



Encontros Bibli

CONJUNTO DE DADOS SOBRE GESTÃO DE DADOS DE PESQUISA NA BIODIVERSIDADE: PRÁTICAS, DESAFIOS E PERSPECTIVAS

Dataset on Research Data Management in Biodiversity: Practices, Challenges, and Perspectives

Paulo Roberto Santos Costa

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência Da Informação,
João Pessoa, PB, Brasil
prbcosta@gmail.com

<https://orcid.org/0000-0001-5833-0676> 

Janaína Nascimento de Araújo

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência Da Informação,
João Pessoa, PB, Brasil
janabiblio@gmail.com

<https://orcid.org/0009-0007-1105-3000> 

Guilherme Ataíde Dias

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência Da Informação,
João Pessoa, PB, Brasil
guilhermeataide@gmail.com

<https://orcid.org/0000-0001-6576-0017> 

Marckson Roberto Ferreira de Sousa

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência Da Informação,
João Pessoa, PB, Brasil
marckson.dci.ufpb@gmail.com

<https://orcid.org/0000-0003-2001-1631> 

Anderson Rafael Castro Simões

Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência Da Informação,
João Pessoa, PB, Brasil
arafaelcs@gmail.com

<https://orcid.org/0000-0002-8268-2783> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: Descrever, documentar e disponibilizar um conjunto de dados sobre práticas, desafios e percepções relacionadas à Gestão de Dados de Pesquisa no contexto da biodiversidade, visando apoiar estudos em Ciência Aberta, gestão da informação científica e formulação de políticas institucionais de dados.

Método: Realizou-se um levantamento por meio de questionário on-line, implementado no Google Forms, aplicado no período de 09/10/2024 a 18/12/2024, a 653 pesquisadores vinculados a programas de pós-graduação na área da biodiversidade de universidades públicas brasileiras avaliados com notas 6 e 7 pela CAPES. Foram obtidas 150 respostas válidas (taxa de retorno de 22,9%). Os dados foram preservados em formato bruto, estruturados em arquivo CSV e depositados no repositório DATAPB, com identificador DOI.

Resultados: O dataset resultante compreende 21 variáveis que caracterizam o perfil demográfico dos participantes e suas práticas de Gestão de Dados de Pesquisa, incluindo familiaridade com planos de gestão de dados, adoção de padrões de metadados, uso de repositórios, verificação da qualidade, integração e análise de dados, bem como obstáculos técnicos e institucionais ao compartilhamento.

Conclusões: O conjunto de dados constitui uma fonte relevante para análises sobre maturidade em Gestão de Dados de Pesquisa na área de biodiversidade, podendo subsidiar estratégias institucionais, ações de capacitação, aprimoramento de repositórios e políticas de interoperabilidade, além de contribuir para pesquisas em Ciência da Informação e Ciência Aberta.

PALAVRAS-CHAVE: Gestão de dados de pesquisa. Biodiversidade. Ciência aberta. Repositórios de dados.

ABSTRACT

Objective: To describe, document, and make available a dataset on practices, challenges, and perceptions related to Research Data Management in the context of biodiversity, aiming to support studies in Open Science, scientific information management, and the formulation of institutional data policies.

Methods: A survey was conducted using an online questionnaire implemented in Google Forms and administered between October 9, 2024, and December 18, 2024, to 653 researchers affiliated with graduate programs in biodiversity at Brazilian public universities rated 6 and 7 by CAPES. A total of 150 valid responses were obtained (22.9% response rate). The data were preserved in raw format, structured in a CSV file, and deposited in the DATAPB repository with a DOI identifier.

Results: The resulting dataset comprises 21 variables that characterize the demographic profile of participants and their Research Data Management practices, including familiarity with data management plans, adoption of metadata standards, use of repositories, quality assurance procedures, data integration and analysis, as well as technical and institutional barriers to data sharing.

Conclusions: The dataset constitutes a relevant source for analyses of Research Data Management maturity in the field of biodiversity, supporting institutional strategies, training initiatives, repository improvement, and interoperability policies, in addition to contributing to research in Information Science and Open Science.

KEYWORDS: Research data management. Biodiversity. Open science. Data repositories.

