EXPLORANDO A COBERTURA TEMÁTICA DOS REPOSITÓRIOS ASSOCIADOS AOS DADOS DE PESQUISA DESCRITOS EM *DATA PAPERS*

Crislaine Zurilda Silveira
Universidade Federal de Santa Catarina (UFSC), Brasil
crislaine.silveira@ufsc.br
https://orcid.org/0000-0003-3081-9968

Thiago Magela Rodrigues Dias
Universidade Federal de Santa Catarina (UFSC), Brasil
thiagomagela@gmail.com

Patricia da Silva Neubert
Universidade Federal de Santa Catarina (UFSC), Brasil
patricia.neubert@ufsc.br
https://orcid.org/0000-0002-8909-1898

DOI: *10.22477/xiv.biredial.390*
EJE TEMÁICO: *Datos abiertos*

RESUMEN

Este estudo investiga o comportamento de pesquisadores brasileiros em relação ao depósito de dados de pesquisa vinculados a data papers, tendo em vista a crescente demanda pela disponibilização pública de dados incentivada por agências de fomento e periódicos científicos. O problema central reside na pouca compreensão sobre como autores brasileiros utilizam os repositórios de dados, especialmente no contexto dos data papers, justificando-se a necessidade de caracterizar a cobertura temática dos repositórios utilizados. O objetivo principal foi analisar os repositórios de dados empregados por autores brasileiros, considerando sua tipologia e área de cobertura. Para isso, foi realizada uma pesquisa bibliográfica nas bases Scopus, Web of Science e OpenAlex, identificando 634 data papers, dos quais, 349 mencionavam o uso de repositórios. A análise envolveu a verificação individual de cada data paper. Como resultados, foram identificados 60 repositórios diferentes, sendo 41 disciplinares e 19 multidisciplinares, com os repositórios multidisciplinares concentrando 50,16% dos dados. Mendeley Data, Zenodo e Figshare destacaram-se entre os mais utilizados. Os resultados sugerem que tanto a área do conhecimento quanto o formato e a natureza dos dados influenciam na escolha dos repositórios, e apontam para a necessidade de estudos futuros que explorem a relação entre editoras, áreas do conhecimento e estratégias de compartilhamento de dados científicos.

Palabras-clave: Ciência Aberta. Produção científica. Artigo de dados. Repositório de dados.

ABSTRACT

This study investigates the behavior of Brazilian researchers regarding the deposit of research data linked to data papers, in view of the growing demand for public data availability encouraged by funding agencies and scientific journals. The central problem lies in the limited understanding of how Brazilian authors utilize data repositories, especially in the context of data papers, justifying the need to characterize the thematic coverage of the repositories used. The main objective was to analyze the data repositories employed by Brazilian authors, considering their typology and area of coverage. To this end, a bibliographic search was conducted in the Scopus, Web of Science, and OpenAlex databases, identifying 634 data papers, of which 349 mentioned the use of repositories. The analysis involved the individual verification of each data paper. As results, 60 different repositories were

identified, with 41 being disciplinary and 19 multidisciplinary, and the multidisciplinary repositories concentrating 50.16% of the data. Mendeley Data, Zenodo, and Figshare stood out among the most used. The findings suggest that both the field of knowledge and the format and nature of the data influence the choice of repositories, and point to the need for future studies exploring the relationship between publishers, fields of knowledge, and strategies for scientific data sharing.

Keywords: Open Science. Scientific production. Data paper. Data repository.

INTRODUÇÃO

Ao longo da última década, tem havido uma demanda internacional crescente para tornar os dados produzidos mais disponíveis para a comunidade de pesquisa e o público. Essa demanda tem sido fortemente influenciada pelas instituições de financiamento, editores de periódicos e pelos movimentos que envolvem a ciência que vem promovendo a abertura de dados, incentivando o uso de repositórios de dados públicos (Thoegersen; Borlund, 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, alguns repositórios também incluem metainformações sobre dados, licença e identificador único persistente (Vázquez *et al*., 2021).

Os repositório de dados junto com os *data papers* fazem parte do ecossistema de publicação de dados de pesquisa, no qual os *data papers* vão um passo além de armazenar os dados junto com seus metadados em um repositório, representando uma garantia de acessibilidade e qualidade, pois também exigem a revisão por pares, e que os dados que são descritos estejam disponíveis publicamente em algum repositório estável e online (Mcgillivray, 2022; Puerta-Piñero; Pérez-Luque; Rodríguez-Echeverría, 2020).

