

A PESQUISA BRASILEIRA EM BIODIVERSIDADE: DATA PAPERS E DATASETS DO GBIF¹

BRAZILIAN BIODIVERSITY RESEARCH: GBIF DATA PAPERS AND DATASETS

Fabiano Couto Corrêa da Silva²
Amanda Santos Witt³
Bryan Nicollas Soares Costa⁴
Laura Vilela Rodrigues Rezende⁵

Resumo: O *Global Biodiversity Information Facility* consiste em uma rede, de alcance internacional, de infraestrutura de dados direcionada para fornecer acesso aberto sobre biodiversidade, incluindo *datasets* e *data papers*. A pesquisa analisa o uso desses recursos por pesquisadores e instituições brasileiras no período de 2010 a 2024, a partir dos filtros tipo de publicação (foco em *data papers*), tipo de acesso (acesso aberto), país ou área do investigador (Brasil). Identificaram-se 536 documentos publicados utilizando *datasets* do GBIF, e após aplicar o filtro da pesquisa, foram recuperados 27 *data papers*, cuja área de cobertura dos dados em sua maioria, foi o Brasil. Os *data papers* foram publicados em 8 periódicos distintos, tais como o *Biodiversity Data Journal* e o *Gigabyte*, e por meio do DOI, verificou-se sua rastreabilidade, o que possibilitou localizar 51 *datasets* usados por esses *data papers*. A análise dos dados foi realizada com base em análise temática e no documento Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU). Conclui-se que instituições brasileiras como Embrapa Florestas, Instituto Osvaldo Cruz e

¹ Este texto foi submetido, avaliado, aprovado, apresentado e premiado no XXIV ENANCIB.

² Professor Adjunto do Departamento de Ciência da Informação/FABICO da Universidade Federal do Rio Grande do Sul (UFRGS), atuando no curso de graduação de Biblioteconomia e o Programa de Pós-Graduação em Ciência da Informação (PPGCIN), ambos da mesma instituição. E-mail: fabianocc@gmail.com. ORCID: <https://orcid.org/0000-0001-5014-8853>.

³ Doutoranda no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC). Bibliotecária-documentalista na Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: amandawitt.asw@gmail.com. ORCID: <https://orcid.org/0000-0001-9334-3018>.

⁴ Mestrando pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS). E-mail: bryannsc@gmail.com. ORCID: <https://orcid.org/0000-0001-6640-7960>.

⁵ Professora associada na Faculdade de Informação e Comunicação da Universidade Federal de Goiás. E-mail: laura_rezende@ufg.br. ORCID: <https://orcid.org/0000-0002-8891-3263>.

Universidades Federais contribuem significativamente para a pesquisa em biodiversidade nos ambientes aquático e terrestre, no *Global Biodiversity Information Facility*.

Palavras-Chave: Data papers. Datasets. Biodiversidade. GBIF. Brasil.

Abstract: *The Global Biodiversity Information Facility is an international network of data infrastructure aimed at providing open access to biodiversity, including datasets and data papers. The research analyzes the use of these resources by Brazilian researchers and institutions from 2010 to 2024, using the filters type of publication (focus on data papers), type of access (open access), country or area of the researcher (Brazil). 536 documents published using GBIF datasets were identified, and after applying the search filter, 27 data papers were retrieved, most of which covered Brazil. The data papers were published in 8 different journals, such as Biodiversity Data Journal and Gigabyte, and their DOI was used to verify their traceability, which made it possible to locate 51 datasets used by these data papers. The data was analyzed using thematic analysis and the United Nations (UN) Sustainable Development Goals (SDGs). It was concluded that Brazilian institutions such as Embrapa Florestas, the Osvaldo Cruz Institute and Federal Universities contribute significantly to biodiversity research in the aquatic and terrestrial environments in the Global Biodiversity Information Facility.*

Keywords: Data papers. Datasets. Biodiversity. GBIF. Brazil.

1 INTRODUÇÃO

A crescente relevância dos *data papers* (artigos de dados) na comunicação científica ressalta a importância do uso e compartilhamento de *datasets* (conjuntos de dados) para promover a transparência, reprodutibilidade e colaboração na pesquisa. Os *data papers* permitem que os pesquisadores publiquem descrições detalhadas dos seus conjuntos de dados, garantindo que outras pessoas possam acessar e utilizar essas informações de maneira eficiente e confiável. Nesse contexto, o *Global Biodiversity Information Facility* (GBIF) se destaca como uma plataforma internacional, financiada por governos de várias partes do mundo, que disponibiliza vastos conjuntos de dados sobre biodiversidade, sendo essencial para pesquisas científicas que necessitam de informações nessa área (Pinheiro, 2003).

