFICOS: UMA ANÁLISE A PARTIR DOS DATA PAPERS PUBLICADOS POR PESQUISADORES BRASILEIROS

#### *THE AVAILABILITY OF RESEARCH DATA IN SCIENTIFIC JOURNALS: AN ANALYSIS BASED ON DATA PAPERS PUBLISHED BY BRAZILIAN RESEARCHERS*

**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:** Os *data papers* fornecem descrições abrangentes de conjuntos de dados, portanto, devem sempre ter um conjunto de dados associado. Por isso, é importante entender como funciona o compartilhamento dos dados brutos aos quais eles se referem. Esse trabalho tem por objetivo verificar se a disponibilidade dos conjuntos de dados de pesquisa associados aos *data papers* está condizente com as orientações dos periódicos. Para tanto, foi realizada a identificação de como são publicados os conjuntos de dados e verificado se o compartilhamento destes dados segue as diretrizes específicas do periódico quanto ao compartilhamento em repositórios de dados. Foi realizada uma pesquisa bibliográfica cujo objetivo foi identificar os *data papers* com pelo menos um autor filiado às instituições brasileiras, nas bases de dados Scopus e Web of Science. Em seguida foram realizadas duas pesquisas documentais, sendo a primeira nos 634 *data papers* identificados na coleta anterior para identificar o tipo de disponibilidade dos dados de pesquisa e a segunda nas políticas dos periódicos onde os *data papers* foram publicados para verificar se as diretrizes foram seguidas. Embora muitos periódicos exijam a disponibilidade de dados em repositórios, cerca da metade dos *data papers* o fez dessa maneira, a identificação de diversas abordagens de compartilhamento de dados, indica uma falta de padronização. A publicação de *data papers* por autores brasileiros ainda carece de maior aderência às políticas editoriais relacionadas ao compartilhamento de dados, particularmente no que tange à escolha adequada dos repositórios.

**Palavras-chave:** Ciência Aberta; data papers; publicação de dados de pesquisa; repositórios de dados.

**Abstract:** Data papers provide comprehensive descriptions of datasets and must always be associated with a dataset. Therefore, it is essential to understand how raw data sharing works. This study aims to verify whether the availability of research datasets associated with data papers aligns with journal guidelines. To achieve this, the publication methods of datasets were identified, and compliance with the specific journal policies regarding repository sharing was analyzed. A bibliographic search was conducted to identify data papers with at least one author affiliated with Brazilian institutions, using the Scopus and Web of Science databases. Subsequently, two documentary analyses were carried out: the first involved examining 634 data papers identified in the previous step to determine the type of data availability, and the second assessed the journal policies to verify adherence to sharing guidelines.

Although many journals require data to be stored in repositories, only about half of the data papers followed this approach. The identification of various data-sharing approaches highlights a lack of standardization. The publication of data papers by Brazilian authors still lacks greater adherence to editorial policies regarding data sharing, particularly concerning the appropriate choice of repositories.

**Keywords:** Open Science; data papers; research data publication; research data repository.

## 1 INTRODUÇÃO

Os *data papers* são publicações científicas cuja principal função é descrever conjuntos de dados, assegurando sua preservação e reusabilidade, sem necessariamente uma análise interpretativa dos dados (Chavan; Penev, 2011).

Considerando que o objetivo central dos *data papers* é dar visibilidade aos conjuntos de dados de pesquisa, é mister a necessidade de entender como funciona o compartilhamento dos dados aos quais eles se referem e verificar se os autores seguem as recomendações dos periódicos quanto à sua disponibilização.

Visando compreender de que maneira os dados de pesquisas são compartilhados por meio dos artigos científicos, Federer *et al.* (2018), fez uma pesquisa no periódico Plos One e identificou diferentes maneiras, as quais são: a) a disponibilização dos dados como material suplementar; b) a disponibilização mediante a solicitação ao autor; c) a disponibilização no próprio artigo e d) de maneira combinada, na qual os dados são disponibilizados em repositórios de dados e como material suplementar.