1 APRESENTAÇÃO

A Gestão de Dados de Pesquisa (GDP) é um tema importante na era da Ciência Aberta, podendo ser aplicada a diversas áreas do conhecimento. A Ciência Aberta promove princípios de transparência e colaboração entre áreas do conhecimento, incentivando práticas que tornam os dados mais acessíveis e reutilizáveis (Silva, 2019). No entanto, a implementação de boas práticas de GDP enfrenta desafios, como a falta de infraestrutura para pesquisa, limitações técnicas e resistência cultural. Segundo Cox e Verbaan (2018), essas práticas envolvem o conjunto de processos e políticas que são implementadas para garantir que os dados integrantes de uma atividade de pesquisa possam ser gerenciados de forma eficiente, assegurando sua integridade, acesso e reprodutibilidade ao longo do seu ciclo de vida.

Os mesmos autores evidenciam a necessidade de o pesquisador desenvolver habilidades para a plena execução das práticas de GDP. Vilar e Zabukovec (2018), ao investigarem o estado do tema em seu país (Eslovênia), destacam a importância do desenvolvimento desta competência, evidenciando também que a evolução a esta formação precisa melhorar em muitas áreas. Possivelmente, começando no nível acadêmico institucional e envolvendo esforços continuados em políticas de conscientização e normatização. Segundo os autores, todos os envolvidos na pesquisa, desde os cientistas

até o pessoal de apoio, precisam compreender e estar preparados para implementar essas práticas.

No contexto da biodiversidade, ressalta-se que os dados de pesquisa possuem uma natureza diversificada e complexa, podendo atingir volumes significativos e englobando registros ecológicos, genéticos e geográficos. Estes dados subsidiam pesquisas científicas, políticas públicas e iniciativas de conservação. Apesar disso, muitos pesquisadores enfrentam dificuldades para alinhar suas práticas no âmbito da Ciência Aberta (Gadelha *et al.*, 2020).

Este artigo descreve um conjunto de dados (*dataset*) coletado de pesquisadores vinculados a programas de pós-graduação em Biodiversidade de universidades públicas brasileiras, especificamente daqueles cursos avaliados com notas 6 e 7 pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

A seguir, são evidenciados alguns aspectos que explicam o porquê destes dados de pesquisa serem valiosos para a comunidade científica na área da biodiversidade:

- **Por que esses dados são úteis?** Os dados coletados refletem diretamente às percepções e práticas dos pesquisadores quanto à gestão de seus próprios dados de pesquisa. A partir das respostas ao instrumento de pesquisa, é possível identificar a familiaridade dos pesquisadores da área da biodiversidade com o tema GDP, especialmente no que diz respeito a: planos de gestão de dados; padrões de metadados; utilização de repositórios de dados; obstáculos ao compartilhamento de dados; plataformas para manipulação e análise de dados; e os desafios na questão integração de dados. Ainda são possíveis de serem identificadas prováveis lacunas e barreiras que limitam a adoção de padrões e ferramentas de GDP. Além disso, busca-se contribuir de forma indireta no aprimoramento das práticas de GDP, fomentando discussões sobre estratégias mais eficazes para o avanço da pesquisa e da colaboração interdisciplinar no campo da biodiversidade.
- **Quem pode se beneficiar desses dados?** Pesquisadores da área de biodiversidade: podem usar as informações para entender e melhorar suas próprias práticas de gestão de dados, alinhando-as aos princípios da ciência aberta; Gestores de programas de pós-graduação: podem identificar necessidades de capacitação e implementar políticas que incentivem a adoção de boas práticas de gestão de dados;

Gestores públicos e acadêmicos: podem usar os dados para desenvolver políticas públicas e institucionais que promovam a transparência e a interoperabilidade de dados; Desenvolvedores de ferramentas e plataformas de software: podem adaptar tecnologias às necessidades específicas dos pesquisadores da área, facilitando o compartilhamento e a análise de dados; Estudantes e educadores: podem utilizá-los como material didático ou referência em cursos sobre Gestão de Dados de Pesquisa e Ciência Aberta.