Na contemporaneidade, evidenciam-se novas formas de produzir ciência

pautadas no compartilhamento e reuso de dados científicos. Estes deixam de ser entendidos apenas como insumos intermediários de uma pesquisa, o que exige a elaboração de métodos e processos que salientem o seu valor (Henning *et al.*, 2019). Práticas direcionadas à abertura de dados podem estar ligadas ao movimento mais amplo da Ciência Aberta (Rocha, 2019), caracterizado pelos fatores conhecimento científico aberto, tais como softwares abertos e dados científicos abertos; infraestruturas de pesquisa compartilhadas (virtuais ou físicas), que envolvem equipamentos ou conjuntos de instrumentos científicos e variados recursos como é o caso das plataformas e repositórios; engajamento aberto dos atores sociais com fins de ampliar a participação social em ciência aproximando cientistas e sociedade e, por fim, o diálogo aberto com outros sistemas de conhecimento (Unesco, 2022).

Este estudo é norteado pela seguinte problemática: a plataforma de relevância internacional GBIF, tem se destacado no Brasil, com relação a disponibilização de *data papers* e *datasets* no âmbito da biodiversidade, nos últimos dez anos, por pesquisadores e instituições brasileiras? Dessa forma, objetiva-se traçar um panorama geral da publicação de *data papers* correlacionados aos *datasets*, com fins de compreender melhor como a disponibilização e utilização de conjuntos de dados podem impactar a pesquisa científica no Brasil, promovendo a transparência, reprodutibilidade e colaboração necessárias para avanços significativos na área da biodiversidade.

2 PANORAMA CIENTÍFICO EM BIODIVERSIDADE

É preciso, primeiramente, definir o que são *data papers* e *datasets*, que constituem o foco deste estudo. Um *data paper* pode ser entendido como um

documento de metadados pesquisável, o qual descreve um conjunto ou grupo de dados publicado, nos moldes de um artigo em um periódico, revisado por pares, diferenciando-se de um artigo convencional pelo fato de descrever dados e circunstâncias de sua coleta em vez de apresentar hipóteses e conclusões (Schöpfel *et al.*, 2020). Documentar o fluxo do processo de produção de dados e assegurar esteja de acordo com os princípios FAIR, apesar de ser mais trabalhoso, é uma atividade fundamental para a reprodutibilidade da pesquisa (Gay, 2021). O *data paper* é um instrumento que permite o reconhecimento da contribuição científica dos produtores de dados, tornando estes facilmente citáveis, além de incrementar a relevância e o alcance da reutilização desses dados (Gay, 2021). Os *datasets* configuram conteúdos arbitrados, são uma fonte referencial que permite aos autores receberem crédito pela publicação, assim como acontece nas publicações convencionais, são dotados de DOI (*digital object identifier*) e, portanto, podem ser citados (Nassi-Calò, 2014).

Na atualidade, a forma de se conduzir a ciência tem sido alterada pelo movimento conhecido como Ciência Aberta, que abrange uma série de práticas a respeito dos usos da internet e dos elementos da web social para transformar, globalmente, os processos acadêmicos, incluindo desde a formulação de questões e hipóteses até a disseminação dos resultados das investigações científicas (Aventurier; Alencar, 2016). Dentre as práticas e usos, destacam-se, por exemplo, acesso aberto, dados abertos e plataformas colaborativas, que através das novas tecnologias digitais e ferramentas colaborativas incentivam esforços conjuntos de pesquisa e avaliação, difundindo e possibilitando a utilização do conhecimento produzido de modo mais rápido e amplo possível (Caballero-Rivero; Sánchez-Tarragó; Santos, 2019).

Logo, o panorama vigente preconiza o incentivo à reprodutibilidade dos resultados dos estudos, clareza metodológica, acesso aos resultados das investigações e economia de recursos, tais como, tempo, dinheiro, recursos humanos e equipamentos, visando elevar o impacto social da ciência (Caballero-Rivero; Sánchez-Tarragó; Santos, 2019).

Dessa forma, a análise do uso de *datasets* do GBIF por pesquisadores brasileiros pode trazer à tona diversas tendências, desafios e oportunidades. Identificar como esses *datasets* são utilizados pode revelar lacunas na pesquisa, áreas com maior atividade científica e demandas específicas de dados. É fundamental considerar a necessidade de implementação de políticas mais efetivas voltadas à disponibilização de dados de pesquisa, visando facilitar o uso e reuso dos dados por toda a comunidade científica, como destacado em estudo sobre compartilhamento de dados no contexto da ciência brasileira (Araújo *et al.*, 2021).

Além disso, a pesquisa acadêmica brasileira tem demonstrado um crescente interesse pelo estudo do uso e compartilhamento de dados científicos. Isso pode ser observado na aproximação entre comunicação científica e divulgação científica, conforme identificado em estudo sobre comunidades científicas e infraestrutura tecnológica no Brasil para uso de recursos eletrônicos de comunicação e informação na pesquisa (Lustosa; Farias; Farias, 2024).

Os dados ecológicos demonstram sua relevância e utilidade especialmente no que diz respeito ao monitoramento e à conservação da biodiversidade. Ademais, seu acesso na modalidade livre bem como sua reutilização permitem identificar locais em que determinadas espécies estão ameaçadas de extinção, contribuindo com o processo de tomada de decisões para reverter o quadro (Fox *et al.*, 2019).