Jiao, Li e Fang (2022) também examinaram essas diferentes maneiras de compartilhamento, mas enfatizaram a análise nos repositórios de dados de pesquisa. No entanto, as pesquisas desses autores, se limitam a análise das declarações de disponibilidade de dados<sup>1</sup>, disponíveis nos artigos científicos publicados pelo periódico Plos One. Neste trabalho, o objetivo é realizar essa análise a partir de um tipo documental diferente dos artigos científicos, os *data papers*. Com isso, o diferencial deste trabalho é identificar a disponibilidade dos dados de pesquisa, tendo como recorte, aqueles *data papers* publicados por autores filiados às instituições brasileiras.

Portanto, vislumbrando uma possível contribuição para os editores de periódicos, na elaboração das políticas relacionadas à disponibilidade dos dados de pesquisa vinculados às

---

<sup>1</sup> Refere-se a uma declaração na qual os autores explicam como os leitores podem acessar os dados de pesquisa utilizados em determinado artigo (Federer *et al.*, 2018).

publicações, e para os pesquisadores, cuja disponibilidade pode contribuir para a encontrabilidade e reusabilidade dos dados de suas pesquisas e das publicações associadas aos conjuntos de dados, este trabalho tem por objetivo verificar de que maneira os conjuntos de dados de pesquisa associados aos *data papers* estão sendo disponibilizados e se o compartilhamento destes dados segue as diretrizes específicas do periódico.

## 2 A DISPONIBILIDADE DE DADOS EM *DATA PAPERS*

Os *data papers* fornecem descrições abrangentes de conjuntos de dados, vinculam-se sua localização, criando um caminho para o uso dos dados de pesquisa, incentivando assim seu compartilhamento, a reprodutibilidade da pesquisa e aprimorando o seu valor percebido dentro do ecossistema acadêmico, além de oferecer um mecanismo para crédito de citação aos autores (Li; Huang; Jeng, 2025).

Os *data papers* devem sempre ter um conjunto de dados depositado num repositório de dados online e estável, descritos por metadados, pois esses são parte essencial do ecossistema da publicação dos dados de pesquisa, dando visibilidade e possibilitando o reuso. Uma característica desse tipo de publicação, é que os dados devem estar disponíveis publicamente, pois representam uma garantia de acessibilidade e qualidade, além da revisão por pares (Candela *et al.*, 2015; Puerta-Piñero; Pérez-Luque; Rodríguez-Echeverría, 2020).

Pode-se dizer que esse caminho para o uso dos dados é operacionalizado pelas políticas de dados adotadas pelos periódicos científicos. Federer *et al.* (2018) identificou alguns mecanismos de disponibilidade de dados em periódicos, estes variaram desde o fornecimento dos dados mediante solicitação, a disponibilização no artigo e/ou materiais suplementares, até o uso de um repositório de dados. Na primeira abordagem, mediante solicitação, o inconveniente é que muitos pesquisadores não respondem a tais solicitações enviadas por outros ou não estão dispostos a compartilhar (Federer *et al.*, 2018; Federer, 2022). O compartilhamento de dados dentro do artigo é insuficiente porque os dados compartilhados tendem a ser dados resumidos quando analisados, não permitindo o acesso aos dados brutos que foram coletados pelos pesquisadores (Federer *et al.*, 2018).

A desvantagem de disponibilizar os dados de pesquisa como material suplementar, é que esses requerem curadoria e preservação, e o compartilhamento dessa maneira não permite que os leitores encontrem o seu *link* independentemente da publicação principal (Candela *et al.*, 2015).

Com isso, os repositórios de dados, acabam sendo a melhor opção, pois além de serem recomendados pelos editores, fornecem armazenamento a longo prazo e uso de identificadores persistentes, tirando o ônus do pesquisador de manter os links continuamente ativos, de modo que os dados estejam mais propensos a se manterem acessíveis ao longo do tempo, além de fazerem uso de metadados para a melhor descrição dos objetos (Federer *et al.*, 2018; Federer, 2022).

