26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
ISSN 2177-3688

GT 8 – Dados, Informação e Tecnologia

PESQUISA-AÇÃO NA GESTÃO E DISPONIBILIZAÇÃO DE DADOS EM BIODIVERSIDADE

ACTION RESEARCH IN BIODIVERSITY DATA MANAGEMENT AND SHARING

Guilherme Ataíde Dias – Universidade Federal da Paraíba (UFPB)
Paulo Roberto Santos Costa – Universidade Federal da Paraíba (UFPB)
Alba Lígia de Almeida Silva – Universidade Federal da Paraíba (UFPB)

Modalidade: Trabalho Completo

Resumo: Os dados de pesquisa são insumos essenciais para a geração de novos conhecimentos e devem ser gerados de forma que possam ser encontrados, acessados, interoperados e reutilizados. Contudo, a vasta maioria dos dados produzidos pela pesquisa científica permanece armazenada em dispositivos de laboratórios ou dos próprios pesquisadores, sem passar por processos adequados de curadoria e disponibilização em repositórios, configurando o fenômeno denominado cauda longa da pesquisa científica. Nesse contexto, o objetivo desta investigação é desenvolver e implementar um processo participativo de publicação de dados de pesquisa em repositório institucional junto a uma comunidade de pesquisadores em biodiversidade, por meio de ações de letramento em gestão de dados que respondam às demandas identificadas na própria comunidade. A pesquisa configura-se como qualitativa, empregando a pesquisa-ação como método, por excelência adequado a problemas que envolvem demandas comuns a uma coletividade. O coletivo de pesquisa foi composto por aproximadamente sessenta pesquisadores vinculados a instituições de ensino superior do estado da Paraíba, atuantes na área da biodiversidade. As ações de letramento realizadas compreenderam oficinas para elaboração de planos de gestão de dados, reuniões para discussão de padrões de metadados, encontros para familiarização com o repositório institucional e capacitações para o tratamento e visualização de dados. Como resultado, foram depositados quarenta e um conjuntos de dados no repositório institucional, evidenciando que o engajamento colaborativo da comunidade foi determinante para o êxito do processo. Conclui-se que a abordagem participativa inerente à pesquisa-ação se mostrou pertinente para lidar com as demandas da comunidade.

Palavras-chave: pesquisa-ação; biodiversidade; dados de pesquisa; gestão de dados; repositórios de dados.

Abstract: Research data are essential inputs for the generation of new knowledge and must be produced in a way that allows them to be found, accessed, interoperated, and reused. However, the vast majority of data generated by scientific research remains stored on laboratory devices or on researchers' own equipment, without undergoing adequate curation processes or being made available in repositories — a phenomenon known as the long tail of scientific research. In this context, the objective of this investigation is to develop and implement a participatory process for publishing research data in an institutional repository alongside a community of biodiversity researchers, through data management literacy actions that address the demands identified within the community itself. The research is qualitative in nature, employing action research as its method, which is par excellence suited to problems involving demands shared by a collective. The research group comprised approximately sixty researchers affiliated with higher education institutions in the state of Paraíba, working in the field of biodiversity. The literacy actions carried out included workshops for the development of data management plans, meetings to discuss metadata standards, sessions for familiarization with the institutional repository, and training in data processing and visualization. As a result, forty-one datasets were deposited in the institutional repository, demonstrating that the collaborative engagement of the community was decisive for the success of the process. It is concluded that the participatory approach inherent to action research proved pertinent for addressing the community's demands.
Keywords: action research; biodiversity; research data; data management; data repositories.

1 INTRODUÇÃO

Os dados de pesquisa são insumos essenciais para a geração de novos conhecimentos. Esses dados resultam dos processos de investigação científica e devem ser gerados de maneira que possam ser encontrados, acessados, interoperados e reutilizados por humanos e agentes artificiais (Borgman, 2015; Wilkinson *et al.*, 2016). Embora essa situação em que os dados possam ser plenamente aproveitados seja a ideal, nem sempre ela se concretiza de forma efetiva. A vasta maioria dos dados de pesquisa não está disponibilizada em repositórios nem foi submetida a processos rigorosos de curadoria; esses dados encontram-se frequentemente armazenados em dispositivos dos laboratórios ou dos próprios pesquisadores. A situação em que conjuntos de dados, não necessariamente volumosos, são gerados por muitos projetos de pesquisa, não publicados e retidos em laboratórios ou pelos pesquisadores, é conhecida como a cauda longa da pesquisa científica. O fenômeno oposto à cauda longa é caracterizado pela *Big Science*, na qual poucos grandes projetos geram dados organizados e centralizados (Borgman, 2015; Sayão; Sales, 2019).