Na prática essa relação se estabelece quando os autores enviam os dados de pesquisa para os repositórios ao mesmo tempo em que enviam o *data paper* para a revisão nos periódicos. Os dados depositados nos repositórios devem ser disponibilizados abertamente no momento da publicação do *data paper*, sendo assim, os repositórios são componentes essenciais da infraestrutura para compartilhamento e preservação de dados de pesquisa (Lee; Kim, 2021).

Com isso, tendo em vista identificar o comportamento dos autores brasileiros quanto ao depósito de dados de pesquisa a partir dos dados publicados em *data papers*, este trabalho tem como objetivo: caracterizar a cobertura temática dos repositórios de dados utilizados para depositar os dados de pesquisas vinculados aos *data papers*. Essa caracterização se fez a partir da identificação da média de repositórios de dados utilizados e por meio da análise da sua cobertura temática.

MÉTODOS

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 Scopus e Web of Science, para identificar os *data papers* publicados por autores filiados às instituições brasileiras; b) a segunda, na base do OpenAlex, para extrair os relatórios com as informações sobre *data papers* identificados na primeira coleta.

Em seguida foi realizada uma pesquisa documental, na qual os 634 *data papers* identificados, foram analisados individualmente para verificar as formas de disponibilização dos dados de pesquisa associados. A verificação de disponibilidade em repositórios se deu a partir da menção do nome do repositório ou do(s) link(s) de acesso aos dados de pesquisa no *data paper*. A partir dessa identificação, foi verificada a quantidade de conjuntos de dados por *data papers*, a partir do número de identificadores, *Digital Object Identifier* (DOI) ou Handles, mencionados como depósitos em cada *data paper*, essa ação foi possível pois os pesquisadores frequentemente usam links da web e DOIs para se referir a um repositório (Jiao, Li, Fang, 2022).

Umas das dificuldades encontradas para identificar os repositórios foi a de que, conforme observado por Jiao, Li e Fang (2022), os pesquisadores frequentemente usaram links da web, DOIs, ou apenas às iniciais para se referir a eles, ou seja, o nome do repositório não é necessariamente mencionado. Nesses casos, para realizar a identificação, foi necessário acessar as URLs dos conjuntos de dados.

Em seguida, utilizando o *Registry of Research Data Repositories* (R3data), foram identificadas as áreas de cobertura e o tipo de repositório, se disciplinar ou multidisciplinar. As áreas de cobertura do R3data obedecem 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), ele fornece os metadados necessários para descrever os repositórios indexados no R3data e 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). Neste trabalho, serão utilizadas apenas as quatros classes principais.

A análise dos dados se deu com a utilização dos métodos mistos, onde na abordagem quantitativa foi utilizada a estatística descritiva, com ênfase na análise da frequência de *data papers* e na frequência de data sets (conjuntos de dados), bem como o cálculo das médias.

RESULTADOS

Do total de *data papers* identificados, 634, em 54,70% (349) deles os dados foram depositados em repositórios em detrimento de outras formas de disponibilização, como por exemplo no próprio *data paper*, como material suplementar, etc. Essa porcentagem supera os 15,4% identificados por Federer *et al*. (2018), quando avaliou às formas de compartilhamento de dados de pesquisa, a partir das declarações de disponibilidade de dados[1] dos artigos científicos publicados na *Plos One* entre os anos 2014 a 2016. Essa diferença na porcentagem da disponibilização dos dados em repositórios pode ter relação com a tipologia de documento usado nessa pesquisa, os *data papers*, uma vez que os repositórios já fazem parte desse ecossistema de publicação.

Com o objetivo de verificar a média de repositórios citados nessas publicações para o depósito dos conjuntos de dados, foi realizada a análise dos *data papers* individualmente, com isso, pode-se perceber, na tabela 1, que houve uma variação na quantidade de repositórios, que foi de 1 a 8 repositórios por *data paper*.