O crescente interesse pelo tema da biodiversidade e o desenvolvimento de várias iniciativas nesse sentido nos últimos anos, estão atrelados à proposição, no ano de 2015, pela Organização das Nações Unidas, de uma agenda voltada para o desenvolvimento sustentável, denominada Agenda 2030, o que culminou na criação de 17 Objetivos de Desenvolvimento Sustentável. Isto demonstra a crescente preocupação e os esforços globais direcionados para acabar com a pobreza, visando, dentre outros elementos, assegurar sustentabilidade, justiça social e melhores condições de vida para todas as pessoas de todos os lugares do mundo (Organização das Nações Unidas, [2024]).

3 METODOLOGIA

Para identificar e analisar o estado atual da publicação de *data papers* e a disponibilização de *datasets* por instituições e centros de pesquisa do Brasil utilizamos a interface de busca do GBIF (2024), acessível pelo site oficial da rede, aplicando filtros específicos para delimitar as buscas. Utilizamos filtros específicos na plataforma do GBIF para identificar *data papers* associados aos *datasets* brasileiros e verificamos a correspondência entre eles utilizando identificadores únicos como o *Digital Object Identifier* (DOI) e outras referências bibliográficas.

Na primeira etapa, procedeu-se à localização de *data papers* que, em suas referências, citaram *datasets* do GBIF no estudo. Para isso, utilizaram-se filtros específicos para delimitar a busca: tipo de documento; título do periódico; país ou área do investigador, selecionando Brasil; tipo de acesso, privilegiou-se acesso aberto para verificar os documentos e data de publicação, compreendendo os anos de 2014 a 2024. Este recorte temporal foi escolhido para capturar uma visão

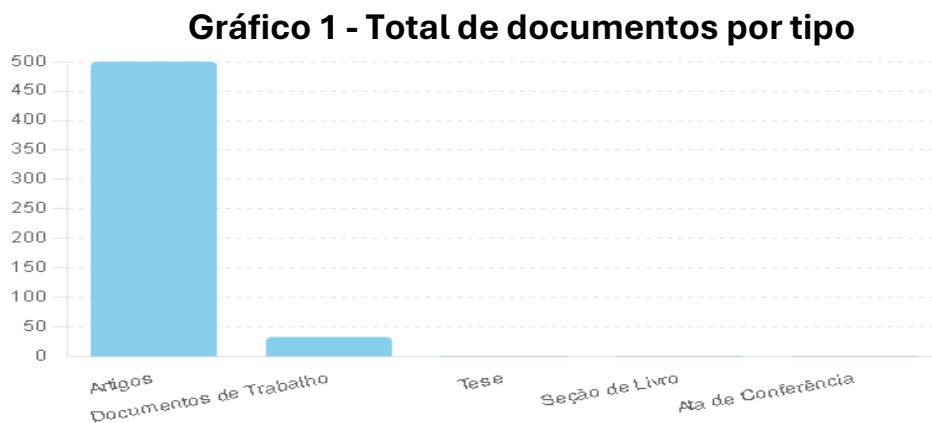
abrangente das tendências e padrões de publicação de *data papers* ao longo de mais de uma década, permitindo uma análise detalhada das mudanças e evoluções na contribuição brasileira para a base de dados do GBIF. Acrescenta-se ainda que o referido intervalo de tempo também coincide com o aumento da digitalização e do acesso aberto a dados de biodiversidade, refletindo as práticas contemporâneas de compartilhamento de dados científicos.

A segunda etapa abrangeu a localização de *datasets* publicados por instituições brasileiras, objetivando identificar as principais publicadoras, suas principais características bem como o número de citações feitas a esses *datasets*, identificando possíveis áreas de destaque ou necessidade de maior incentivo em pesquisas focadas em biodiversidade.

Os dados coletados foram examinados sob o viés da análise temática, por se tratar de um método que permite verificar dados qualitativos, envolvendo a leitura destes em busca de padrões de significado para encontrar temas. O documento Objetivos do Desenvolvimento Sustentável (ODS) foi utilizado na pesquisa como parâmetro para a categorização ambiental do tema principal dos *data papers* analisados. Cabe salientar a importância dos ODS para a sustentabilidade mundial, uma vez que consistem em um conjunto de 17 metas globais para alcançar um mundo mais sustentável até 2030. O pacto foi estabelecido pela Organização das Nações Unidas (ONU) em 2015 e faz parte da Agenda 2030, plano de ação global que preconiza o desenvolvimento sustentável. Os ODS são interconectados e abarcam diversos temas críticos para o desenvolvimento humano visando a erradicação da pobreza e priorizando o progresso de forma horizontal para todos (ONU Brasil, 2024).

4 RESULTADOS

A pesquisa inicial revelou o total de 536 documentos publicados utilizando *datasets* do GBIF. Com relação ao tipo de trabalho científico produzido, dividem-se em 500 artigos, 33 documentos de trabalho (relatórios técnicos, *white papers* ou *working papers*, etc.), 1 tese, 1 seção de livro e 1 ata de conferência (ver Gráfico 1).