Depositar conjuntos de dados em repositórios estáveis, públicos e que seguem padrões internacionais é fundamental para assegurar os princípios FAIR, que orientam que os dados sejam localizáveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e reutilizáveis (*Reusable*) (Wilkinson *et al.*, 2016). A ausência dessa prática compromete a transparência científica, limita o reuso dos dados e enfraquece a integridade da comunicação científica. Portanto, os repositórios de dados se tornam agentes mediadores-chave entre publicações e dados de pesquisa, o que torna vital entender como eles são usados pelos pesquisadores (Jiao; Li; Fang, 2022).

Portanto, analisar a disponibilidade dos conjuntos de dados pode contribuir para a otimização, a preservação e a confiabilidade de todo o ecossistema de comunicação científica que abarca os dados de pesquisa.

### **3 ASPECTOS METODOLÓGICOS**

Foi realizada uma pesquisa bibliográfica cujo objetivo foi identificar os *data papers* com pelo menos um autor filiado às instituições brasileiras. Essa pesquisa, ocorreu em duas etapas: a) a primeira nas bases de dados *Scopus* e *Web of Science*, para identificar os *data papers* publicados por autores filiados às instituições brasileiras; b) a segunda, no catálogo da OpenAlex, para extrair os relatórios com dados sobre *data papers* identificados na primeira coleta de modo a enriquecer os dados da pesquisa.

Foi realizada uma pesquisa documental, na qual os 634 *data papers* identificados na coleta anterior, foram analisados individualmente para: a) identificar a quantidade de conjuntos dados mencionada por cada *data paper*, essa contabilização se deu pela quantidade do *Digital Object Identifier* (DOI) ou Handles, mencionados como depósitos em cada *data paper*, procedimento possível pois os pesquisadores frequentemente usam links da web e DOIs para se referir a um repositório (Jiao, Li, Fang, 2022); e, b) verificar as formas de disponibilização dos dados de pesquisa associados aos *data papers*.

Nesta etapa foram identificadas o tipo de disponibilização: a) em **repositórios**: 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*, 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 por fim, e) como **material suplementar e repositório**: quando mencionava as duas formas.

A partir da identificação dos 26 periódicos, nos quais foram publicados os 349 *data papers* que publicaram os dados em repositórios, outro levantamento documental foi realizado, cujo objetivo foi verificar nas políticas de dados dos periódicos as orientações referentes a disponibilidades dos dados e os repositórios recomendados. A tipologia dos repositórios indicados para o depósito dos conjuntos de dados foi verificada no *Registry of Research Data Repositories (R3data)*.

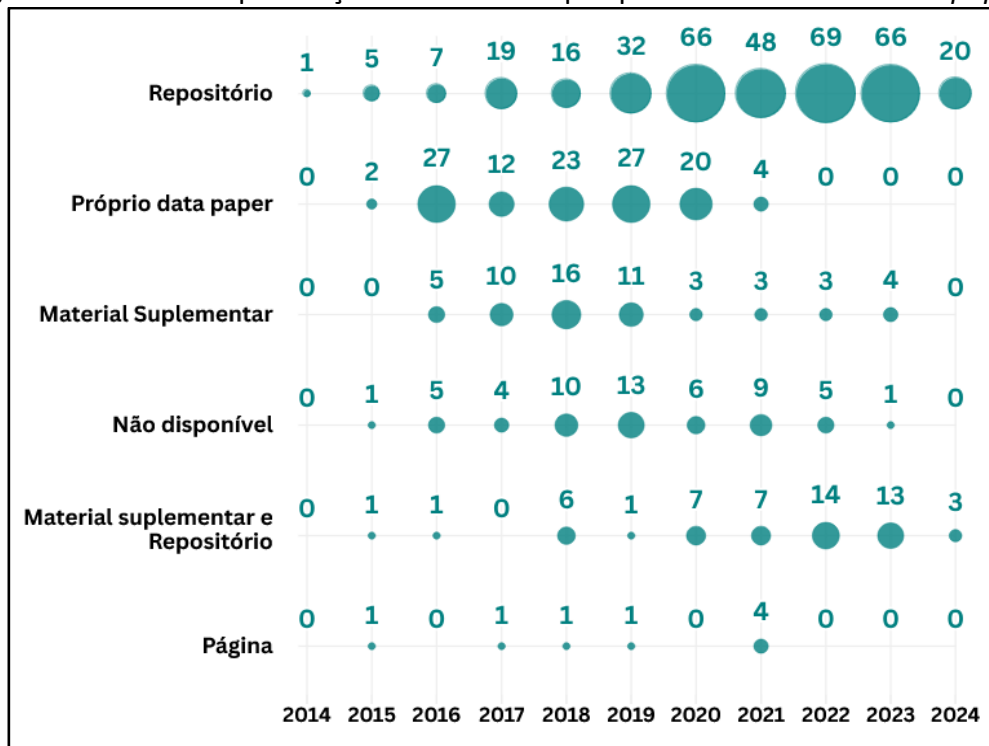
A análise dos dados se deu com a utilização dos métodos mistos, nos quais a abordagem quantitativa utiliza estatística descritiva, contabilizando a frequência, a porcentagem e o cálculo da média.