- **Como esses dados podem ser usados para mais *insights* e desenvolvimento de experimentos?** Identificação de lacunas: a análise pode revelar processos com menor adoção de boas práticas de GDP; Comparações regionais: os dados podem ser utilizados para avaliar diferenças geográficas nas práticas de GDP; Desenvolvimento de soluções personalizadas: insights sobre os principais desafios enfrentados, como falta de infraestrutura ou políticas adequadas, podem embasar a criação de ferramentas, repositórios ou metodologias específicas para a área da Biodiversidade; Modelagem de experimentos colaborativos: os dados podem ajudar a formular hipóteses sobre como mudanças em políticas ou práticas afetam o compartilhamento e a reutilização de dados; Estudos futuros: podem servir como uma linha de base para estudos futuros, permitindo avaliar o impacto de iniciativas voltadas à melhoria da gestão de dados no campo da Biodiversidade.

Além de sua relevância para a área da biodiversidade, este conjunto de dados também oferece subsídios importantes para a Ciência da Informação, especialmente no que tange à gestão de dados de pesquisa. Ele pode contribuir para a formulação de estratégias institucionais para repositórios de dados, o aprimoramento de metadados descritivos e a promoção de boas práticas de compartilhamento e preservação de dados científicos. O estudo reforça a importância da interoperabilidade de dados e da adoção de padrões na Ciência Aberta, ampliando o impacto da pesquisa para além do domínio da biodiversidade.

2 MÉTODOS E INSTRUMENTOS

Nesta seção, apresentamos os procedimentos metodológicos adotados, detalhando o contexto da coleta de dados, os instrumentos utilizados e aspectos éticos relacionados. Para garantir a representatividade da amostra e a robustez dos dados coletados, utilizamos critérios específicos na seleção dos participantes e recorremos a ferramentas que possibilitaram a obtenção de dados qualificados sobre as práticas e desafios da GDP.

2.1 Contexto da Coleta

Os dados foram coletados junto a pesquisadores vinculados a programas de pós-graduação em biodiversidade em universidades públicas brasileiras. Esses pesquisadores estão associados a instituições de diversas regiões do país, contribuindo para uma visão abrangente das práticas e desafios enfrentados pela GDP no Brasil.

Os cursos escolhidos compreenderam os que possuíam notas 6 e 7 segundo avaliação da CAPES, em sua avaliação quadrienal 2021-2024, considerados de excelência nacional. A identificação dos cursos a serem alvos da pesquisa foi realizada através da Plataforma Sucupira (<https://sucupira.capes.gov.br/>), utilizando critérios específicos de filtragem:

- Nota CAPES: apenas programas com notas 6 e 7, que indicam excelência acadêmica.
- Modalidade de Programa: acadêmico, excluindo programas profissionais.
- Área de Avaliação: biodiversidade, garantindo a aderência dos programas ao escopo de interesse.
- Modalidade de Ensino: educação presencial, desconsiderando cursos na modalidade a distância.
- Situação: em funcionamento, assegurando que os programas listados se encontram ativos.
- Com base nesses filtros, a pesquisa realizada em 23/09/2024 relacionou diversos programas de pós-graduação *stricto sensu* distribuídos em diferentes instituições de ensino superior do Brasil, conforme apresentado no Quadro 1.

A seleção de programas de pós-graduação com notas 6 e 7 pela CAPES garante que os participantes da pesquisa representem instituições com forte impacto na produção científica. Essa escolha permite uma análise mais qualificada sobre as práticas de gestão de dados de pesquisa, refletindo as políticas e desafios enfrentados por pesquisadores inseridos em programas de excelência acadêmica.