Fonte: Elaborado pelos autores (2024).

Ao filtrar os estudos por “país ou área do investigador” e selecionar o Brasil, foram localizados 27 estudos, estando todos publicados em forma de artigos. Destes, 26 são de acesso aberto, e 22 identificam o país ou área de cobertura dos dados, sendo 21 artigos cobrindo o Brasil e 1 o Chile. O restante não informou esse dado.

No âmbito da Biodiversidade, foram identificadas temáticas variadas nos artigos analisados. Com base nos Objetivos do Desenvolvimento Sustentável (ODS), os temas foram divididos em dois ambientes, o aquático e o terrestre, correspondentes aos ODS 14 e 15, respectivamente.

Quadro 1 - Temáticas abrangidas pelos artigos de acordo com os ODS 14 e 15

ODS	No. de artigos	Abrangência temática
ODS 14 Ambiente aquático	3	Fauna ictiológica Biodiversidade marinha Planorbídeos
ODS 15 Ambiente terrestre	24	Insetos Plantas Florestas Árvores Flores Invertebrados do solo Serpentes

Fonte: Elaborado pelos autores (2024).

A respeito dos ambientes analisados, percebe-se que abrangem uma diversidade de assuntos:

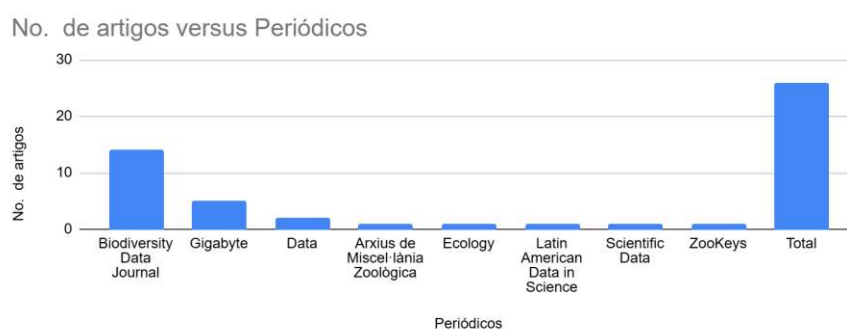
- Ambiente aquático: os artigos versam sobre a fauna ictiológica (peixes) incluindo temas como macroinvertebrados bentônicos, produtores primários estuarinos, biodiversidade marinha da Antártida e planorbídeos (com foco nas espécies de moluscos *biomphalaria*, responsáveis pela disseminação da esquistossomose);
- Ambiente terrestre: Em maior número destacam-se oito estudos sobre insetos, incluindo vagalumes, abelhas, besouros, e formigas, sendo quatro deles sobre mosquitos, com destaque para os transmissores de doenças: flebotomíneos (incluem os principais vetores de transmissão da *leishmaniose*, doença infecciosa causada por protozoários do gênero *Leishmania*) e *culicidae* (família de insetos habitualmente chamados de muriçocas, mosquitos ou pernilongos). O restante dos trabalhos

inclui assuntos variados como plantas; florestas; árvores; flores; invertebrados do solo e serpentes.

Ademais, verificou-se uma variedade de biomas e regiões analisados nas pesquisas, tais como, Mata Atlântica, Cerrado, Pantanal, Caatinga e Amazônia, demonstrando a riqueza brasileira em termos de biodiversidade e o amplo potencial do uso de dados para a elaboração de *data papers* referentes à diversidade biológica do país. Tal cenário descrito vai ao encontro das proposições do ODS 14 - Vida na água cuja premissa é a conservação e uso de forma sustentável dos oceanos, mares e recursos marinhos e do ODS 15 - Vida terrestre, que visa a proteção, restauração e promoção do uso sustentável dos ecossistemas terrestres, combatendo a degradação dos solos e perda da biodiversidade (Organização das Nações Unidas Brasil, 2024).

São apresentados a seguir os periódicos que publicaram os artigos analisados neste estudo (Gráfico 2).

Gráfico 2 - Periódicos que publicam *data papers*



Fonte: Elaborado pelos autores (2024).

Constatou-se que, dos oito periódicos analisados, um é brasileiro, o Data Science, da Data in Science Editora LTDA, de Lavras-MG, o qual publicou um

artigo deste estudo; cinco deles tem como foco a Biodiversidade em geral (*Biodiversity Data Journal*, *Gigabyte*, *Arxius de Miscel·lània Zoològica*, *Ecology* e *ZooKeys*) abrangendo subtemas como taxonomia, flora, fauna, etc., e três são multidisciplinares (*Data*, *Latin American Data in Science* e *Scientific Data*). O periódico que mais publicou artigos, 14 mais especificamente, foi o *Biodiversity Data Journal*, voltado para a área das Ciências Biológicas e que desenvolveu a ferramenta de escrita colaborativa denominada *ARPHA Writing Tool* (AWT).