A limitação deste estudo é que ele analisou apenas a tipologia dos repositórios e não os repositórios em si, além disso, a consulta às diretrizes dos periódicos, considerou apenas a política vigente, não abarcando as políticas editoriais anteriores, o que pode influenciar na análise da disponibilidade, pois às orientações podem de ter mudado ao longo do tempo. Nos casos em que os dados foram disponibilizados de forma combinada, como material suplementar e em repositório, é importante ressaltar que não foi realizada a análise qualitativa dos conjuntos de dados para verificar se eram de fato os mesmos arquivos ou não.

#### 4 RESULTADOS

Conforme figura 1, a disponibilidade dos dados de pesquisa em *data papers* foi sendo realizada de diversas maneiras ao longo dos anos<sup>2</sup>. No ano de 2016 houve um destaque para a disponibilização dos dados de pesquisa a partir do próprio *data paper*. O que pode indicar que os periódicos ainda estavam consolidando um entendimento da melhor maneira de disponibilizar os conjuntos de dados.

**Figura 1** - Formas de publicação dos dados de pesquisa associados aos *data papers*



Fonte: elaborada pelos autores.

Embora o uso dos repositórios para disponibilizar os conjuntos de dados venha ocorrendo desde 2014, sua consolidação se deu de maneira progressiva a partir de 2017, enquanto as outras formas de disponibilização estão sendo reduzidas. Esses dados podem se relacionar com a adoção, pelos periódicos, de orientações claras a respeito da disponibilidade do conjunto de dados em repositórios. A tendência de crescimento do uso de repositórios ao longo dos anos também foi identificada nos estudos de Jiao, Li e Fang (2022).

Dentre as formas identificadas para a disponibilização dos dados de pesquisa nos 38 periódicos que publicaram os 634 *data papers* analisados, em 1,26% (8) dos *data papers*, os

<sup>2</sup> Cabe ressaltar que os dados referentes a 2024 se referem até o mês de maio. Mês no qual foi realizada a coleta dos dados desta pesquisa.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

dados foram disponibilizados em **páginas pessoais ou institucionais**. Em 8,36% (53) dos *data papers*, os conjuntos de dados foram disponibilizados de maneira combinada, **tanto em repositórios como em material suplementar**. Os dados não foram disponibilizados em 8,52% (54) dos *data papers*. Os dados foram disponibilizados como **material suplementar** em 8,68% (55) dos *data papers*. Os dados foram disponibilizados no **próprio data paper**, em 18,14% (115). E em 55,05% (349) dos casos os conjuntos de dados foram disponibilizados em **repositórios**, totalizando 26 periódicos, os quais estão arrolados na Tabela 1.