Tabela 1 - *Quantidade de repositórios utilizados*

Quantidade de repositório	F* data paper	%	F** data set	%	Média
1	304	87,11%	374	58,62%	1,23
2	33	9,46%	84	13,17%	2,55
3	9	2,58%	54	8,46%	6,00
4	1	0,29%	4	0,63%	4,00
5	1	0,29%	85	13,32%	85,00
6	-	-	-	-	-
7	-	-	-	-	-
8	1	0,29%	37	5,80%	37,00
Total	**349**	**100%**	**638**	**100%**	**1,83**

Fonte:. Elaborado pelos autores (2025).
Notas: os traços (-) significam zeros.
F* data paper: refere-se a quantidade de data papers que mencionou a utilização de repositórios
F ** data sets: refere-se a quantidade de datasets (ou conjuntos de dados) identificados no repositório específico

A grande maioria, 87,11% (304) dos autores publicaram os dados associados aos *data papers* em apenas um repositório de dados, com uma média de 1,23 conjuntos de dados. Com isso, 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

[1] 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)

conjunto de dados.

Embora haja um predomínio de uso de um único repositório, na tabela 1, 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 ambos[2] pertencem a área da Medicina.

O primeiro *data paper* que mais concentrou repositórios e conjuntos de dados utilizou os seguintes repositórios: Sequence Read Archive (7), BioProject (7), Gene Expression Omnibus (7), Proteomics Identifications Database (2), Mass Spectrometry Interactive Virtual Environment (4), MetaboLights (3), Plasmodium Genomics Resource (5) e o Immunology Database and Analysis Portal (2). O segundo *data paper* que mais concentrou repositórios e conjuntos de dados utilizou os seguintes repositórios: Sequence Read Archive (3), BioProject (6), BioSample (6), Lancaster University Research Directory (24), European Nucleotide Archive (6). O que ambos têm em comum, é que com o exceção do Lancaster University Research Directory que é um repositório multidisciplinar, todos os demais são repositórios disciplinares da área das Ciências da Vida, conforme figura 1, o que sugere que a área pode influenciar na quantidade e na seleção dos repositório de dados associados aos *data papers*.

Em seguida, na tabela 2, são arrolados todos os repositórios onde os conjuntos de dados foram depositados, bem sua média e sua tipologia.

Tabela 2 - Características dos repositórios

Nome do Repositório	F* data set	F** data paper	Média	Tipo ***
1. Mendeley Data	106	100	1,06	M
2. Sequence Read Archive	69	12	5,75	D
3. Genbank	64	23	2,78	D
4. Zenodo	62	57	1,09	M
5. Figshare	62	55	1,13	M
6. Pangaea	31	13	2,38	D
7. Lancaster University Research Directory	24	1	24	M
8. Open Science Framework	23	7	3,29	M

[2] 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* data set	F** data paper	Média	Tipo ***
9. Github	17	17	1	M
10. BioProject	17	6	2,83	D
11. Proteomics Identifications Database	13	10	1,30	D
12. Global Biodiversity Information Facility	13	12	1,08	D
13. Gene Expression Omnibus	12	6	2	D
14. European Nucleotide Archive	10	4	2,50	D
15. Mass Spectrometry Interactive Virtual Environment	8	5	1,60	D
16. Science Data Bank	7	6	1,17	M
17. BioSample	7	2	3,50	D
18. GigaScience Database	7	5	1,40	D
19. Cambridge Crystallographic Data Center	7	4	1,75	D
20. Dryad	6	6	1	M
21. Synapse	6	5	1,20	M
22. Plasmodium Genomics Resource	5	1	5	D
23. Dados Abertos do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro	4	2	2	D
24. Atmospheric Science Data Center	4	1	4	D
25. National Centers for Environmental Information	4	3	1,33	D
26. Harvard DataVerse	3	3	1	D
27. Sea scientific open data publication	3	3	1	D
28. MetaboLights	3	1	3	D
29. ArrayExpress	3	3	1	D
30. RCSB Protein Data Bank	3	2	1,50	D
31. Data repository	3	3	1	D
32. herbarium of Federal University of Mato Grosso do Sul	2	2	1	D
33. NCBI DataSets	2	2	1	D
34. Immunology Database and Analysis Portal	2	1	2	D