Os periódicos têm em comum o fato de salientarem a adoção dos princípios FAIR, acrônimo para *Findable, Accessible, Interoperable, Reusable*, diretrizes essas que visam a melhoria da Encontrabilidade, Acessibilidade, Interoperabilidade e Reutilização de ativos digitais (GO FAIR, 2025). Outro aspecto relevante é o *Integrated Publishing Toolkit* (IPT), ferramenta de software aberta e gratuita que permite publicar e compartilhar *datasets* de biodiversidade por meio da rede GBIF. Uma instância do IPT, incluindo os dados e metadados registrados por meio do IPT, são conectados ao Registro do GBIF, indexados para consulta por intermédio da rede e do portal do GBIF, sendo disponibilizados para uso público (Global Biodiversity Information Facility, [2025]).

Para uma análise mais aprofundada no GBIF, foram também explorados os *datasets* do GBIF referenciados pelos *data papers* examinados. Ao acessar o link para a busca de *datasets* no GBIF, observamos que existem diversos conjuntos de dados frequentemente atualizados e utilizados por pesquisadores no Brasil. A análise inicial indica uma forte contribuição das instituições públicas brasileiras

na publicação desses *datasets*, os quais abrangem uma variedade de dados de biodiversidade, incluindo flora, fauna e ecossistemas marinhos.

Ao filtrar os estudos por “país ou área do investigador” e selecionar o Brasil, foram localizados 27 estudos, estando todos publicados em forma de artigos. Destes, 26 são de acesso aberto, e 22 identificam o país ou área de cobertura dos dados, sendo 21 artigos cobrindo o Brasil e 1 (um) o Chile. O restante não informou esse dado.

A análise dos artigos identificou 51 conjuntos de dados citados, sendo necessário verificar aqueles que não mostram os DOIs para garantir a rastreabilidade. Constatou-se que três *data papers* não apresentaram o DOI na interface do GBIF em dados referidos no estudo, mas foi possível verificar que cada realizou uma menção ao próprio *dataset* institucional:

1. “Herbarium of the Pontifical Catholic University of Paraná (HUCP), Curitiba, Southern Brazil” (Kersten; Salesbram; Acra, 2017), o qual fez uma análise sobre a coleção em si do *dataset* da instituição PUC, denominado “Herbário da Pontifícia Universidade Católica do Paraná – HUCP”;
2. “World Checklist of Opiliones species (Arachnida). Part 2: Laniatores – Samooidea, Zalmoxoidea and Grassatores incertae sedis” (Kury; Souza, 2021), que traz uma análise do *dataset* publicado por dois pesquisadores do Museu Nacional da Universidade Federal do Rio de Janeiro, que compõe a autoria do artigo;

3. “The retrospective analysis of Antarctic tracking data project” (Ropert-Coudert *et al.*, 2020), apresenta dados do projeto do Comitê Científico para Pesquisa Antártica, disponível no GBIF pelo *Scientific Committee on Antarctic Research* - ademais, os dados são disponibilizados no GitHub e Zenodo.

Outro *data paper* utilizou *dataset* próprio, tendo feito menção a ele. Categorizado como acesso fechado o GBIF, o *data paper* “Wild bees of Chile: a database on taxonomy, sociality, and ecology” (López-Aliste, 2021), fez uso do *dataset* utilizado Wild bees of Chile - The PUCV collection, da Pontificia Universidad Católica de Valparaíso, Chile. O referido documento tem sete autores e um deles é brasileiro, da Escola Politécnica da Universidade de São Paulo (USP) razão pela qual o *data paper* foi recuperado nas buscas.

Portanto, foram identificados 51 *datasets* no total, demonstrando o predomínio do uso de *datasets* secundários em sua maioria, totalizando 47, para a elaboração de *data papers* e quatro elaborados a partir de dados do próprio projeto. Isto ratifica a importância da disponibilização de *datasets* em acesso aberto para uso e reuso pela comunidade científica, evitando a duplicidade de esforços na coleta de dados que já foram coletados por outros pesquisadores, em conformidade com os princípios da Ciência Aberta.

Verificou-se que a instituição com o maior número de citações dos seus *datasets* foi a Embrapa Florestas, que obteve 29 citações oriundas de apenas um *datapaper*, intitulado “Soil macrofauna communities in Brazilian land-use systems” (Brown *et al.*, 2024), publicado no *Biodiversity Data Journal*. A Embrapa Florestas, situada no Paraná, é uma unidade da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) criada em 1978 e que tem um importante papel no que

tange a criação de tecnologias para o setor florestal brasileiro. Seu foco é estabelecer melhorias na eficiência produtiva, reduzir custos de produção, aumentar a oferta de produtos florestais e agrícolas e por fim, conservar o meio ambiente (Embrapa Florestas, 2024).

