

Elaboração, aplicação e avaliação de um plano de gestão de dados: o caso DataPB

Adriana Alves Rodrigues¹

<https://orcid.org/0000-0003-2378-6934>

Joana Ferreira de Araújo¹

<https://orcid.org/0000-0002-5693-3964>

Pedro Felipy Cunha da Silva¹

<https://orcid.org/0000-0001-9063-4629>

Vivianne de Queiroz Leal¹

<https://orcid.org/0009-0006-2967-5150>

Guilherme Ataíde Dias¹

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

Alzira Karla Araújo da Silva¹

<https://orcid.org/0000-0003-3499-2530>

Wagner Junqueira de Araújo¹

<https://orcid.org/0000-0002-2301-4996>

¹Universidade Federal da Paraíba, João Pessoa, PB, Brasil

Resumo: Na ciência contemporânea, as Instituições de Ensino Superior, Agências de Fomento, Órgãos de Governo e Instituições de Pesquisa tem exigido a elaboração de um Plano de Gestão de Dados, documento que estabelece os procedimentos mandatórios para gerenciamento dos dados da pesquisa, com intuito de auxiliar durante toda investigação acadêmica. Diante disso, o objetivo dessa pesquisa foi analisar o processo de elaboração e aplicabilidade do Plano de Gestão de Dados na área da Biodiversidade no âmbito da Universidade Federal da Paraíba. A investigação é qualitativa, de natureza aplicada e intervencionista. Utilizou-se o método da pesquisa-ação para descrever e avaliar a aplicação do Plano. Os resultados apontam para um processo que envolveu as etapas de: (a) planejamento, ao realizar uma pesquisa prévia sobre os Planos de Gestão de Dados existentes, definir metas, organizar as tarefas a serem realizadas, indicar responsabilidades e prazos de execução; (b) ação, por meio da aplicação-teste do Plano desenvolvido; (c) descrição, momento de sistematizar o processo de

planejamento, elaboração, sugestões de melhorias; e (d) avaliação, a partir da análise do Plano, com vistas a corrigir inconsistências e aprimorar o resultado final. Conclui-se que o registro e partilha do processo de elaboração, aplicação e avaliação de um Plano de Gestão de Dados faz-se salutar para o intercâmbio de ideias e experiências, além de contribuir para que outros pesquisadores e Instituições de Ensino Superior possam se inspirar e também adotar o instrumento para gestão de dados de suas investigações e, ampliar o escopo de possibilidades investigativas e de transparência dos dados das pesquisas, através da publicação.

Palavras-chave: plano de gestão de dados; *e-science*; dados de pesquisa; *big data*

1 Introdução

O progresso contínuo das tecnologias da informação no âmbito científico tem facilitado e ampliado a aquisição de grandes conjuntos de dados, os quais podem ser empregados como componentes essenciais para a condução de uma ampla variedade de pesquisas científicas. Esse fenômeno tem influenciado a adaptação das práticas dos cientistas em relação aos dados de pesquisa, uma vez que esses se tornaram recursos fundamentais e indispensáveis no cenário da ciência contemporânea, essenciais nos processos de geração de novos conhecimentos científicos (Curty; Aventurier, 2017). As pesquisas científicas demandam o uso intensivo e a gestão adequada de vastos conjuntos de dados, bem como a análise criteriosa desses dados, o que inclui o desenvolvimento de métodos científicos apropriados que atendam às exigências metodológicas das investigações (Azam, 2014).

A crescente aquisição de dados de pesquisa e os diversos contextos associados ao seu compartilhamento são aspectos fundamentais que impulsionam o progresso da ciência em várias áreas, de forma abrangente. Esses aspectos não apenas maximizam a visibilidade das pesquisas científicas, mas também contribuem para ampliar os debates acadêmicos, facilitam o reuso dos dados e promovem a transparência, fortalecendo assim o impacto das pesquisas. Adicionalmente, o compartilhamento de dados ajuda a evitar a duplicação de esforços de pesquisa e cria cenários inovadores para futuras investigações, entre outras possibilidades.