**Tabela 1 - Tipo de disponibilidade dos dados conforme os periódicos**

<b>Tipo de disponibilidade</b>	<b>F* periódicos</b>	<b>Periódicos (<i>Data Paper</i>)**</b>
<b>Repositórios</b>	26	<i>Data in brief</i> (164), <i>Scientific data</i> (82), <i>Earth system science data</i> (34), <i>Data</i> (16), <i>Biodiversity Data Journal</i> (15), <i>Gigascience</i> (5), <i>BMC research notes</i> (5), <i>Chemical data collections</i> (5), <i>Geoscience data journal</i> (30), <i>Frontiers in ecology and evolution</i> (2), <i>Microbiology resource announcements</i> (2), <i>Frontiers in neuroscience</i> (2), <i>Proteomics</i> (1), <i>Frontiers in genetics</i> (1), <i>Annals of forest Science</i> (1), <i>Big earth data</i> (1), <i>Ecology</i> (1), <i>Ecosistemas</i> (1), <i>Frontiers in marine Science</i> (1), <i>Frontiers in psychology</i> (1), <i>Frontiers in public health</i> (1), <i>Global ecology and biogeography</i> (1), <i>Journal of open humanities data</i> (1), <i>Marine genomics</i> (1), <i>Scientific reports</i> (1), <i>ZooKeys</i> (1).
<b>Dados não foram disponibilizados</b>	14	<i>Data in brief</i> (27), <i>Scientific data</i> (7), <i>Chemical data collections</i> (7), <i>Frontiers in Marine Science</i> (2), <i>Frontiers in Earth Science</i> (2), <i>Earth system science data</i> (1), <i>Biodiversity Data Journal</i> (1), <i>Frontiers in ecology and evolution</i> (1), <i>Marine genomics</i> (1), <i>Cahiers Agricultures</i> (1), <i>Frontiers in microbiology</i> (1), <i>Frontiers in molecular biosciences</i> (1), <i>Human mutation</i> (1), <i>Taxon</i> (1).
<b>Material suplementar</b>	12	<i>Ecology</i> (19), <i>Scientific data</i> (13), <i>Data in Brief</i> (12), <i>Biodiversity Data Journal</i> (2), <i>Chemical data collections</i> (2), <i>Earth system science data</i> (1), <i>Data</i> (1), <i>Frontiers in genetics</i> (1), <i>ZooKeys</i> (1), <i>Frontiers in bioinformatics</i> (1), <i>Nature Conservation</i> (1), <i>Naturwissenschaften</i> (1).
<b>Repositórios como em material suplementar</b>	7	<i>Ecology</i> (31), <i>Global Ecology and Biogeography</i> (8), <i>Biodiversity Data Journal</i> (7), <i>Proteomics</i> (3), <i>Data in Brief</i> (2), <i>Scientific data</i> (1), <i>Journal of biogeography</i> (1).
<b>Páginas pessoais ou institucionais</b>	5	<i>Data in brief</i> (3), <i>BMC bioinformatics</i> (2), <i>Biodiversity Data Journal</i> (1), <i>Frontiers in psychology</i> (1), <i>Comparative cytogenetics</i> (1).
<b>Próprio data paper</b>	2	<i>Data in brief</i> (107), <i>Chemical data collections</i> (8)

**Fonte:** elaborado pelos autores.

Notas: \*F (frequência): refere-se a quantidade de periódicos.

\*\*Quantidade de *data papers* publicados pelos periódicos, conforme tipo de disponibilidade.

Dito de outra forma, em pouco mais da metade dos *data papers* publicados, os repositórios foram a opção dos autores para disponibilizar os dados de pesquisa. No caso dos artigos científicos publicados na Plos One, Federer *et al.* (2018) havia identificado que o uso dos repositórios está longe do método dominante de partilha de dados, correspondendo a cerca de 20% dos artigos que foram analisados.

Esta variedade de disponibilização dos conjuntos de dados de pesquisa levou Candela *et al.*, (2015) a afirmar que em geral os *data papers* são os principais objetos de preocupação dos periódicos, tornando os conjuntos de dados objetos secundários. Jiao, Li e Fang (2022) corroboram com essa afirmação destacando que esses diferentes mecanismos de compartilhamento, impedem que os dados cumpram seus objetivos na comunicação acadêmica, além de prejudicar sua preservação.