Obteve destaque também o Instituto Oswaldo Cruz (IOC), com cinco *datasets* citados pelos seguintes *data papers*: 1) “Sampling collections and metadata of planorbidae (Mollusca: Gastropoda) in Brazil: a comprehensive analysis of the Oswaldo Cruz Institute’s Mollusk Collection from 1948 to 2023” (Lima *et al.*, 2023); 2) “Occurrence records and metadata for sand flies (Diptera, Psychodidae, Phlebotominae) collected in the lands of indigenous people in the Brazilian Amazon” (Shimabukuro *et al.*, 2022) o qual utilizou dois *datasets*; “Sand fly (Diptera: Psychodidae: Phlebotominae) records in Acre, Brazil: a dataset Online catalogue of the Coleção de Flebotomíneos” (Godoy *et al.*, 2022); “(FIOCRUZ/COLFLEB), a biological collection of American sand flies (Diptera: Psychodidae, Phlebotominae) held at Fiocruz Minas, Brazil” (Andrade-Filho *et al.* 2022).

O IOC tem como missão realizar pesquisa, ensino, e desenvolvimento tecnológico, além de inovação, serviços de referência e de coleções biológicas, voltados para a promoção da saúde. Por meio dos seus 66 laboratórios, produz conhecimentos sobre variadas doenças, com foco em transmissão, tratamento, controle e prevenção de agravos, dentre estes, AIDS, tuberculose, hepatite, hanseníase, meningite e também doenças tropicais e parasitárias, tais como,

malária, febre amarela, dengue, doença de Chagas, leishmaniose, etc. (Instituto Oswaldo Cruz, 2025). No cenário dos estudos ambientais, desenvolve investigação de fármacos e vacinas, bem como métodos de diagnóstico e de estratégias terapêuticas, com fins de atender as demandas da saúde pública brasileira (Instituto Oswaldo Cruz, 2025).

Outras Instituições brasileiras de renome figuram nas buscas desses *datasets*, com uma mais referências: Instituto Nacional de Pesquisas da Amazônia (INPA); Museu Paraense Emílio Goeldi; Museu Nacional da Universidade Federal do Rio de Janeiro (UFRJ); Universidade de São Paulo (USP); Universidade Federal de Viçosa (UFV); Universidade Federal de Minas Gerais (UFMG); Universidade Federal de Lavras (UFLA); Universidade Federal do Paraná (UFPA); Universidade Federal de Rio Grande (FURG) e Pontifícia Universidade Católica do Paraná.

Em síntese, destaca-se que 6 (seis) Universidades Federais aparecem no ranking; 1 (uma) Universidade Estadual, a Universidade de São Paulo e 1 (uma) Universidade Privada, a Pontifícia Universidade Católica do Paraná. São importantes e renomadas instituições, tanto nacional quanto internacionalmente, as quais se destacam pelos atributos de qualidade do ensino, produção científica, inovação e impacto social.

Os *datasets* analisados são um exemplo de como os dados de biodiversidade são coletados, processados e disponibilizados através do GBIF, permitindo a utilização por pesquisadores ao redor do mundo para estudos ecológicos, monitoramento ambiental e avaliações de impacto (GBIF, 2024).

O filtro aplicado para *data papers* revelou que, 27 registros identificados, indicando uma tendência crescente de publicação de *data papers* pelos pesquisadores brasileiros e de uso dos *datasets* do GBIF. Esses artigos são publicados em periódicos renomados e frequentemente acessados, demonstrando a relevância dos *datasets* do GBIF para a comunidade científica brasileira.

A análise sobre os periódicos que publicam *data papers* de biodiversidade indica que o *Biodiversity Data Journal* (BDJ) é um dos principais canais de divulgação desses artigos. O BDJ facilita a publicação de dados de biodiversidade de várias formas, incentivando o compartilhamento e a reutilização dos dados através de uma plataforma de acesso aberto. Este periódico é altamente reconhecido e frequentemente utilizado para a publicação de *data papers* relacionados a dados de biodiversidade. O BDJ possui uma estrutura que acelera o processo de revisão e publicação, tornando-o uma escolha popular entre os pesquisadores. Além do BDJ, periódicos importantes na área incluem o Gigabyte e outros de alto impacto que também contribuem significativamente para a disseminação científica dos *data papers*.

5 CONSIDERAÇÕES FINAIS

Este estudo destacou a importância e a relevância dos *data papers* na comunicação científica, especialmente no contexto da biodiversidade. A utilização e o compartilhamento de *datasets* por meio do *Global Biodiversity Information Facility* (GBIF) promovem a transparência, reprodutibilidade e

colaboração na pesquisa científica, elementos fundamentais para o avanço do conhecimento. *Data papers* são artigos científicos que descrevem a coleta, o processamento e a qualidade dos dados, oferecendo uma visão detalhada e padronizada dos conjuntos de dados disponibilizados. Já os *datasets* são os conjuntos de dados propriamente ditos, que contêm as informações brutas coletadas durante as pesquisas.

A análise dos *data papers* publicados por pesquisadores brasileiros revelou uma contribuição significativa das universidades brasileiras, particularmente do sudeste do país (Minas Gerais, São Paulo e Rio de Janeiro) para a base de dados do GBIF. Liderando o *ranking* está a Embrapa Florestas, do sul do Brasil, que teve 29 conjuntos de dados usados por apenas um *data paper*. Os *datasets* analisados englobam uma ampla variedade de dados de biodiversidade, incluindo flora, fauna e ecossistemas marinhos, evidenciando o papel crucial dessas instituições na coleta e na disponibilização de informações essenciais para a pesquisa científica.