Nome do Repositório	F* data set	F** data paper	Média	Tipo ***
35. Archive of Data on Disability to Enable Policy	1	1	1	D
36. DataSuds	1	1	1	M
37. IsoArcH	1	1	1	D
38. Cirad dataverse	1	1	1	M
39. Digital CSIC	1	1	1	M
40. DataverseNL	1	1	1	M
41. Fairdata IDA Research Data Storage Service	1	1	1	M
42. Global Health Data Exchange	1	1	1	D
43. ProteomeXchange	1	1	1	D
44. Metagenomics analysis server	1	1	1	D
45. British Antarctic Survey	1	1	1	D
46. TRY	1	1	1	D
47. Ag Data Commons	1	1	1	D
48. Environmental Data Initiative Repository	1	1	1	D
49. International Neuroimaging Data-Sharing Initiative	1	1	1	D
50. OTELo Research Data Repository	1	1	1	D
51. GFZ Data Service	1	1	1	D
52. ICOS Carbon Portal	1	1	1	D
53. British Oceanographic Data Centre	1	1	1	D
54. Digital Rocks Portal	1	1	1	D
55. Dagshub	1	1	1	M
56. Repositório da UNESP	1	1	1	M
57. Repositório da Universidade de São Paulo	1	1	1	M

Nome do Repositório	F* data set	F** data paper	Média	Tipo ***
58. Repositório de Dados de Pesquisa da Fundação Getulio Varga	1	1	1	M
59. Repositório de Dados de Pesquisa da Unicamp (REDU)	1	1	1	M
60. Repository of Open Access Data sets	1	1	1	M
Total	**638**	**409****	**1,56**	

Fonte: Dados da pesquisa (2025).
Notas: F* data sets: refere-se a quantidade de data sets (ou conjuntos de dados) publicados no repositório específico
F** data paper: refere-se a quantidade de data papers que mencionou o repositório específico
Tipo***: refere-se ao tipo de cobertura do repositório: D indica repositório é disciplinar e M indica que o repositório é multidisciplinar
Total de data papers 409****Esse número se distingue do total de data papers listados na Tabela 1. Isso ocorreu porque um único *data paper* abrangeu múltiplos repositórios, e, nesse caso, cada repositório foi contabilizado separadamente como um data paper.

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 (M), Sequence Read Archive (D), No-repository web site, Institucional repository, Github (M), 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, 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 (M), GitHub (M), Dataverse (M), Zenodo (M) e Bioproject (D).

Ao realizar uma comparação entre esses dois estudos, pode-se observar que sete entre os dez repositórios se repetem, são eles: Figshare (M), Gene Expression Omnibus (D), Genbank (D), Dryad (M), Sequence Read Archive (D), Github (M) e o Dataverse (M). Esses resultados, podem estar atrelados ao fato de ser o mesmo periódico analisado, 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 (M), Open Science Framework (M), Zenodo (M) e Bioproject (D). Esse resultado pode demonstrar que os autores reconhecem a importância que estes repositórios têm para as suas áreas, uma vez que esses se mantêm entre os dez que mais concentram o depósito de dados.

Entretanto, embora haja uma constância os repositórios utilizados neste trabalho e nos estudos de Federer *et al.* (2018) e Jiao, Li, Fang (2022), pode-se observar que o repositório Mendeley Data não aparece neles, enquanto que neste estudo é o repositório que mais concentra conjuntos de dados e é utilizado por cerca de um terço dos *data papers*. O que pode indicar

uma ascensão dos repositórios multidisciplinares, quando analisados os *data papers* publicados por autores filiados às instituições brasileiras.