A pesquisa apresentada neste texto está relacionada ao processo de depósito, no repositório DATAPB, dos dados de pesquisa associados às investigações de pesquisadores na

área da biodiversidade. O DATAPB é gerido pelo Programa de Pós-Graduação em Ecologia e Monitoramento Ambiental (PPGEMA) da Universidade Federal da Paraíba (UFPB), Campus IV, e mantido por uma rede multidisciplinar de colaboradores voluntários (Universidade Federal da Paraíba, 2024). Este repositório foi desenvolvido por meio do produto de software *Harvard Dataverse*, constituindo uma aplicação *Web* de código aberto voltada para o compartilhamento, a preservação, a citação, a exploração e a análise de dados de pesquisa, facilitando a disponibilização de dados para outras pessoas (The Dataverse Project, 2026).

As pessoas envolvidas nesta investigação são pesquisadores associados ao projeto "Biodiversidade da Paraíba: Status, Ameaças e Oportunidades", apoiado pela Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ-PB), por meio de edital vinculado ao Programa de Apoio a Núcleos de Excelência (PRONEX). A percepção da necessidade de realizar ações de gestão de dados que contribuíssem para o seu depósito em um repositório emergiu no início da execução do projeto mencionado. Dessa forma, foram realizadas diversas ações com o intuito de tornar efetivamente viável o compartilhamento dos conjuntos de dados produzidos pelos pesquisadores envolvidos na iniciativa. Os conjuntos de dados trabalhados são originários das atividades de pesquisa desenvolvidas por cientistas da área da biodiversidade que atuam nos diversos ecossistemas do estado da Paraíba, compreendendo dados provenientes dos herbários e de coleções zoológicas da UFPB, da Universidade Federal de Campina Grande (UFCG) e da Universidade Estadual da Paraíba (UEPB).

A partir de uma avaliação inicial dos conjuntos de dados disponíveis, verificou-se a necessidade de conduzir algumas ações inerentes aos Ciclos de Vida de Dados (CVD) e essenciais para que os dados estejam efetivamente disponíveis para publicação (Dias *et al.*, 2020), possibilitando assim seu uso, reúso e compartilhamento, a saber: elaborar Planos de Gestão de Dados (PGD); verificar a consistência dos dados; descrever os elementos de metadados em padrão adequado ao domínio da biodiversidade; determinar a licença de publicação de dados para os respectivos casos concretos; submeter os conjuntos de dados ao repositório DATAPB. A implementação dessas ações demonstrou ser um desafio para sua execução, visto que se verificou a necessidade de letramento dos pesquisadores em biodiversidade no contexto da gestão de dados de pesquisa.

A partir destas considerações preliminares, indaga-se: Como realizar a publicação de dados de pesquisa em um repositório institucional junto a uma comunidade de pesquisadores em biodiversidade que apresenta demandas de letramento em gestão de dados?

Como objetivo principal da investigação busca-se desenvolver e implementar um processo participativo de publicação de dados de pesquisa em repositório institucional junto a uma comunidade de pesquisadores em biodiversidade, por meio de ações de letramento em gestão de dados que respondam às demandas identificadas na própria comunidade.

Esta pesquisa impacta a comunidade internacional de pesquisadores na área de biodiversidade por dar visibilidade a conjuntos de dados que anteriormente estavam menos disponíveis, possibilitando desta forma a geração de novos conhecimentos e divulgando os trabalhos dos pesquisadores das instituições envolvidas.

2 GESTÃO DE DADOS DE PESQUISA: FUNDAMENTOS, FERRAMENTAS E ENGAJAMENTO COMUNITÁRIO

Para que a publicação de dados de pesquisa se concretize efetivamente, faz-se necessário apresentar alguns fundamentos, ferramentas e práticas sociais importantes para um entendimento deste processo. As subseções subsequentes trazem informações que contribuem para uma melhor compreensão do que está sendo discutido.

2.1 A GESTÃO DE DADOS DE PESQUISA E A CIÊNCIA ABERTA

A Ciência Aberta configura-se como um movimento de alcance global que preconiza uma transformação paradigmática nos modos de produção, compartilhamento e comunicação do conhecimento científico. Para além do acesso livre a publicações, esse movimento enfatiza a disponibilização dos dados brutos que servem como base para os resultados da pesquisa, visando aumentar a transparência, a reprodutibilidade e o reuso do patrimônio científico (Silva; Marques, 2023). No contexto da biodiversidade, no qual o volume de dados gerados por sensores, expedições de campo e coleções biológicas cresce significativamente, a Gestão de Dados de Pesquisa (GDP) proporciona meios para evitar a perda de dados científicos (Gadelha *et al.*, 2020).