Neste contexto delineado, o fenômeno do *big data* reconfigura práticas, ações e a abordagem de gerir grandes volumes de dados e desafia a forma como se vive em sociedade (Mayer-Schonberger; Cukier, 2013). Além disso, demanda assertividade e celeridade nos processos infocomunicacionais, que impulsionam o acesso aos conjuntos de dados de modo aberto. Assim, os dados são produzidos em todos os contextos, em larga escala e velocidade contínua, possibilitando o desenvolvimento dos mais diversos produtos de informação. Seguindo essa mesma lógica de raciocínio, Ferrer-Sapena e Sánchez-Pérez, (2013, p. 150) complementam ao afirmarem que: “[...] agora queremos que estejam acessíveis também os dados, de maneira livre e gratuita”¹.

Na ciência contemporânea, os dados que antes ocupavam um lugar secundário nas investigações científicas, se tornam os protagonistas, galgando novas perspectivas metodológicas e de investigação (Sales; Sayão, 2019). São considerados “[...] não apenas alicerces de pesquisas realizadas, mas como subsídios essenciais para outros pesquisadores, instituições acadêmicas e agências de fomento” (Fernandes; Oliveira, 2018, p. 257). Este cenário é evidenciado pela publicação de normativas específicas para a gestão de dados, conforme estabelecido pela Portaria n. 693, de 20 de agosto de 2009, do Ministério da Ciência, Tecnologia e Inovação (Brasil, 2009). Tal portaria define, entre outros aspectos, no contexto do Programa de Pesquisa em Biodiversidade (PPBio), uma política de dados que inclui protocolos para a criação, gestão, acesso e utilização desses dados.

A ciência orientada por dados (*data-driven science*) representa as características de um novo paradigma, no qual os dados emergem como novos substratos do conhecimento. Eles são coletados de extensos conjuntos de dados com o objetivo de desvelar novas competências, vivências e aprendizados (Himanen *et al.*, 2019). Distinta das metodologias científicas convencionais, a prática da ciência orientada por dados diferencia-se pelo significativo volume de dados processados e pela metodologia mais automatizada de extração, utilização, processamento, análise e disseminação de informações. Nesse contexto, o acesso a dados abertos no meio científico reformula as práticas em todas as disciplinas

do saber, dada a ampla gama de possibilidades relativas à coleta, armazenamento, segurança, recuperação, compartilhamento, uso e os recursos providos pelos conjuntos de dados de pesquisa.

Conforme exposto por Peset, Ferrer-Sapena e Subirats-Coll (2011), os dados abertos constituem um movimento que advoga pela disponibilização de dados, predominantemente não-textuais, em formatos passíveis de reutilização, oriundos de diversas organizações científicas. Esse cenário demanda uma gestão de dados eficiente, que satisfaça os requisitos de armazenamento, compartilhamento, utilização e reutilização dos dados de pesquisa que circulam nas redes digitais. É imperativo considerar que a variedade de dados provenientes de investigações científicas, assim como seus formatos e a maneira como são disponibilizados, variam de acordo com o campo de conhecimento específico.

Observa-se que diferentes aplicativos e plataformas eletrônicas vêm sendo utilizadas para coletar, armazenar e compartilhar dados de pesquisa das mais diferentes fontes. Tal conceito pode ser verificado no trabalho de Djick, Poell e Wall (2018), quando aborda o conceito de Plataformização da sociedade, como um cenário a partir do qual todas as ações humanas passam a ser realizadas por meio de aplicativos. Este cenário torna premente a reflexão sobre um novo tipo de profissional da informação, capaz de gerenciar os recursos provenientes das atividades de pesquisa criados em ambientes digitais, alguém capaz de realizar a curadoria dos dados (Higgins, 2008), especificamente os científicos (Green; MacDonald; Rice, 2009), com suas características próprias (Sales; Sayão, 2015).

Gerir os dados científicos exige, portanto, novas habilidades para os profissionais envolvidos com o processamento desse recurso informacional, tais como o estabelecimento de normas que padronizam sua captura, descrição, tratamento técnico, prazos de guarda nas diferentes fases de uma pesquisa, formatos de armazenamento, regras para descarte, preservação, reutilização dos dados originais e as novas demandas envolvidas no compartilhamento desse tipo de material (Costa; Cunha, 2019).