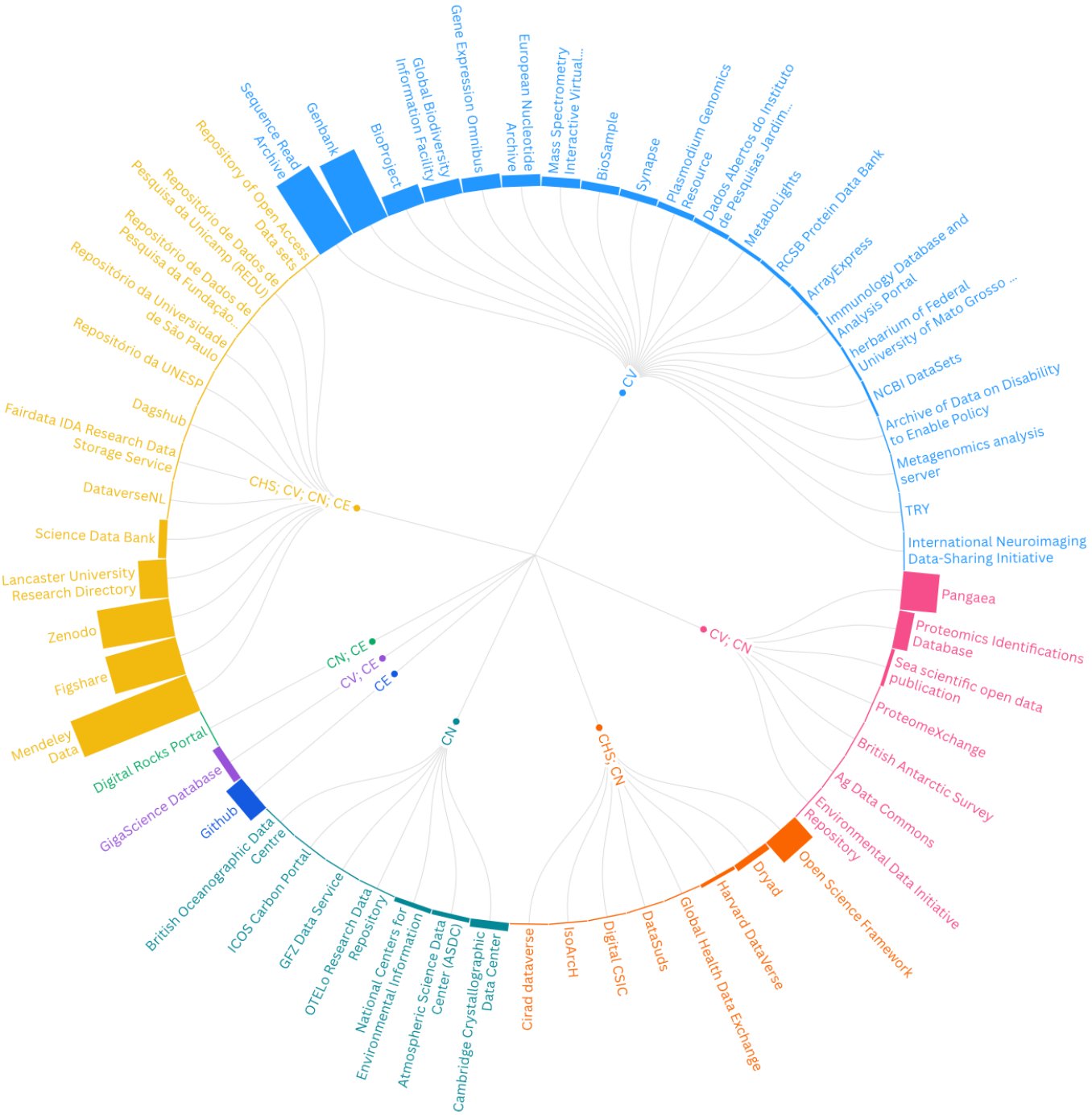
Foram identificados 60 repositórios diferentes, sendo 41 disciplinares, que juntos concentram 49,84% (318) conjuntos de dados e 19 multidisciplinares, que mesmo estando em menor quantidade, concentram 50,16% (320) conjunto de dados. Esses achados sugerem que o volume de dados em repositórios multidisciplinares ou generalistas, pode estar relacionado a ausência de repositórios para uma comunidade específica (Assante *et al*., 2016) ou ainda que devido às recomendações dos periódicos, que sugerem prioritariamente o uso de repositórios específicos, os autores têm aderido a esse tipo de repositorio para realizar o depósitos dos dados disciplinares (Candela *et al*., 2015).

Embora os repositórios disciplinares sejam em maior quantidade, quase o dobro dos multidisciplinares, esses possuem menos *datasets* e tendem a concentrar uma maior quantidade de conjuntos por *data paper*, por isso suas médias de depósito são altas, variando entre 5,75 a 1. Além disso, foi possível verificar que os conjuntos de dados tendem a ser depositados em repositórios específicos tais como: Sequence Read Archive (69), criado em 2007, Genbank (64), criado em 1982 e Pangea (31), criado em 1992. Sendo que os dois primeiros são repositórios voltados ao depósito de dados de sequenciamento genético e o terceiro, é um repositório com mais de 30 anos de história, e está voltado para a publicação e disseminação de dados georreferenciados das ciências da Terra, ambientais e da biodiversidade. Logo, pode-se supor que a área e o tempo de criação do repositório podem influenciar na escolha pelo local para depositar os dados de pesquisa.