A GDP pode ser definida como um conjunto de atividades técnicas e gerenciais que acompanham os dados ao longo de todo o seu ciclo de vida - desde o planejamento inicial e a coleta até a preservação a longo prazo e o descarte seguro (Bonetti, 2023). Conforme destacam Sayão e Sales (2012), a ciência atravessa um momento em que a investigação é impulsionada pelo uso intensivo de dados processados por software e armazenados em infraestruturas digitais. Para que essa gestão seja efetiva, os dados devem seguir os princípios FAIR (*Findable*, *Accessible*, *Interoperable* e *Reusable*), garantindo que sejam facilmente

encontrados, acessíveis sob condições claras, interoperáveis entre sistemas e prontos para o reúso pela comunidade das formas que se fizerem necessárias (Wilkinson *et al.*, 2016).

2.2 O PLANO DE GESTÃO DE DADOS (PGD)

Um dos instrumentos operacionais mais importantes da GDP é o Plano de Gestão de Dados (PGD). Trata-se de um documento que descreve formalmente como os dados serão tratados durante e após o projeto de pesquisa (Silva, 2019; Bonetti, 2023). A elaboração de um PGD permite ao pesquisador antecipar decisões críticas sobre formatos de arquivos, padrões de metadados e políticas de compartilhamento, servindo como um norte metodológico para a gestão de dados. Michener, (2015) propõem diretrizes essenciais para um bom PGD, que incluem desde a identificação das responsabilidades pela gestão até a definição de estratégias sólidas de *backup* e preservação.

Atualmente, agências de fomento nacionais e internacionais têm exigido o PGD como requisito para a aprovação de projetos (Veiga *et al.*, 2022). Ferramentas como o *DMPTool* [1] e *DMPOnline*[2] têm auxiliado pesquisadores a preencher esses planos de forma padronizada, oferecendo modelos que facilitam a conformidade com as exigências dos financiadores. Para além de uma obrigação burocrática, o PGD é uma prática científica que contribui para que o investimento em pesquisa retorne para a sociedade dados documentados e inteligíveis.

2.3 PADRONIZAÇÃO SEMÂNTICA COM O *DARWIN CORE* (DWC)

Para que os dados de biodiversidade possam "conversar" entre si e ser integrados em plataformas globais, é indispensável o uso de padrões de metadados. O *Darwin Core* (DwC) é o padrão internacional de grande aceitação para o compartilhamento de dados sobre ocorrências biológicas (*Darwin Core Maintenance Group*, 2026). Mantido pelo *Biodiversity Information Standards* (TDWG), o DwC consiste em um glossário de termos que define entidades como *Occurrence* (ocorrência), *Event* (evento de coleta), *Location* (localização) e *Taxon* (classificação taxonômica) (Gadelha *et al.*, 2020).

A aplicação do DwC permite transformar bases de dados heterogêneas, muitas vezes criadas em laboratórios isolados, em conjuntos de dados interoperáveis. O processo de curadoria exige um mapeamento criterioso das variáveis originais para os termos DwC,

[1] https://dmptool.org/.
[2] https://dmponline.dcc.ac.uk/.

garantindo que o conjunto de dados seja legível por máquinas e integrável por agregadores como o GBIF[3] (*Global Biodiversity Information Facility*) e o SiBBr [4](Sistema de Informação sobre a Biodiversidade Brasileira). Esse alinhamento é o que viabiliza, por exemplo, análises de modelagem de nicho ecológico em escala global para prever impactos de mudanças climáticas.

2.4 REPOSITÓRIOS E A INFRAESTRUTURA DO *HARVARD DATAVERSE*

Após serem tratados e padronizados, os dados precisam de um local seguro para serem depositados. Os repositórios de dados surgem como ambientes digitais equipados com infraestrutura para dar suporte à gestão e disponibilização dessa produção científica. O *Harvard Dataverse*[5] destaca-se como uma plataforma de código aberto amplamente utilizada por instituições brasileiras, como a Unifesp[6], o IBICT[7] e a UFPB (por meio do repositório DATAPB [8] associado ao projeto LMI/IDEAL[9]).

Diferentemente de serviços de armazenamento em nuvem convencionais, o *Dataverse* suporta esquemas de metadados disciplinares, versionamento automático de arquivos e a atribuição de identificadores persistentes, como o DOI (*Digital Object Identifier)*. Além disso, a plataforma permite organizar os dados em comunidades (*dataverses*) que refletem a estrutura da instituição, facilitando a comunicação científica entre diferentes departamentos e centros de pesquisa (Veiga *et al.*, 2022). O uso de repositórios com foco específico em dados é uma das boas práticas que garantem que os dados permaneçam acessíveis e citáveis, fortalecendo a visibilidade do pesquisador e da instituição, contribuindo assim para a ciência aberta.