Desta forma, para um melhor gerenciamento dos dados de pesquisa, a criação do Plano de Gestão de Dados (PGD) é um aspecto fundamental para qualquer organização que deseja aproveitar ao máximo as informações disponíveis. Com o aumento exponencial da quantidade de dados gerados diariamente, tornou-se essencial desenvolver estratégias eficazes para coletar, armazenar, processar, compartilhar e descartar dados. O PGD pode ser entendido como um arsenal de preceitos que estipula como os dados serão gerenciados antes, durante e depois da pesquisa. É um processo complexo que envolve a coleta, armazenamento, processamento, compartilhamento e descarte de dados. É fundamental desenvolver estratégias eficazes para gerenciar os dados de forma segura, eficiente e eficaz. Isso inclui a definição de políticas, procedimentos e protocolos para garantir a qualidade, segurança e privacidade dos dados. A implementação de um PGD é fundamental para garantir que os dados sejam gerenciados de forma eficiente e segura. Isso ajuda a reduzir o risco de perda ou violação de dados, e garante que os dados sejam utilizados para alcançar os objetivos do projeto ou programa.

Além disso, o PGD também é fundamental para garantir a conformidade com as leis e regulamentos aplicáveis. Isso inclui a definição das políticas e procedimentos para gerenciar os dados, bem como a definição das medidas de segurança para proteger os dados.

A partir da discussão empreendida, consideramos a importância de partilhar experiências voltadas para o desenvolvimento de PGD que melhor orientem os processos relacionados ao contexto dos dados abertos, surge esta investigação. Nesta direção, o objetivo do estudo é registrar as etapas de elaboração, aplicação e avaliação de um PGD desenvolvido para a área da Biodiversidade, no âmbito da Universidade Federal da Paraíba (UFPB), com o intuito de realizar a gestão de dados aferindo visibilidade e organicidade sobre os dados de pesquisa desta área de conhecimento.

No decorrer desta pesquisa, adotou-se o método pesquisa-ação, que se adequa à resolução de investigações no domínio da Ciência da Informação. Esse método tem demonstrado a capacidade de gerar resultados práticos,

especialmente em programas de Pós-graduação (Almeida; Perucchi; Freire, 2019), bem como oferecer uma abordagem dinâmica que se adequa aos profissionais em contexto real (Galvão; Ricarte; Daura, 2011). Dessa maneira, o presente estudo detalha o processo de formulação e aplicação de um Plano de Gestão de Dados, elaborado especificamente para o campo da Biodiversidade, no âmbito da UFPB.

2 Plano de gestão de dados

De uma maneira ampla, pode-se compreender o Plano de Gestão de Dados (PGD), *Data Management Plan (DMP)*, em inglês, como um conjunto de políticas mandatórias que determina o tratamento e ordenamento para a gestão de dados de pesquisa, durante e após o término da investigação. Isto é, tais produções dos PGDs necessitam da integração com o Ciclo de Vida dos Dados² (CVD), percurso que vai da descoberta, organização, coleta, até a descrição de como os dados serão preservados (Michener, 2015; Dudziak, 2016). A função primordial do PGD consiste em assistir os pesquisadores na coleta, processamento e descrição dos dados de pesquisa, bem como oferecer diretrizes essenciais para o compartilhamento, utilização e reutilização desses conjuntos de dados. Instituições situadas em nações como Holanda, Reino Unido, Austrália e Estados Unidos, dentre outras, frequentemente exigem que os pesquisadores elaborem os respectivos PGDs para as investigações propostas.

Apesar de ser uma atividade ainda recente no país, algumas Instituições de Ensino Superior (IES) e agências de fomento começaram a requerer o PGD junto aos pesquisadores. Tal medida visa expandir a organicidade dos dados de pesquisa, além de promover sua preservação e curadoria de forma aprimorada.

No Brasil, destacam-se iniciativas, como da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que compreendeu a importância desse documento para a comunidade científica, pesquisadores e sociedade civil. A FAPESP percebe que “[...] a gestão apropriada desses dados facilita a

reprodutibilidade da pesquisa e permite promover novas pesquisas, graças à possibilidade de reuso e compartilhamento” (FAPESP, 2023, local. 1).