Os repositórios multidisciplinares, mesmo em menor número, possuem uma maior quantidade de dados depositados, embora possuam maior variação entre as médias identificadas, cuja faixa se deu de 24 a um conjunto. Esses concentram, 36,05% (230) dos dados em três repositórios, Mendeley (106), Zenodo (62) e Figshare (62).

Com o objetivo de identificar as áreas de cobertura dos repositórios, na figura 1, estão representadas, bem como a quantidade de conjuntos de dados.

Figura 1 - Áreas de cobertura dos repositórios



Fonte: Elaborada pelos autores (2025).
Notas: CHS; CV; CN; CE - Ciências Humanas e Sociais; Ciências da Vida; Ciências Naturais; Ciências da Engenharia
CV - Ciências da Vida
CV; CN - Ciências da Vida; Ciências Naturais
CHS; CV; CN - Ciências Humanas e Sociais; Ciências da Vida; Ciências Naturais
CN - Ciências Naturais
CE - Ciências da Engenharia
CV; CE - Ciências da Vida; Ciências da Engenharia
CHS; CN - Ciências Humanas e Sociais; Ciências Naturais
CN; CE - Ciências Naturais; Ciências da Engenharia
CHS; CV - Ciências Humanas e Sociais; Ciências da Vida

Em relação às áreas de cobertura, as Ciências da Vida concentraram 21 repositórios, as Ciências Humanas e Sociais; Ciências da Vida; Ciências Naturais; Ciências da Engenharia, concentraram 13 repositórios, as Ciências da Vida; Ciências Naturais e Ciências Naturais concentraram sete repositórios cada, às Ciências Humanas e Sociais; Ciências da Vida; Ciências Naturais concentram 5 repositórios, Ciências Humanas e Sociais; Ciências Naturais concentraram 2 repositórios e as Ciências da Engenharia, as Ciências da Vida; Ciências da Engenharia, Ciências Naturais; Ciências da Engenharia e as Ciências Humanas e Sociais; Ciências da Vida concentram um repositório cada e um repositório não teve a área identificada.

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 Ciência da Vida, se considerada a quantidade de conjuntos de dados depositados nesses repositórios, 168, o que representa 26,33% do total de dados depositados, pode-se afirmar que há indícios de uso pelos pesquisadores dessas áreas.

Neste estudo não foi encontrado nenhum repositório com uma cobertura específica da área das Ciências Humanas e Sociais, no entanto, cabe ressaltar que essa área abarca diversas áreas com características de publicação e objetos de estudos bem diversos (McGillivray *et al*., 2022). Isso pode indicar que os autores podem recorrer a repositórios multidisciplinares.

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 mais concentra um quantitativo maior de repositórios por *data paper*. Enquanto que às Ciências Humanas e Sociais, não foram representadas por nenhum repositório específico.

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 sejam *data papers*, e dados de pesquisa, por isso é vital entender como eles são usados pelos pesquisadores.

CONCLUSÕES

Pouco mais da metade dos *data papers* publicados faz menção ao depósito dos dados da pesquisa em repositório, o que sugere que a outra metade, ou não disponibiliza os conjuntos de dados ou disponibiliza como material complementar ao *data paper*, o que pode dificultar seu acesso e reuso. Além disso, se observa que há a tendência, entre aqueles que utilizam repositórios de dados, em disponibilizar apenas um conjunto de dados, mas entre aqueles que depositam mais de um conjunto de dados esses o fazem em mais de um repositório, frequentemente em um repositório especializado.

Em relação às áreas de cobertura, as Ciências da Vida concentraram 21 repositórios, sendo a grande maioria deles reconhecidos na área, e em geral, a disponbilidade é de mais de um conjunto por *data paper*.