2.5 QUALIDADE DOS DADOS E FERRAMENTAS AUXILIARES DE CURADORIA

Um dos grandes desafios da GDP é lidar com a "sujeira" informacional. Dados coletados em campo ou extraídos de acervos históricos costumam apresentar inconsistências como erros tipográficos, coordenadas geográficas invertidas ou nomenclaturas taxonômicas

[3] https://www.gbif.org/.
[4] https://www.sibbr.gov.br/.
[5] https://library.harvard.edu/services-tools/harvard-dataverse.
[6] https://domusdados.unifesp.br/.
[7] https://repositoriopesquisas.ibict.br/.
[8] https://dataverse.ideal.ufpb.br/.
[9] https://pt.ird.fr/lmi-ideal-inteligencia-artificial-analise-de-dados-e-observacao-da-terra-aplicadas-ao-laboratorio.

desatualizadas. Para que os dados tenham adequação ao uso, eles devem passar por processos rigorosos de certificação e limpeza (Gadelha *et al.*, 2020).

Nesse cenário, algumas ferramentas auxiliares destacam-se como importantes recursos. O *OpenRefine*[10] é uma ferramenta para a limpeza e transformação de dados volumosos, permitindo a normalização de textos e a reconciliação com bases externas de forma eficiente. Outra ferramenta, como o *Data Cleaner*[11], auxilia na identificação de padrões e valores ausentes, garantindo que o conjunto de dados final tenha a qualidade necessária para análises científicas robustas. A curadoria digital atua, portanto, como um processo de refinamento que agrega valor aos dados brutos, transformando-os em ativos digitais confiáveis.

2.6 VISUALIZAÇÃO DA INFORMAÇÃO E *DASHBOARDS* DINÂMICOS

A Visualização da Informação (VI) busca simplificar a complexidade de grandes conjuntos de dados por meio de representações gráficas interativas, facilitando a identificação de padrões, tendências e *outliers* que passariam despercebidos em tabelas estáticas. A VI utiliza a percepção visual humana para aumentar a capacidade cognitiva no tratamento de informações complexas (Albuquerque; Bittencourt; Pinto Filho; Santa Rita, 2024).

O conceito de *Storytelling* com dados, popularizado por Knaflic (2019), propõe que a visualização não deve apenas mostrar números, mas contar uma história clara que direcione a atenção do público para os *insights* mais relevantes. No campo da biodiversidade, a criação de *dashboards* dinâmicos (utilizando ferramentas como o *Power BI*[12] ou *Looker Studio*[13]) permite que curadores explorem suas coleções visualmente, filtrando espécies por localização geográfica ou visualizando a distribuição geográfica em mapas interativos. Esses painéis atuam como *interfaces* de transparência, permitindo que gestores e a sociedade compreendam rapidamente o estado da biodiversidade local. Ressalta-se que a Visualização da Informação, quando utilizada durante o processo de curadoria de dados, permite que se observe mais claramente possíveis não conformidades relacionadas aos dados de pesquisa, o que permite sua correção.

[10] https://openrefine.org/.
[11] https://datacleaner.github.io/.
[12] https://www.microsoft.com/pt-br/power-platform/products/power-bi.
[13] https://sites.google.com/gsaladeaula.com.br/lookerstudio/in%C3%ADcio.

2.7 O ENGAJAMENTO COMUNITÁRIO EM REPOSITÓRIOS DE DADOS

O desenvolvimento de um repositório de dados de pesquisa constitui um empreendimento multidisciplinar que envolve a disponibilidade de insumos estritamente técnicos, tais como a infraestrutura subjacente de *hardware*, de *software* e os subsistemas de transmissão de dados. Estes insumos são muito importantes para a efetivação bem-sucedida do repositório. Assim como a existência de um alinhamento entre demandas imediatas da comunidade de usuários e os objetivos institucionais de longo prazo, contudo, o fator que provavelmente representa o maior condicionante para o sucesso efetivo do repositório seja o envolvimento de uma comunidade de usuários que faça uso efetivo deste recurso como ponto de convergência para o depósito dos dados fruto de suas atividades de pesquisa. Nesse sentido Heery e Anderson (2005, p. 15) trazem que:

> Para garantir o êxito, é fundamental que os repositórios atendam às necessidades de curto prazo dos usuários, ao mesmo tempo em que sustentem os objetivos estratégicos de longo prazo de financiadores e instituições. Faz-se necessário explorar os requisitos dos usuários e priorizá-los no desenvolvimento dos repositórios, bem como elaborar cenários que investiguem de que modo usuários reais se beneficiariam da implantação de repositórios. O processo de desenvolvimento dos repositórios deve engajar efetivamente a comunidade de usuários, por meio de métodos como estudos de usabilidade e processos participativos de desenvolvimento, com a construção e o refinamento contínuo de casos de uso.

Por conseguinte, nesse contexto entende-se que o efetivo envolvimento da comunidade de pesquisadores não está restrito à etapa de implantação do repositório de dados, mas constitui-se de um processo contínuo que é inerente a todas as etapas de desenvolvimento e consolidação deste recurso.