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 1. Com isso, pode-se dizer que, em 26,66% (169) dos *data papers* os conjuntos de dados conjuntos 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 interessados acabam não tendo acesso aos conjuntos de dados, exceto se entrarem em contato direto com os pesquisadores e se forem atendidos, o que acaba se tornando uma barreira para a verificabilidade e o reuso dos dados de pesquisa. Federer (2022) afirma que as diferenças disciplinares podem contribuir para variações nas taxas de disponibilidade de dados, o que pode explicar os resultados do presente estudo.

Na pesquisa de Jiao, Li e Fang (2022) os resultados revelaram que os métodos dominantes adotados na Plos One, os conjuntos de dados são compartilhados dentro do artigo e/ou nos materiais suplementares, seguidos pela categoria repositório. 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 pesquisadores, embora seja importante ressaltar que a tipologia documental analisada pode ter influência direta nos resultados.

#### 4.1 Compartilhamento de dados em repositórios

Com o objetivo de identificar se as orientações descritas pelos periódicos em relação a tipologia de repositório de dados foram seguidas, foram analisados 638 conjuntos de dados, vinculados a 349 *data papers*, publicados em 26 periódicos, cujos dados foram disponibilizados em repositórios de dados, conforme Tabela 2.

**Tabela 2 - Disponibilidade dos conjuntos de dados segundo a tipologia dos repositórios**

continua...

Periódico	Orientação do periódico	Data paper	Total de data sets			Repositório			
			Multidisciplinar		Disciplinar				
	Tipo de orientação	F*	F	%	MD**	F	%	F	%
Data in brief	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	164	204	31,97%	1,24	139	21,79%	65	10,19%
Scientific data	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	82	249	39,03%	3,04	117	18,34%	132	20,69%
Earth system science data	Repositórios temáticos	34	45	7,05%	1,32	25	3,92%	20	3,13%
Data	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	16	20	3,13%	1,25	19	2,98%	1	0,16%
Biodiversity Data Journal	Repositórios temáticos	15	19	2,98%	1,27	3	0,47%	16	2,51%
BMC research notes	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	5	8	1,25%	1,60	4	0,63%	4	0,63%
Chemical data collections	Recomenda o Mendeley Data	5	8	1,25%	1,60	1	0,16%	7	1,10%
Gigascience	Recomenda o GigaDB	5	24	3,76%	4,80	-	-	24	3,76%
Geoscience data journal	Repositórios indexados no R3Data ou que estejam numa lista específica do periódico	3	21	3,29%	7	2	0,31%	19	2,98%
Frontiers in ecology and evolution	Recomenda o Figshare	2	12	1,88%	6	-	-	12	1,88%
Frontiers in neuroscience	Recomenda o Figshare	2	2	0,31%	1	2	0,31%	-	-
Microbiology resource announcements	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	2	9	1,41%	4,50	-	-	9	1,41%
Annals of forest science	Repositórios indexados no R3Data	1	1	0,16%	1	1	0,16%	-	-
Big earth data	Repositórios indexados no R3Data	1	1	0,16%	1	-	-	1	0,16%

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

Periódico	Orientação do periódico	Data paper	Total de data sets			Repositório			
			F*	F	%	MD**	Multidisciplinar		Disciplinar
	Tipo de orientação					F	%	F	%
Ecology	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	1	1	0,16%	1	1	0,16%	-	-
Ecosistemas	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	1	1	0,16%	1	1	0,16%	-	-
Frontiers in genetics	Recomenda o Figshare	1	2	0,31%	2	1	0,16%	1	0,16%
Frontiers in marine science	Recomenda o Figshare	1	1	0,16%	1,00	1	0,16%	-	-
Frontiers in psychology	Recomenda o Figshare	1	1	0,16%	1	1	0,16%	-	-
Frontiers in public health	Recomenda o Figshare	1	1	0,16%	1	-	-	1	0,16%
Global ecology and biogeography	Recomenda o Dryad	1	1	0,16%	1	-	-	1	0,16%
Journal of open humanities data	Possui uma lista própria de repositórios	1	1	0,16%	1	1	0,16%	-	-
Marine genomics	Preferencialmente em repositórios específicos	1	1	0,16%	1	-	-	1	0,16%
Proteomics	Repositórios que fazem parte do consórcio ProteomeXchange	1	3	0,47%	3	-	-	3	0,47%
Scientific reports	Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar	1	1	0,16%	1	1	0,16%	-	-
ZooKeys	Recomenda o Dryad	1	1	0,16%	1	-	-	1	0,16%
<b>Totais</b>		<b>349</b>	<b>638</b>	<b>100%</b>	<b>1,83</b>	<b>320</b>	<b>50,16%</b>	<b>318</b>	<b>49,84%</b>

Fonte: elaborado pelos autores.