A identificação de 500 *data papers* no total e após o filtro por Brasil, possibilitou a recuperação de 27 estudos, dos quais apenas um em acesso fechado, reforça a tendência crescente de publicação de *data papers* por pesquisadores brasileiros. A predominância de periódicos como o *Biodiversity Data Journal* (BDJ) como canal de divulgação sublinha a importância deste meio na disseminação científica, facilitando o compartilhamento e a reutilização dos dados. Os resultados obtidos demonstram a necessidade de políticas mais efetivas para a disponibilização de dados de pesquisa, visando facilitar o uso e reúso desses dados por toda a comunidade científica. A implementação dessas políticas pode fortalecer a infraestrutura de pesquisa, aumentando a visibilidade e o impacto das instituições brasileiras no cenário científico global.

O Brasil em função de sua extensão territorial, diversidade geográfica e climática, apresenta vasta diversidade biológica, o que amplifica o seu papel na produção, armazenamento e disseminação de dados sobre biodiversidade.

O acesso livre e a padronização desses dados, registrados pelas instituições brasileiras, incrementam o cenário da pesquisa no país e permitem análises em escala global, fortalecendo redes de cooperação internacional.

Em conclusão, este estudo não apenas sublinha a importância do GBIF como uma plataforma vital para a biodiversidade global, mas também ressalta a contribuição brasileira para essa iniciativa. A continuidade e expansão dessas práticas são essenciais para promover avanços científicos e enfrentar os desafios relacionados à biodiversidade de maneira colaborativa e inclusiva.

REFERÊNCIAS

- ANDRADE-FILHO, José Dilermando *et al.* Online catalogue of the Coleção de Flebotomíneos (FIOCRUZ/COLFLEB), a biological collection of American sand flies (Diptera: Psychodidae, Phlebotominae) held at Fiocruz Minas, Brazil. **Gigabyte**, [s.l.], May 2022, online. Disponível em: <https://doi.org/10.46471/gigabyte.52>. Acesso em: 08 jun. 2024
- ARAÚJO, D. G. S.; DIAS, G. A.; AUTRAN, M. M. M. Compartilhamento de dados no contexto da ciência brasileira: um estudo integrativo. **Informação & Informação**, Londrina, v. 26, n. 3, p. 100, 2021. Disponível em: <https://doi.org/10.5433/1981-8920.2021v26n3p100>. Acesso em: 08 jun. 2024
- AVENTURIER, P.; ALENCAR, M. C. de. Os desafios de dados de pesquisa abertos. **RECIIS**, [s.l.], v. 10, n. 3, 2016. Disponível em: <https://doi.org/10.29397/reciis.v10i3.1069>. Acesso em: 08 jun. 2024
- BIODIVERSITY DATA JOURNAL - BDJ. **Foco e escopo**. [S.l.], 2025. Disponível em: <https://bdj.pensoft.net/about>. Acesso em: 08 jun. 2024.
- BROWN, G. G. *et al.* Soil macrofauna communities in Brazilian land-use systems. **Biodiversity Data Journal**, [s.l.], v. 12, e115000, 2024. Disponível em: <https://doi.org/10.3897/BDJ.12.e115000>. Acesso em: 08 jun. 2024.

CABALLERO-RIVERO, A.; SÁNCHEZ-TARRAGÓ, N.; SANTOS, R.N.M. Práticas de Ciência Aberta da comunidade acadêmica brasileira: estudo a partir da produção científica.

Transinformação, [s.l.], v.31, e190029, 2019. Disponível em: <http://dx.doi.org/10.1590/2318-0889201931e190029>. Acesso em: 08 jun. 2024.

DATA IN SCIENCE. **Saber mais sobre o jornal**. [S.l.], 2025. Disponível em: <https://www.datainscience.com.br/>. Acesso em: 08 jun. 2024.

EMBRAPA FLORESTAS. **Apresentação**. [S.l.], 2025. Disponível em: <https://www.embrapa.br/florestas/apresentacao>. Acesso em: 08 jun. 2024.

FOX, Richard *et al.* Opinions of citizen scientists on open access to UK butterfly and moth occurrence data. **Biodiversity and Conservation**, [s.l.], v. 28, p. 3321-3341, 2019. Disponível em: <https://link.springer.com/article/10.1007/s10531-019-01824-6#citeas>. Acesso em: 10 set. 2024.

GAY, V. Un data paper en SHS : pourquoi, pour qui, comment? In: Christine Kosmopoulos; Joachim Schöpfel. **Publier, partager, réutiliser les données de la recherche: les data papers et leurs enjeux**. [S. l.]: [S. n.], 2021. Disponível em: <https://hal.archives-ouvertes.fr/hal-03434216>. Acesso em: 08 jun. 2024.