Quadro 1 - Retorno da pesquisa submetida à Plataforma Sucupira

Instituição de Ensino Superior	Programa de Pós-Graduação
INPA	BIOLOGIA (ECOLOGIA)
PUC/RS	ECOLOGIA E EVOLUÇÃO DA BIODIVERSIDADE
UEM	ECOLOGIA DE AMBIENTES AQUÁTICOS CONTINENTAIS
UERJ	ECOLOGIA E EVOLUÇÃO
UESC	ECOLOGIA E CONSERVAÇÃO DA BIODIVERSIDADE
UFG	ECOLOGIA E EVOLUÇÃO
UFLA	ECOLOGIA APLICADA
UFMG	BIOLOGIA VEGETAL
UFMG	ECOLOGIA (CONSERVAÇÃO E MANEJO DA VIDA SILVESTRE)
UFPE	BIOLOGIA ANIMAL
UFPE	BIOLOGIA DE FUNGOS
UFPE	BIOLOGIA VEGETAL
UFPR	ECOLOGIA E CONSERVAÇÃO
UFRGS	ECOLOGIA
UFRJ	BIODIVERSIDADE E BIOLOGIA EVOLUTIVA
UFRJ	CIÊNCIAS BIOLÓGICAS (ZOOLOGIA)
UFRJ	ECOLOGIA
UFRN	ECOLOGIA
UFSM	BIODIVERSIDADE ANIMAL
UNB	ECOLOGIA
UNESP-BOT	BIOLOGIA VEGETAL
UNESP-BOT	CIÊNCIAS BIOLÓGICAS (ZOOLOGIA)
UNESP-RC	ECOLOGIA, EVOLUÇÃO E BIODIVERSIDADE
UNICAMP	BIOLOGIA VEGETAL
USP	CIÊNCIAS BIOLÓGICAS (BOTÂNICA)
USP	CIÊNCIAS BIOLÓGICAS (ZOOLOGIA)
USP	ECOLOGIA
USP/RP	ENTOMOLOGIA
USP/RP	BIOLOGIA COMPARADA

Fonte: Elaborado pelos autores (2025).

A partir da identificação dos programas de pós-graduação na área de Biodiversidade, foi realizada a coleta de dados junto a pesquisadores vinculados a esses cursos. Na próxima seção são detalhados os dados coletados.

2.2 Coleta de Dados

A coleta de dados foi realizada por meio de um questionário *online*, utilizando a plataforma *Google Forms*, entre 09/10/2024 e 18/12/2024, sendo submetida a um total de 653 pesquisadores e obtendo 150 respondentes (taxa de 22,9% de respondentes). Dentre os respondentes:

- 141 responderam (94%): Declaro que fui devidamente esclarecido (a) sobre a pesquisa e dou o meu consentimento. Estou ciente que receberei uma cópia das informações preenchidas neste formulário.
- 9 responderam (6%): Declaro que fui devidamente esclarecido (a) sobre a pesquisa, contudo opto pela não participação.

O instrumento incluiu 21 perguntas de caráter quantitativo e/ou qualitativo, abrangendo os seguintes temas:

- Perfil demográfico do pesquisador: Unidade Federativa (UF) da sua principal instituição de pesquisa; Natureza da sua instituição; Realização de estágio pós-doutoral; Faixa etária do pesquisador; e Tempo como pesquisador.
- Percepção sobre Gestão de Dados de Pesquisa: Familiaridade e utilização de PGDs; Adoção de padrões de metadados; Frequência de utilização de repositórios de dados; Obstáculos enfrentados na disponibilização de dados; e Desafios na integração e análise de dados.

A Figura 1, apresenta o fluxo envolvendo a coleta e disponibilização do dataset no repositório de dados DATAPB (<https://dataverse.ideal.ufpb.br/>). O processo de coleta envolveu o envio do formulário de pesquisa em 4 momentos distintos, no intuito de atingirmos a quantidade satisfatória de respondentes (acima de 20%), segundo Nulty

(2008). Ressalta-se que o reenvio do questionário excluiu os respondentes do envio anterior.

Figura 1 - Fluxo envolvendo a coleta e publicação dos dados de pesquisa



Fonte: Elaborado pelos autores (2025).

O processo de coleta foi documentado de maneira detalhada para permitir sua replicação em estudos futuros. Os critérios de seleção dos participantes, os filtros aplicados na Plataforma Sucupira e a estrutura do questionário estão descritos integralmente no *dataset* publicado, permitindo que outros pesquisadores repliquem ou ampliem a investigação em diferentes contextos institucionais.

Para garantir a transparência e a conformidade da pesquisa, foram observados aspectos éticos relacionados à participação dos respondentes, evidenciados na seção subsequente.

2.3 Aspectos Éticos

A coleta seguiu as normas éticas em pesquisa, garantindo a confidencialidade e a anonimização das respostas. Todos os participantes foram informados sobre os objetivos do estudo, sua confidencialidade e o direito de recusa, assegurando a conformidade com princípios éticos na pesquisa científica.