Notas: F data paper\* (data papers depositado em repositório): quantidade de *data papers* cujos dados foram depositados em repositórios de dados  
MD\*\*(média): média de conjuntos de dados (*data sets*) por *data paper*.

Dos 26 periódicos identificados, 10 deles sugerem o depósito em repositórios multidisciplinares específicos, sendo que: seis sugerem o *Figshare*, dois sugerem o *Dryad*, um recomenda o *GigaScience Data Base* e um sugere o *Mendeley Data*. A recomendação pelo uso de repositórios disciplinares, temáticos ou específicos foi realizada por três periódicos, o *Earth System Science*, o *Biodiversity Data Journal*, e o *Marine genomics*, no entanto, esses periódicos não especificam qual ou quais os repositórios disciplinares seriam ideais para o depósito dos dados de pesquisa. O periódico *Proteomics*, que restringiu tal seleção aos repositórios que fazem parte do consórcio *ProteomeXchange*<sup>3</sup>.

Os demais periódicos (12), deixam a seleção do repositório a critério dos autores, uma vez que apenas recomendam o depósito “Sempre que possível em repositórios específicos, caso não haja usar multidisciplinar” ou ainda que o repositório esteja indexado no R3data ou lista específica definida pelo periódico. Esses resultados indicam que não há uma padronização na recomendação clara dos periódicos em utilizar repositórios disciplinares, pois apenas dois periódicos realizaram essa recomendação.

Ao realizar a comparação entre a sugestão do periódico e a efetivação do depósito dos dados de pesquisa, segundo a tipologia do repositório foi possível verificar que:

a) nos seis periódicos que indicaram o repositório multidisciplinar *Figshare*, em três títulos, os dados foram de fato depositados em repositórios multidisciplinares, em dois a opção foi pelo repositório disciplinar e um depositou os conjuntos nas duas tipologias de repositórios;

b) nos periódicos que sugeriram repositórios específicos, dois periódicos que sugeriram o *Dryad* e um indicou o *GigaScience Data Base*, ambos seguiram a recomendação do periódico, compartilharam os dados em repositórios disciplinares e

c) o único que indicou o *Mendeley Data*: publicou nas duas formas.

Com isso, dos 10 periódicos que especificamente indicaram o uso de repositórios multidisciplinares, em apenas três, os autores o fizeram especificamente, em quatro

---

<sup>3</sup> O *ProteomeXchange* uma colaboração internacional cujo objetivo é compartilhar conjuntos de dados de proteínas. Este foi um esforço internacional criado para evitar duplicação e facilitar o acesso e a reanálise de dados proteômicos (Stroe, 2024).

publicaram em repositórios disciplinares e três publicaram nas duas tipologias de repositórios.

Dos quatro periódicos que indicaram o uso de repositórios disciplinares, apenas os *data papers* publicados em dois dos títulos seguiram essa orientação e os demais publicaram em ambos os repositórios, multidisciplinares e disciplinares.

Nos demais periódicos, onde a indicação pela tipologia do repositório, não foi estabelecida (12), em cinco periódicos os conjuntos de dados foram depositados em repositórios multidisciplinares, em dois periódicos o depósito ocorreu em repositórios disciplinares e, em cinco periódicos o depósito ocorreu nos dois tipos de repositórios.

Com isso, é possível verificar que o uso apenas do repositório multidisciplinar ocorreu em oito periódicos, o mesmo se repetiu em repositórios disciplinares (8) e em 10 periódicos o depósito foi realizado em ambas às tipologias de repositórios. De maneira geral, esses resultados sugerem, que a falta de clareza nas orientações quanto aos tipos de repositórios a serem utilizados, têm levado os autores a selecionarem os repositórios com base em critérios alheios às recomendações dos periódicos.