3 PROCEDIMENTOS METODOLÓGICOS

A escolha do método é um fator precípuo para o sucesso de uma investigação científica; nesse sentido, optou-se por empregar a pesquisa-ação pelas características do problema que envolve demandas comuns a toda uma comunidade. Assim, Thiollent (1996, p. 14) expõe que:

> [...] a pesquisa-ação é um tipo de pesquisa social com base empírica que é concebida e realizada em estreita associação com uma ação ou com a resolução de um problema coletivo e no qual os pesquisadores e os participantes representativos da situação ou do problema estão envolvidos de modo cooperativo ou participativo.

No contexto da investigação realizada, a pesquisa-ação enquadra-se como uma opção metodológica viável, visto que se aplica de maneira adequada para atingir o objetivo proposto.

Nesse sentido, Thiollent (1996) apresenta seis aspectos da pesquisa-ação que se entende estarem em consonância com a pesquisa conduzida, a saber: 1) “[...] há uma ampla e explícita interação entre pesquisadores e pessoas implicadas na situação investigada” (Thiollent, 1996, p. 16). Os pesquisadores, nesse caso, estão envolvidos explicitamente com as pessoas implicadas na situação investigada e pertencem, quase que na sua totalidade, à mesma instituição de pesquisa e a um mesmo projeto de pesquisa. 2) “[...] desta interação resulta a ordem de prioridade dos problemas a serem pesquisados e das soluções a serem encaminhadas sob forma de ação concreta” (Thiollent, 1996, p. 16). A partir da interação entre os envolvidos, emergiu que a prioridade imediata seria abordar o problema de tratar os dados para o seu depósito no repositório. Esta foi uma situação que direcionou os demais esforços da comunidade no sentido do seu atingimento; 3) “[...] o objeto de investigação não é constituído pelas pessoas e sim pela situação social e pelos problemas de diferentes naturezas encontrados nesta situação” (Thiollent, 1996, p. 16). O objeto em questão configura-se pela realidade prática posta na ambiência estudada e refere-se ao desafio e à necessidade de letramento de uma comunidade de pesquisadores em biodiversidade no contexto da gestão de dados de pesquisa; 4) “[...] o objetivo da pesquisa-ação consiste em resolver ou, pelo menos, em esclarecer os problemas da situação observada” (Thiollent, 1996, p. 16). O objetivo apresentado na pesquisa está alinhado com o que a pesquisa-ação se propõe; 5) “[...] há, durante o processo, um acompanhamento das decisões, das ações e de toda a atividade intencional dos atores da situação” (Thiollent, 1996, p. 16). Ao longo das ações conduzidas, as tomadas de decisões foram acompanhadas pelo pesquisador-coordenador e pelos pesquisadores-atores junto ao grupo; 6) “[...] a pesquisa não se limita a uma forma de ação (risco de ativismo): pretende-se aumentar o conhecimento dos pesquisadores e o conhecimento ou o “nível de consciência” das pessoas e grupos considerados” (Thiollent, 1996, p. 16). De modo a alcançar os objetivos, foram realizadas diversas ações com o intuito de incrementar os conhecimentos do grupo com relação ao processo de publicação de dados de pesquisa.

A pesquisa configura-se como qualitativa, visto que a pesquisa-ação, por excelência, configura-se como tal. No que se refere aos objetivos, a pesquisa é exploratória e descritiva (Richardson, 2017), respectivamente em momentos específicos da investigação.

Pelas características específicas da pesquisa-ação, optou-se por não especificar os conceitos estatísticos típicos de uma pesquisa quantitativa de população e amostra. No caso

concreto desta pesquisa, trabalhou-se com a ideia de coletivo de pesquisa composto pelos integrantes do processo, sejam estes o pesquisador-coordenador, os pesquisadores-atores ou os demais pesquisadores envolvidos no processo investigativo. O critério de seleção dos pesquisadores envolvidos na pesquisa deu-se em virtude do grau de envolvimento com o problema. Este coletivo de pesquisa reúne cerca de sessenta pesquisadores que participaram das ações de formação de acordo com suas necessidades formativas e disponibilidade de tempo. A condução da pesquisa realizou-se ao longo do período compreendido entre julho de 2023 e dezembro de 2025.

Dentre as ações de letramento em gestão de dados realizadas com o fito de desenvolver um processo participativo de publicação de dados de pesquisa, destacam-se as seguintes: a realização de duas oficinas, com os pesquisadores, para a elaboração de PGDs; a realização de três reuniões, com os pesquisadores, para a discussão do padrão de metadados *Darwin Core*, voltado para a área da biodiversidade (Figura 1), e para a escolha dos elementos de metadados necessários à descrição de cada um dos conjuntos de dados; a realização de oficinas individuais para a familiarização com o repositório de dados DATAPB e com o produto de *software* Dataverse; e a realização de encontros individuais para a familiarização com ferramentas voltadas para o tratamento de dados e com produtos de *software* voltados para a elaboração de *dashboards*.