A pesquisa, pelo fato de envolver a aplicação de questionários a seres humanos, foi submetida em 14/10/2024 e aprovada em 06/12/2024 pelo Comitê de Ética em Pesquisa da Universidade Federal da Paraíba - UFPB com (Número do parecer: 7.286.870). Sua submissão se deu através da Plataforma Brasil (<https://plataformabrasil.saude.gov.br/>), detentora de uma base nacional unificada que registra e acompanha todas as pesquisas

que envolvem seres humanos, garantindo a transparência e o acesso aos dados públicos das pesquisas aprovadas.

Para garantir a privacidade dos participantes, os dados foram anonimizados antes do depósito no repositório DATAPB. Informações pessoais, como endereços de e-mail, foram removidas ou codificadas, assegurando que nenhuma identificação direta ou indireta fosse possível. Além disso, os dados foram revisados para evitar a presença de padrões que pudessem levar à identificação dos respondentes.

3 TABELA DE ESPECIFICAÇÕES

Os dados coletados para esta pesquisa estão vinculados à área de Ciência da Informação, com foco específico em Gestão de Dados de Pesquisa. No Quadro 2, são evidenciadas as características gerais do *dataset*:

Quadro 2 - Características gerais do *dataset*

Área de Conhecimento	Ciência da Informação
Área de assunto específica	Gestão de Dados de Pesquisa
Idioma	Português
Tipo de dados	Arquivo CSV em UTF-8, delimitado por vírgula
Tamanho	109 KB
Como os dados foram adquiridos	Questionário online via Google Forms
Estado dos dados	Brutos
Parâmetros para coleta de dados	Cursos presenciais ativos de pós-graduação (acadêmico) em biodiversidade, presenciais, que possuem notas 6 e 7 segundo avaliação da CAPES, em sua avaliação quadrienal 2021-2024 (data da consulta: 23/09/2024)
Descrição da coleta de dados	Questionário com 20 perguntas enviado por e-mail em 4 etapas para um total de 653 pesquisadores, com 150 respostas obtidas.
Localização da fonte de dados	https://doi.org/10.48472/DATAPB/YDW3IM
Acessibilidade de dados	Nome do repositório de dados: DATAPB (https://dataverse.ideal.ufpb.br/) Número de identificação de dados: 10.48321/D17E7F63DD
Artigo de pesquisa relacionado	Gestão de Dados de Pesquisa: uma investigação sob a perspectiva da Biodiversidade, no contexto da Ciência Aberta

Fonte: Elaborado pelos autores (2025).

Este conjunto de dados foi estruturado seguindo os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), garantindo sua localização e recuperação por meio de um DOI permanente, seu armazenamento em um repositório de acesso aberto (DATAPB), e a adoção de um formato amplamente interoperável (CSV). A documentação do *dataset* segue padrões reconhecidos, assegurando que as informações possam ser reutilizadas por diferentes comunidades de pesquisa e integradas a outros sistemas de informação científica.

3.1 Descrição do conjunto de dados

O Quadro 3, apresenta a estrutura detalhada dos dados coletados na pesquisa, detalhando os campos do questionário e suas respectivas categorias. As respostas foram categorizadas de acordo com escalas de frequência, múltipla escolha ou através de segmentação etária e temporal.

Quadro 3 - Estrutura detalhada do *dataset*

Campo	Comentário	Tipo
Carimbo de data/hora	Data e hora na qual a resposta foi enviada pelo respondente.	Data, Hora, Minuto e Segundo
Nome de usuário	E-mail do respondente. Por questões de sigilo, este campo teve seu conteúdo anonimizado.	Texto
Consentimento de participação na pesquisa	Sinaliza o posicionamento do respondente, no tocante à aceitação ou não de participar da pesquisa.	Categórica
Unidade Federativa (UF) da sua principal instituição de pesquisa?	UF na qual o pesquisador possui maior vínculo	Categórica
Natureza da sua instituição?	Evidencia se a instituição do pesquisador é pública estadual ou federal	Categórica
Você já realizou estágio pós-doutoral?	Sim ou Não para a confirmação se o pesquisador realizou ou não estágio pós-doutoral	Categórica
Faixa etária do pesquisador(a)?	Faixa etária seguindo o critério de segmentação: 18-25 anos, 26-35 anos, 36-45 anos, 46-55 anos, 56-65 anos e acima de 65 anos	Categórica