Pode-se citar também o PGD-BR vinculado ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), fundamentado nos Princípios FAIR³. O projeto teve início em 2011 agregando algumas⁴ instituições parceiras. Mais a frente e devido à grande aceitação da primeira versão, os sócios fundadores conseguiram o subsídio financeiro da Fundação Alfred P. Sloan e criaram, em 2014, uma versão atualizada da ferramenta. Desta maneira, o PGD-BR atendeu às requisições para que as agências de fomento contemplassem os PGDs, tais como a *National Science Foundation* (NSF) e o *National Institutes of Health* (NIH) (Fundação Alfred Sloan, 2025; IBICT, 2025). Assim, a ferramenta online e gratuita proporciona instruções para a construção de PGD que visam atender às demandas das agências de fomento e IES, bem como disponibilizar o PGD para que outros pesquisadores tenham acessos aos dados científicos disponíveis na plataforma. Dentre os princípios que norteiam o trabalho do PGD-BR estão: aprimoramento contínuo de uso dos recursos disponíveis, requisitos e prioridades de desenvolvimento com foco no usuário, entusiasmo e compromisso com o processo aberto, garantia de qualidade dos códigos e materiais desenvolvidos.

Soma-se, ainda, o FioDMP, PGD desenvolvido pela equipe do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, com apoio do Grupo de Trabalho Ciência Aberta, da Fiocruz - instituição que se destaca por suas contribuições para o movimento de Ciência Aberta no país (Veiga *et al.*, 2019). A solução orienta a construção do PGD com base em uma sequência de perguntas categorizadas em seis seções, quais sejam: informação administrativa (corresponde ao pesquisador e a pesquisa proposta); descrição dos dados coletados e/ou reuso de dados existentes; documentação e qualidade dos dados; armazenamento e *backup*; requisitos legais, éticos e códigos de conduta; e compartilhamento de dados (Fiocruz, 2025; Veiga *et al.*, 2019, Veiga *et al.*, 2022).

No cenário internacional também podem ser identificados intentos. A Universidade de Michigan, por meio do seu Instituto de Pesquisa Social, disponibiliza o *ICPSR: Sample Data Management Plan for Social and Political*

Science Data, uma ferramenta de apoio aos candidatos à bolsas de pesquisa naquela instituição (Inter-university Consortium for Political and Social Research, 2025). O *NEH-ODH: Data Management Plans from Successful Grant Applications*, foi a solução criada pelo Escritório de Humanidades Digitais norte-americano para orientar proponentes de financiamento público nesta área (National Endowment for the Humanities, 2021). O *NIH: Scientific Data Sharing*, que auxilia na criação de uma política de gerenciamento e compartilhamento de dados, voltada para subsidiar a distribuição de conjuntos de diferentes campos do conhecimento nos Estados Unidos (National Institutes of Health, 2023).

Na Europa, a organização *Science Europe* desenvolveu manuais e modelos de planos de gestão de dados científicos para orientar instituições, pesquisadores, revisores e toda a comunidade científica sobre as melhores práticas de ciência aberta (Science Europe, 2021). Essas, entre outras iniciativas, inspiram o desenvolvimento de PGD e fortalecem o movimento de abertura da ciência.

Para a elaboração de um PGD, Michener (2015, tradução nossa)⁵ aponta oito aspectos-chave que devem estar contidos nos planos, quais sejam:

- (1) Determinar os requisitos do patrocinador de pesquisa;
- (2) Identificar os dados a serem coletados;
- (3) Definir como os dados serão organizados;
- (4) Explicar como os dados serão documentados;
- (5) Descrever como a qualidade dos dados será assegurada;
- (6) Apresentar uma estratégia sólida de armazenamento e preservação de dados;
- (7) Definir as políticas de dados do projeto;
- (8) Descrever como os dados serão divulgados;
- (9) Atribuir funções e responsabilidades;
- (10) Preparar um orçamento realista. Em conformidade com o autor, tais aspectos podem garantir a segurança dos dados além de amparar os pesquisadores na gestão de dados de pesquisa.

A efetivação do PGD sugere uma atuação de maneira colaborativa que integre tanto as IES quanto a comunidade acadêmica, considerando que as investigações científicas requerem a admissão de novas metodologias, bem como estratégias de gestão de dados, manejo tecnológico e práticas atualizadas. Neste sentido, Sales e Sayão (2022) recomendam que a elaboração de um PGD contenha uma mescla de funcionalidades com o intuito para a acomodação dos dados de

modo eficaz e que, ao mesmo tempo, ofereça suporte técnico para os pesquisadores. Baseados nos Princípios FAIR⁶, os autores propõem o que denominam de “serviços de gestão de dados de pesquisa” que são definidos como “[...] o conjunto de serviços informacionais, computacionais, científicos e administrativos oferecidos no âmbito da gestão de dados de pesquisa” (Sales; Sayão, 2022, p. 3-4).