Figura 1 - Curso de *Darwin Core*



Fonte: elaboração própria (2023).

Ao longo do processo investigativo, o pesquisador-coordenador e os pesquisadores-atores, com vistas a uma melhor imersão no ambiente em que a pesquisa se desenvolvia, efetuaram diversas visitas e reuniões nos locais em que se encontram acondicionadas as

coleções de espécimes biológicos sob a responsabilidade dos pesquisadores do grupo envolvido na investigação (Figura 2).

Figura 2 - Coleção de peixes



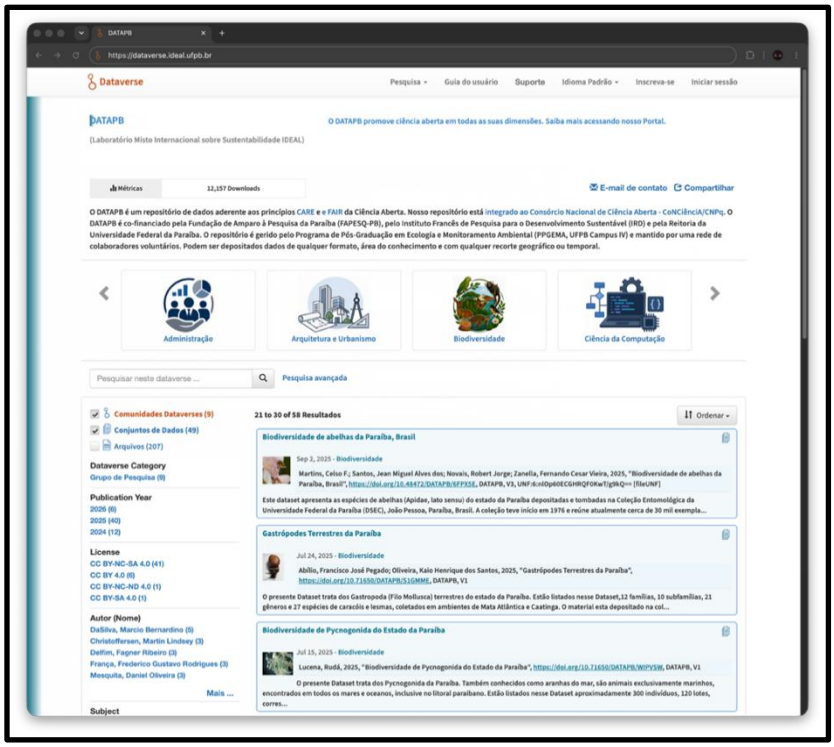
Fonte: elaboração própria (2023).

Esta ação possibilitou o entendimento das demandas do grupo facilitando a construção de soluções que facilitaram a publicação dos dados.

4 RESULTADOS E DISCUSSÃO

Como um dos resultados do processo de investigação contabiliza-se um total de quarenta e um (41) conjuntos de dados depositados no DATAPB (Figura 3). Tal resultado indica que a atuação colaborativa entre os atores envolvidos no empreendimento desempenhou um papel fundamental nos processos de publicação de dados de pesquisa no repositório. Nesse sentido, coaduna-se com o pensamento de Heery e Anderson (2005) ao indicarem que repositórios bem-sucedidos são aqueles estruturados no âmbito de comunidades de prática, ou seja, alicerçados a partir de coletivos de pesquisadores, e não de meros arranjos administrativos.

Figura 3 - Coleções de dados disponibilizados no DATAPB



Fonte: adaptada de Universidade Federal da Paraíba (2024a).

As ações de letramento em gestão de dados realizadas com a comunidade de pesquisadores foram importantes para equalizar o nível de conhecimentos de todos sobre o assunto. Verificou-se uma assimetria entre as pessoas com relação ao nível de conhecimento sobre o tema, mas nenhum dos envolvidos era desprovido de um mínimo de conhecimento sobre o assunto. A princípio, alguns pesquisadores demonstraram alguma resistência à ideia de compartilhar o acesso aos seus dados com o pesquisador-coordenador e os pesquisadores-atores durante as ações de letramento conduzidas. Esta resistência foi superada após os contatos iniciais e a constatação por parte dos pesquisadores de que toda a comunidade envolvida nas ações tinha realmente como intuito o letramento em dados e a disponibilização destes no DATAPB.