Campo	Comentário	Tipo
Tempo como pesquisador(a)?	Faixa de tempo seguindo o critério de segmentação: 0-5 anos, 6-10 anos, 11-15 anos, 16-20 anos e acima de 20 anos	Categórica
Você está familiarizado com a elaboração do Plano de Gestão de Dados (PGD) para suas pesquisas ?	A escala de respostas apresentada varia do nível mais elevado de familiaridade, representado por "Estou plenamente familiarizado com o tema", até o nível mais baixo, indicado por "Desconheço totalmente o tema".	Categórica
Suas pesquisas incluem um Plano de Gestão de Dados (PGD) ?	A escala de respostas apresentada varia do nível mais frequente de ocorrência, representado por "Sempre", até a ausência completa de ocorrência ou conhecimento, indicada por "Nunca" e "Não sei informar".	Categórica
Qual o recurso mais comumente utilizado para obter seus dados de pesquisa?	As opções de resposta são: Dados laboratoriais; Literatura científica; Pesquisa de campo; Repositório de dados; Simulações computacionais; Não sei informar.	Categórica
Você enfrenta dificuldades na coleta dos seus dados de pesquisa?	A escala de respostas apresentada varia do nível mais elevado de frequência ou ocorrência, representado por "Sempre", até a ausência total, indicada por "Nenhuma" e "Não sei informar"	Categórica
Você realiza verificações para garantir a qualidade dos dados coletados (ex.: precisão taxonômica ou geográfica)?	A escala de respostas apresentada varia do nível mais elevado de frequência ou ocorrência, representado por "Sempre", até a ausência total, indicada por "Nenhuma" e "Não sei informar"	Categórica
Como você lida com erros identificados nos dados, por exemplo em relação à identificação incorreta de espécies ou localização geográfica imprecisa?	As opções de resposta são: Corrijo manualmente; Outros membros da equipe tratam estes erros; Utilizo ferramentas automatizadas; Identifico a existência de erros, mas não os corrijo; Não sei informar como os erros são tratados; Os erros não são tratados.	Categórica,
O que você acha sobre a adoção de padrões de metadados para descrever seus dados de pesquisa?	A escala de respostas apresentada varia do nível mais consistente de adoção, representado por "Adoto sempre", até a ausência completa, indicada por "Não adoto" e "Não sei informar"	Categórica
Com que frequência você documenta e descreve os dados de pesquisa para que possam ser reutilizados por outros(as) pesquisadores(as) ?	A escala de respostas apresentada varia do nível mais frequente de ocorrência, representado por "Sempre", até a ausência completa de ocorrência ou conhecimento, indicada por "Nunca" e "Não sei informar".	Categórica

Campo	Comentário	Tipo
Os repositórios de dados são importantes para compartilhar seus dados de pesquisa.	A escala de respostas apresentada varia do nível mais elevado de importância atribuída, representado por "Acho extremamente importantes", até a ausência de importância ou incerteza, indicada por "Não acho importante" e "Não sei informar"	Categórica
Quais obstáculos você encontra ao disponibilizar seus dados de pesquisa em repositórios? Selecione as opções que se aplicam.	As opções de resposta são: Custo para publicação dos dados; Falta de políticas éticas e legais para o uso dos dados; Falta de conhecimento técnico; Falta de infraestrutura de pesquisa; Interpretação inadequada dos dados provenientes da minha pesquisa; Nenhum obstáculo; Não sei informar.	Categórica
Ao fazer uma pesquisa com que frequência você reutiliza dados disponíveis em repositórios de dados de biodiversidade (exs.: GBIF, SIBBr)?	A escala de respostas apresentada varia do nível mais frequente de ocorrência, representado por "Sempre", até a ausência completa de ocorrência ou conhecimento, indicada por "Nunca" e "Não sei informar".	Categórica
Existem dificuldades para acessar dados de pesquisa em repositórios de dados?	A escala de respostas apresentada varia do nível mais frequente de ocorrência, representado por "Sempre", até a ausência completa de ocorrência ou conhecimento, indicada por "Nunca" e "Não sei informar".	Categórica
Que desafios você enfrenta ao integrar dados de diferentes fontes (ex.: coleções biológicas, dados moleculares, sensores)?	As opções de resposta são: Conhecimento técnico; Falta de documentação; Formatos de dados incompatíveis; Falta de ferramentas de integração; Problemas de interoperabilidade; Nenhum desafio; Não sei informar.	Categórica
Você utiliza ferramentas específicas para garantir a compatibilidade dos dados que integra em suas pesquisas?	A escala de respostas apresentada varia do nível mais frequente de ocorrência, representado por "Sempre", até a ausência completa de ocorrência ou conhecimento, indicada por "Nunca" e "Não sei informar".	Categórica
Quais as plataformas de software que você utiliza para analisar dados de pesquisa?	As opções de resposta são: Ferramentas de Business Discovery (Tableau, PowerBI, QlikSense, Looker, etc); Plataformas para Processamento de Dados Geográficos (ArcGIS, QGIS, PostGIS, etc); Linguagens de Programação (Python, C, Java, etc); Pacotes Estatísticos (R, SPSS, MATLAB, SATA, etc); Pacotes de Automação de Escritório (MS-Office, LibreOffice, etc); Plataformas específicas para Biodiversidade; Não utilizo ferramentas de software.	Categórica