Ao se realizar uma análise a partir da quantidade de conjuntos de dados que foram depositados nos tipos de repositórios indicados pelos periódicos, percebeu-se não houve uma diferença significativa entre o número de conjuntos de dados depositados em repositórios disciplinares e multidisciplinares, correspondendo a apenas dois conjuntos de dados. Com isso, pode-se inferir que novamente na prática que apesar das orientações dos periódicos, a escolha do tipo de repositório pode depender mais de fatores individuais dos autores ou da natureza dos dados do que das diretrizes dos periódicos.

#### **4.2 A influência da área no compartilhamento dos dados**

A média geral foi de 1,83 conjuntos de dados por *data paper*. As cinco maiores médias foram dos periódicos *Geoscience data journal* (7), *Frontiers in ecology and evolution* (6), *Gigascience* (4,80), *Microbiology resource announcements* (4,50), *Scientific data* (3,04), juntos esses periódicos correspondem a 61,64% (196) do total de conjuntos de dados depositados em repositórios disciplinares (318). O fato dos 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.

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 pela área.

Portanto, os resultados deste trabalho parecem se compatibilizar com Federer (2022) onde simplesmente exigir que os dados sejam compartilhados de alguma forma pode não ter o impacto desejado para 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.

## 5 CONSIDERAÇÕES FINAIS

Esse trabalho se propôs verificar se a disponibilidade dos conjuntos de dados de pesquisa associados aos *data papers* de autores filiados às instituições brasileiras está condizente com as orientações dos periódicos, a partir da identificação de como são publicados os conjuntos de dados e da discussão se a publicação segue as diretrizes do periódico no que condiz aos repositórios de dados.

Todos os periódicos aos quais os conjuntos de dados foram depositados em repositórios possuem orientações voltadas para o depósito dos dados de pesquisa. No entanto, o que se observou na prática é que, embora muitos periódicos exijam a disponibilidade de dados em repositórios, apenas a metade dos *data papers* o fez dessa maneira. A existência de múltiplas formas de disponibilização indica falta de padronização nas orientações. Tais resultados apontam para a necessidade de maior rigor nas políticas editoriais e no fortalecimento da cultura de dados abertos entre os pesquisadores.

O estudo revelou que a publicação de *data papers* por autores brasileiros ainda carece de maior aderência às políticas editoriais relacionadas ao compartilhamento de dados, particularmente no que tange à escolha adequada dos repositórios. Incorporar a verificação da existência, acessibilidade e qualidade dos conjuntos de dados como parte do *peer review* se mostra essencial para fortalecer o ecossistema da publicação de dados.

Para estudos futuros recomenda-se analisar os repositórios especificamente e relacioná-los às temáticas abordadas nos *data papers*. Além de realizar uma busca retroativa nas políticas editoriais dos periódicos para verificar o quanto as orientações referentes às disponibilidades dos dados influenciaram o compartilhamento ao longo do tempo.

## REFERÊNCIAS

- CANDELA, Leonardo; *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: 8 abr. 2025.
- CHAVAN, Vishwas; PENEV, Lyubomir. 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: 8 abr. 2025.
- FEDERER, Lisa; *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://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194768>. Acesso em: 8 abr. 2025.
- FEDERER, Lisa. 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.
- JIAO, Chenyue; LI, Kai; FANG, Zhichao. 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.
- LI, Kai; HUANG, Pao-Pei; JENG, Wei. 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.
- PUERTA-PIÑERO, Carolina; PÉREZ-LUQUE, Antonio Jesus; RODRÍGUEZ-ECHEVERRÍA, Susana. 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: 16 abr. 2025.
- STROE, Oana. Proteomics: a different lens for precision medicine. **EMBL**, 2024. Disponível em: <https://www.embl.org/news/science-technology/proteomics-a-different-lens-for-precision-medicine/>. Acesso em: 03 maio 2025.
- WILKINSON, Mark; *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: 16 abr. 2025.