Várias dúvidas surgiram durante a pesquisa. Dentre os questionamentos mais frequentes manifestados pelos pesquisadores, destacam-se as dúvidas relacionadas com o esquema de direito autoral a ser adotado quando da publicação dos dados no repositório. Outra questão levantada pela comunidade diz respeito a como deixar os dados encontráveis no repositório, mas não acessíveis pelo público. Nesse caso foi realizada uma explicação sobre a possibilidade de embargar os dados. O embargo de dados configura-se como um período escolhido pelo pesquisador em que é possível encontrar os dados no repositório, acessar os

seus elementos de metadados e obter informações gerais sobre o conjunto de dados e acessar o seu PGD, mas os dados propriamente ditos não estão acessíveis até o momento em que os seus detentores optem por permitir o seu acesso. Um dos motivos para o pesquisador se preocupar em embargar os dados está relacionado com a garantia do ineditismo para futuras publicações, bem como a prevenção de possíveis violações dos direitos autorais. Estes motivos para justificar um embargo encontram abrigo na pesquisa de Medeiros e Sánchez-Tarragó (2021).

Verificou-se que um dos fatores que motivou positivamente a comunidade a publicar os seus dados no repositório foi a ciência de que para cada submissão de um conjunto de dados seria atribuído a ele um identificador persistente, no caso um *Digital Object Identifier* (DOI), quando da submissão desses ao repositório. A atribuição de DOI foi associada de pronto à possibilidade de os dados serem referenciados de forma inequívoca por terceiros.

A pesquisa contribuiu também para conferir visibilidade a dados situados na cauda longa da pesquisa científica. No caso concreto da investigação realizada, foram disponibilizados dados de espécies da biodiversidade. A título de exemplo, os dados da coleção ictiológica da UFPB tiveram oito mil e trezentos (8.300) *downloads* em um período de aproximadamente um ano - dados obtidos diretamente no DATAPB -, fato considerado digno de nota, dada a especificidade do assunto.

5 CONSIDERAÇÕES FINAIS

Verifica-se que o objetivo da pesquisa foi alcançado com êxito, tendo sido possível desenvolver e implementar um processo participativo de publicação de dados de pesquisa no âmbito do DATAPB junto a uma comunidade de pesquisadores em biodiversidade por meio de ações de letramento em gestão de dados que atenderam às demandas identificadas na própria comunidade. Confirma-se que a abordagem participativa da pesquisa-ação, consoante com Thiollent (1996), foi pertinente para lidar com as demandas de uma comunidade de pesquisadores no contexto da biodiversidade, assim como se destaca que o engajamento efetivo da comunidade de pesquisadores foi importante para o sucesso do depósito de dados no repositório, transcendendo os aspectos meramente técnicos conforme indicado por Heery e Anderson (2005).

Entende-se que a principal contribuição do estudo para a Ciência da Informação reside na possibilidade de sistematização das ações empreendidas sob a forma de um modelo

processual de publicação participativa de dados de pesquisa, o qual articula as fases de imersão e diagnóstico, de pactuação de prioridades, de letramento em gestão de dados, de depósito e de acompanhamento do uso dos conjuntos de dados. Destaca-se que tal modelo não se propõe a ser reproduzido de forma mecânica, mas a servir como referencial passível de adaptação por iniciativas que atuem junto a comunidades de pesquisadores às quais se imponha a demanda pela publicação de dados de pesquisa, visto que os resultados da pesquisa-ação estão intrinsecamente vinculados ao contexto em que foram produzidos (Thiollent, 1996).

Apontam-se como limitações do estudo a delimitação do coletivo de pesquisa a pesquisadores de uma única área do conhecimento e a sua vinculação a instituições de pesquisa localizadas apenas no estado da Paraíba. Tais limitações indicam, por conseguinte, a pertinência de que futuras investigações ampliem o escopo da pesquisa para comunidades de pesquisadores de outras áreas do conhecimento e de outras instituições, de modo a permitir a aplicação e o refinamento do modelo processual de publicação participativa de dados de pesquisa em diferentes casos concretos.

REFERÊNCIAS

ALBUQUERQUE, Júlio Augusto Enders de; BITTENCOURT, Ibsen Mateus; PINTO FILHO, Jovino; SANTA RITA, Luciana Peixoto. Explorando a visualização da informação: conceitos, técnicas e aplicações em um mundo de dados complexos**. Informação@Profissões**, Londrina, v. 13, n. 1, p. 1-21, 2024. Disponível em: https://ojs.uel.br/revistas/uel/index.php/infoprof/article/view/49270. Acesso em: 12 jun. 2026.

BONETTI, Letícia G. **Metadados e os princípios FAIR**: um estudo dos repositórios de dados de pesquisa em São Paulo. 2023. 195 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de São Carlos, São Carlos, 2023. Disponível em: https://www.researchgate.net/publication/381205872_Metadados_e_os_principios_FAIR_um_estudo_dos_repositorios_de_dados_de_pesquisa_em_Sao_Paulo. Acesso em: 10 jun. 2026.