Campo	Comentário	Tipo
Qual o principal desafio que você enfrenta na análise de dados de pesquisa?	As opções de resposta são: Falta de capacidade computacional; Falta de treinamento técnico; Complexidade dos dados; Não enfrento desafios; Não sei informar.	Categórica

Fonte: Elaborado pelos autores (2025).

Na Figura 2, apresentamos um exemplo das primeiras linhas do *dataset*, evidenciando a estrutura dos dados disponibilizados:

Figura 2 – Trecho do *dataset* resultante da pesquisa

```

1 Carimbo de data/hora, Nome de usuário, Consentimento de participação na pesquisa,
2 "2024/10/09 9:24:26 AM GMT-3", "email1@dominio.com", "Declaro que fui devidamente
3 "2024/10/09 9:29:10 AM GMT-3", "email2@dominio.com", "Declaro que fui devidamente
4 "2024/10/09 10:04:50 AM GMT-3", "email3@dominio.com", "Declaro que fui devidament
5 "2024/10/09 11:48:45 AM GMT-3", "email4@dominio.com", "Declaro que fui devidament
6 "2024/10/09 12:15:10 PM GMT-3", "email5@dominio.com", "Declaro que fui devidament
7 "2024/10/09 12:26:28 PM GMT-3", "email6@dominio.com", "Declaro que fui devidament

```

Fonte: Elaborado pelos autores (2025).

O *dataset* completo está disponível no repositório DATAPB (DOI: <https://doi.org/10.48472/DATAPB/YDW3IM>).

O conjunto de dados apresentado reflete um panorama detalhado sobre as práticas e desafios enfrentados por pesquisadores da área de biodiversidade no contexto da GDP. Os dados coletados possibilitam uma análise abrangente sobre a familiaridade com PGD, estratégias de coleta, verificação e compartilhamento de dados, além de obstáculos técnicos e institucionais. Esse *dataset* representa uma base para futuras investigações envolvendo os temas Ciência Aberta e Gestão de Dados de Pesquisa.

3.2 Visão geral do *dataset*

O *dataset* contém respostas de 150 pesquisadores, dos quais 94% (141) consentiram em participar e 6% (9) optaram por não participar.

No tocante ao perfil dos entrevistados, as respostas abrangem 13 Unidades Federativas, com destaque para RJ (21,3%, 30 respostas), SP (20,6%, 29 respostas) e MG (11,3%, 16 respostas); quanto ao perfil, 74,5% (105) dos participantes têm idade acima de 45 anos, e 78,1% (110) possuem mais de 15 anos de experiência.

No que trata sobre aspectos relacionados à GDP: sobre a familiaridade com Planos de Gestão de Dados (PGD), 22,0% (31) declararam desconhecer o tema, enquanto 20,6%

(29) nunca incluem um PGD em suas pesquisas; a pesquisa de campo é a fonte de dados mais comum (58,2%, 82 respostas), e 92,2% (130) dos pesquisadores verificam sempre a qualidade dos dados, com 71,6% (101) corrigindo erros manualmente; apenas 18,4% (26) reutilizam dados de repositórios sempre, e os obstáculos mais citados ao compartilhamento incluem falta de conhecimento técnico (17,9%, 40) e infraestrutura (19,3%, 43); e pacotes estatísticos (31,4%, 100) e plataformas geoespaciais (23,0%, 73) são as ferramentas mais usadas para análise.