GLOBAL BIODIVERSITY INFORMATION FACILITY - GBIF. **Site institucional** [artigos de dados]. 2024. Disponível em: https://www.gbif.org/pt/resource/search?contentType=literature&topics=DATA_PAPER&relevance=GBIF_PUBLISHED. Acesso em: 08 jun. 2024.

GLOBAL BIODIVERSITY INFORMATION FACILITY - GBIF. **IPT: The Integrated Publishing Toolkit**. [S.l.], [2025]. Disponível em: <https://www.gbif.org/pt/ipt>. Acesso em: 08 abr. 2025.

GODOY, Rodrigo Espindola *et al.* Sand fly (Diptera: Psychodidae: Phlebotominae) records in Acre, Brazil: a dataset. **Gigabyte**, [S.l.], May 2022. Disponível em: <https://doi.org/10.46471/gigabyte.60>. Acesso em: 08 jun. 2024.

INSTITUTO OSWALDO CRUZ (IOC/Fiocruz). **Site institucional**. Rio de Janeiro, 2025. Disponível em: <https://www.ioc.fiocruz.br/>. Acesso em: 08 abr. 2025.

KERSTEN, R. A.; SALESBRAM, J. A. M.; ACRA, L. A. Herbarium of the Pontifical Catholic University of Paraná (HUCP), Curitiba, Southern Brazil. **DATA**, [s.l.], v. 2, n. 10, 2017. Disponível em: <https://doi.org/10.3390/data2010010>. Acesso em: 08 jun. 2024.

KURY, Adriano; SOUZA, Daniele. World Checklist of Opiliones species (Arachnida). Part 2: Laniatores – Samooidea, Zalmoxoidea and Grassatores incertae sedis. **Biodiversity Data Journal**, [s.l.], v. 3, e6482, 2021. Disponível em: <https://doi.org/10.3897/BDJ.3.e6482>. Acesso em: 08 jun. 2024.

LÓPEZ-ALISTE, Manuel *et al.* Wild bees of chile: a database on taxonomy, sociality, and ecology. **Ecological Society of América**, [s.l.], v. 102, issue 8, e03377, August 2021. Disponível em: 08 jun. 2024.

em: <https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.3377>. Acesso em: 08 jun. 2024.

LUSTOSA, M. M.; FARIAS, M. G. G.; FARIAS, G. B. Inteligência artificial e comunicação científica. **Brazilian Journal of Information Science: Research Trends**, Marília, v. 18, e024004, 2024. Disponível em: <https://doi.org/10.36311/1981-1640.2024.v18.e024004>. Acesso em: 08 jun. 2024.

NASSI-CALÒ, Lilian. **Movimento Open Data se consolida internacionalmente**. [s.l.], 2014. Disponível em: <https://blog.scielo.org/blog/2014/07/14/movimento-open-data-se-consolida-internacionalmente/>. Acesso em: 08 jun. 2024.

PINHEIRO, Lena Vania R. Comunidades científicas e infra-estrutura tecnológica no Brasil para uso de recursos eletrônicos de comunicação e informação na pesquisa. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 62-73, set. 2003. Disponível em: <https://doi.org/10.1590/s0100-19652003000300008>. Acesso em: 08 jun. 2024.

ROCHA, Luana M. Pinto. **Os cientistas e a ciência cidadã: um estudo exploratório sobre a visão dos pesquisadores profissionais na experiência brasileira**. 2019. Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019. Disponível em: <https://ridi.ibict.br/bitstream/123456789/1053/1/dissertacao-final-LuanaRocha-Ciencia%20cidadada%20e%20cientistas%20profissionais.pdf>. Acesso em: 10 set. 2024.


ROPERT-COUDERT, Yan *et al.* The retrospective analysis of Antarctic tracking data project. **Scientific Data**, [s.l.], v. 7, article number 94, 2020. Disponível em: <https://www.nature.com/articles/s41597-020-0406-x>. Acesso em: 08 jun. 2024.

SCHÖPFEL, J., et al. Data Papers as a new form of knowledge organization in the field of research data. **Knowledge Organization**, [s.l.], v. 46, n. 8, p. 622–38, 2020. Disponível em: <https://doi.org/10.5771/0943-7444-2019-8-622>. Acesso em: 08 jun. 2024.

SHIMABUKURO, Paloma Helena Fernandes et al. Occurrence records and metadata for sand flies (Diptera, Psychodidae, Phlebotominae) collected in the lands of indigenous people in the Brazilian Amazon. **Gigabyte**, [S.l.], May 2022. Disponível em: <https://doi.org/10.46471/gigabyte.61>. Acesso em: 08 jun. 2024.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS - ONU BRASIL. **Os Objetivos de Desenvolvimento Sustentável no Brasil**. Brasília, [2024]. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 15 jun. 2024.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA - UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. Brasília, DF: Representação da UNESCO no Brasil, 2022. [36].

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 



 tpbci@ancib.org

 [@anciboficial](https://www.instagram.com/anciboficial)