Este resultado sugere que, em relação ao uso de repositórios, o comportamento do pesquisador está associado à área do conhecimento e a quantidade do conjunto de dados. Para estudos futuros recomenda-se verificar o comportamento das áreas do conhecimento de maneira mais específica quanto a quantidade de *datasets* compartilhados e repositorios utilizados, assim como a vinculação dos editores comerciais na criação e manutenção dos repositórios. Portanto, o*s resultados sugerem que tanto a área do conhecimento quanto o formato e a natureza dos dados influenciam na escolha dos repositórios, e apontam para a necessidade de estudos futuros que explorem a relação entre editoras, áreas do conhecimento e estratégias de compartilhamento de dados científicos.*

REFERÊNCIAS

Assante, M. *et al*. (2016). Are scientific data repositories coping with research data publishing? *Data Science Journal*, 15(6), 79–83. http://dx.doi.org/10.5334/dsj-2016-006

Candela, L. *et al*. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747-1762. https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.23358

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: an analysis of Data Availability Statements. *PLoS ONE* 13(5), e0194768. https://doi.org/10.1371/journal.pone.0194768

Jiao, C., Li, K., & Fang, Z. (2022). Data sharing practices across knowledge domains: a dynamic examination of data availability statements in PLOS ONE publications. *Journal of Information Science*, 50(3), 673-689. https://doi-org.ez46.periodicos.capes.gov.br/10.1177/01655515221101830

Lee, J., & Kim, J. (2021). Korean researchers' motivations for publishing in data journals and the usefulness of their data: a qualitative study. *Science Editing*, 8(2), 145-152. https://www.escienceediting.org/upload/kcse-246.pdf

McGillivray, B., Marongiu, P., Pedrazzini, N., Ribary, M., Wigdorowitz, M., & Zordan, E.(2022). Deep Impact: a study on the impact of data papers and datasets in the Humanities and Social Sciences. *Publications*, 10(39), 1-40. https://doi.org/10.3390/publications10040039

Puerta-Piñero, C., Pérez-Luque, A.J., & Rodríguez-Echeverría, S. (2020). A Ecosystems está comprometida com a publicação de artigos de dados (Data Papers). *Ecossistemas*, 29(3), 2118. *https://doi.org/10.7818/ECOS.2118*

R3data Coref (2021). *Reviewing the subject classification in re3data*. Blog. https://coref.project.re3data.org/blog/reviewing-the-subject-classification-in-re3data

Strecker, D. *et al*. (2023). *Metadata Schema for the Description of Research Data Repositories*: version 4.0. https://doi.org/10.48440/re3.014

Thoegersen, J.L., & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: a meta-evaluation of studies on researcher data sharing. *Journal of Documentation*, 78(1), 1-17. https://doi-org.ez46.periodicos.capes.gov.br/10.1108/JD-01-2021-0015

Vásquez, I., Novo-Lourés, M., Pavón, R., Laza, R., Méndez, J. R., & Ruano-Ordás, D. (2021). Improvements for research data repositories: the case of text spam. *Journal of Information Science*, 49(2), 285-301.https://doi.org/10.1177/0165551521998636

ANEXO 1

RESUMEN BIOGRÁFICO DE LOS AUTORES

Crislaine Zurilda Silveira

Doutoranda no Programa de Pós Graduação em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC). Atua como Bibliotecária na mesma instituição. Tem interesse nas temáticas de pesquisa: Ciência Aberta, gestão de dados de pesquisa, comunicação científica, gestão de bibliotecas universitárias e competência em informação.

Patrícia da Silva Neubert

Docente do Programa de Pós-graduação em Ciência da Informação (PGCIN) da UFSC. Atua no campo de Comunicação científica, em temáticas associadas à metodologia da pesquisa e publicações, com ênfase na Comunicação, Filosofia e Sociologia da Ciência, desenvolvendo estudos sobre produtividade científica, periódicos científicos e sua estrutura editorial, marketing científico, bases de dados, Acesso Aberto, indicadores da produção científica e fontes de informação.

Thiago Magela Rodrigues Dias

Docente do Programa de Pós-graduação em Ciência da Informação (PGCIN) da UFSC. Tem experiência na área de Ciência da Computação e Ciência da Informação, atuando principalmente nos seguintes temas: Bibliometria, Cientometria, Extração e Integração de Dados, Análise de

Redes Sociais, Análise de Redes de Colaboração Científica, Acesso Aberto, Recuperação e Organização da Informação, Ciência de Dados, Data Mining, Text Mining e Web Mining.

ANEXO 2

REQUERIMIENTOS DE EQUIPO TÉCNICO PARA LA PRESENTACIÓN DE LA PONENCIA

Indicar si se requiere alguno de los siguientes equipos: computadora, proyector, parlantes, software, conexión a Internet, traducción simultánea, mesas, etc.

Será necessário um computador conectado à internet e um projetor.