REFERÊNCIAS

- COX, A.; VERBAAN, E. **Exploring research data management**. [s. l.]: Facet Publishing, 2018.
- GADELHA, L. M. R. *et al.* **A survey of biodiversity informatics: Concepts, practices, and challenges**. WIREs Data Mining and Knowledge Discovery, v. 11, n. 1, 2020. Disponível em: <https://doi.org/10.1002/widm.1394>. Acesso em: 15 dez. 2024.
- NULTY, D. D. The adequacy of response rates to online and paper surveys: What can be done? **Assessment & Evaluation in Higher Education**, v. 33, n. 3, p. 301–314, 2008. Disponível em: <https://doi.org/10.1080/02602930701293231>. Acesso em: 25 fev. 2025.
- SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.
- VILAR, P.; ZABUKOVEC, V. Research data management and research data literacy in Slovenian science. **Journal of Documentation**, v. 75, n. 1, p. 24-43, 2018. Disponível em: <https://doi.org/10.1108/JD-03-2018-0042>. Acesso em: 15 dez. 2024.

NOTAS

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: P. R. S. Costa, J. N. Araújo

Coleta de dados: G. A. DIAS, M. R. F. SOUSA, P. R. S. Costa, J. N. Araújo

Análise de dados: G. A. DIAS, M. R. F. SOUSA, P. R. S. Costa, J. N. Araújo

Discussão dos resultados: G. A. DIAS, M. R. F. SOUSA, P. R. S. Costa, J. N. Araújo

Revisão e aprovação: P. R. S. Costa, J. N. Araújo, A. R. C. SIMÕES

PREPRINTS

(X) O manuscrito não é um *preprint*.

USO DE INTELIGÊNCIA ARTIFICIAL

Não se aplica

FINANCIAMENTO

Pesquisa realizada com auxílio do CNPq, através do processo 311563/2018-0.



APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

A pesquisa, pelo fato de envolver a aplicação de questionários a seres humanos, foi submetida em 14/10/2024 e aprovada em 11/12/2024 pelo Comitê de Ética em Pesquisa da Universidade Federal da Paraíba - UFPB com (Número do parecer: 7.286.870) – CAE 83912024.3.0000.5188

CONFLITO DE INTERESSES

(X) As pessoas autoras declaram não haver interesses conflitantes.

DISPONIBILIDADE DE DADOS DE PESQUISA E OUTROS MATERIAIS²

(X) Os dados já estão disponíveis em repositórios de dados confiáveis. Fornecer os títulos dos conjuntos de dados e as URLs correspondentes: Research Data Management: An investigation from the perspective of Biodiversity, in the context of Open Science.

DOI: <https://doi.org/10.48472/DATAPB/YDW3IM>

ANUÊNCIA DE AVALIAÇÃO ABERTA

(X) Deseja interagir diretamente com o avaliador caso este também concorde, durante o processo de avaliação do manuscrito?

LICENÇA DE USO

As autorias cedem à *Revista Encontros Bibli* os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença [Creative Commons Attribution](#) (CC BY) 4.0 International. Essa licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. As autorias têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em *site* pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. As ideias expressadas neste artigo são de responsabilidade das pessoas autoras, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Patrícia Neubert, Genilson Geraldo, Camila de Azevedo Gibbon, Amanda Santos Witt, Camila de Cássia Brito, Daniela Capri

HISTÓRICO

Recebido em: 22-07-2025

Aprovado em: 08-12-2025

Publicado em: 27-02-2026

Copyright (c) 2026 Paulo Roberto Santos Costa, Guilherme Ataíde Dias, Janaína Nascimento de Araújo, Marckson Roberto Ferreira de Sousa, Anderson Rafael Castro Simões. Este trabalho está licenciado sob uma licença Creative Commons Atribuição 4.0 Internacional. Autores mantêm os direitos autorais e concedem à revista o direito de primeira publicação, com o trabalho licenciado sob a [Licença Creative Commons Attribution \(CC BY 4.0\)](#), que permite o compartilhamento do trabalho com reconhecimento da autoria. Os artigos são de acesso aberto e uso gratuito.