BORGMAN, Christine L**. Big data, little data, no data**: scholarship in the networked world. Cambridge: MIT Press, 2015.

DARWIN CORE MAINTENANCE GROUP. **Darwin Core**. Alexandria: Biodiversity Information Standards (TDWG), 2026. Disponível em: https://dwc.tdwg.org. Acesso em: 12 jun. 2026.

DIAS, Guilherme Ataíde et al. Publicando dados de pesquisa: contextualizando as principais etapas e elementos envolvidos no processo**. Ciência da Informação**, Brasília, v. 49, n. 3, p. 1-

41, 2020. DOI: 10.18225/ci.inf.v49i3.5501. Disponível em: https://revista.ibict.br/ciinf/article/view/5501. Acesso em: 11 jun. 2026.

GADELHA, Luiz M. R. et al. A survey of biodiversity informatics: concepts, practices, and challenges**. WIREs Data Mining and Knowledge Discovery**, Hoboken, v. 10, n. 5, e1394, 2020. Disponível em: https://doi.org/10.1002/widm.1394. Acesso em: 11 jun. 2026.

HEERY, Rachel; ANDERSON, Sheila. **Digital repositories review**. Bristol: Joint Information Systems Committee (JISC), 2005. Disponível em: https://purehost.bath.ac.uk/ws/portalfiles/portal/498122/digital-repositories-review-2005.pdf. Acesso em: 11 jun. 2026.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados**: um guia sobre visualização de dados para profissionais de negócios. Tradução de João Tortello. 2. ed. Rio de Janeiro: Alta Books, 2019. 256 p.

MEDEIROS, Joyanne de Souza; SÁNCHEZ-TARRAGÓ, Nancy. Acesso aberto imediato ou embargo dos trabalhos de conclusão de curso? Estudo embasado na teoria da ação racional. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 26, n. esp., p. 1-22, 2021. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/81917. Acesso em: 12 jun. 2026.

MICHENER, William K. Ten simple rules for creating a good data management plan. **PLoS Computational Biology**, San Francisco, v. 11, n. 10, e1004525, 2015. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC4619636/. Acesso em: 12 jun. 2026.

RICHARDSON, Roberto Jarry**. Pesquisa social**: métodos e técnicas. 4. ed. rev. atual. ampl. São Paulo: Atlas, 2017.

SAYÃO, Luis Fernando; SALES, Luana Farias. A ciência invisível: os dados da cauda longa da pesquisa científica. In: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire de (org.). **Dados científicos**: perspectivas e desafios. João Pessoa: Editora UFPB, 2019. p. 1-21.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade**: Estudos, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012. Disponível em: https://periodicos.ufpb.br/index.php/ies/article/view/12224. Acesso em: 11 nov. 2026.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 146 p.

SILVA, Vanessa Batista da; MARQUES, Tathiane Amaral. O uso do Dataverse como repositório de dados: uma análise prática. *In*: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 22., 2023, Florianópolis. **Anais** [...]. Florianópolis: FEBAB, 2023. p. 1-10. Disponível em: https://portal.febab.org.br/snbu2023/article/view/2920. Acesso em: 10 jun. 2026.

THE DATAVERSE PROJECT. **About**. Cambridge: Harvard University, Institute for Quantitative Social Science, 2026. Disponível em: https://dataverse.org/about. Acesso em: 12 jun. 2026.
THIOLLENT, Michel**. Metodologia da pesquisa-ação**. 7. ed. São Paulo: Cortez, 1996.

UNIVERSIDADE FEDERAL DA PARAÍBA. **Dataverse DATAPB**. João Pessoa: UFPB, [2024]. Disponível em: https://dataverse.ideal.ufpb.br/. Acesso em: 11 jun. 2026.

VEIGA, Viviane S. de O.; DIB, Simone Faury; HENNING, Patricia Corrêa; NASCIMENTO, Francisco José T. *et al.* **Plano de gestão de dados de pesquisa (PGD):** guia de elaboração. Rio de Janeiro: Fiocruz; ICICT, 2022. 32 p. Disponível em: https://arca.fiocruz.br/handle/icict/54805. Acesso em: 12 jun. 2026.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, art. 160018, 2016. Disponível em: https://www.nature.com/articles/sdata201618. Acesso em: 11 jun. 2026.

AGRADECIMENTOS

Agradecemos ao CNPq, por meio do processo 311563/2018-0, à Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ-PB) e ao *Institut de Recherche pour le Développement* (IRD) pelo aporte de recursos que contribuíram para a realização desta pesquisa.

NOTA

Para a correção gramatical deste trabalho foi utilizada a ferramenta de inteligência artificial *Claude Opus 5*.