Estes serviços são divididos em quatro dimensões: (1) a governança, onde são discutidos os princípios norteadores do projeto de serviços de gestão de dados; (2) as infraestruturas técnicas, nas quais se incluem também as categorias de expertises necessárias; (3) os serviços informacionais, computacionais, científicos e administrativos; e (4) os resultados da efetivação desses serviços manifestados pela “FAIRificação” dos dados (Sales; Sayão, 2022, p. 4).

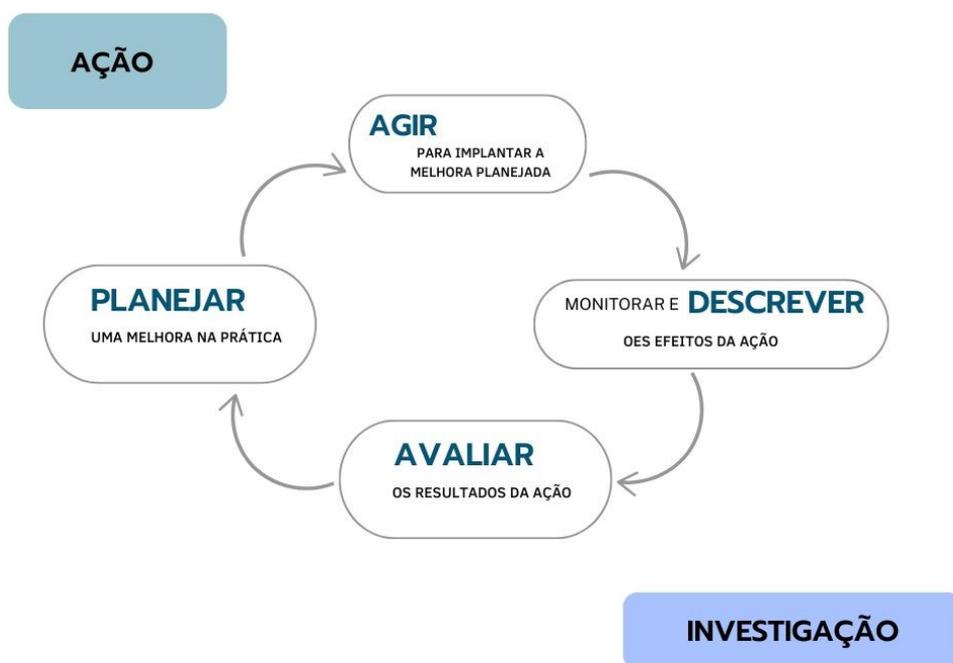
3 Procedimentos metodológicos

A investigação empreendida é qualitativa, de natureza aplicada e intervencionista. Utiliza o método da pesquisa-ação para descrever o processo de desenvolvimento e aplicabilidade de um PGD para a área da Biodiversidade no âmbito da UFPB. Conforme assinala Thiollent (1986, p. 16), este tipo de método é alicerçado com a finalidade de realização dos objetivos práticos inseridos em um ambiente-chave e tem como principal proposta “[...] resolver ou, pelo menos, em esclarecer os problemas da situação observada [...]”, na qual pode ser estabelecida uma configuração com características como: coletiva, participativa e ativa em que as ciências sociais acredita que este método tem como objeto de estudo “[...] analisar suas qualidades, potencialidades, limitações e distorções” (Thiollent, 1986, p. 25). Como define Thiollent (1986, p. 15):

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.

Nesse contexto, além de instituir uma ação na prática, a pesquisa-ação não se restringe apenas a esse modelo aplicável, mas também, em expansão do conhecimento e amplificando os horizontes que conduz a uma ação dialógica entre pensamento, planejamento e ação. Através dos estudos de Kurt Lewin (1890-1947), considerado com um dos pioneiros da pesquisa-ação (Thiollent, 1986; Engel, 2000; Koerich *et al.*, 2009), esse método vem sendo explorado sob diversas perspectivas e intencionalidades, agregando uma vasta composição de abordagens teórico-metodológicas com vistas à práxis investigativa. Para essa pesquisa, aplicam-se as Etapas do Ciclo da Pesquisa-Ação (Figura 1).

Figura 1 - Etapas do Ciclo da Pesquisa-Ação



Fonte: Elaborado pelos autores, adaptado de Tripp (2005).

Tripp (2005, p. 445-446) ressalta que a pesquisa-ação “[...] é um termo genérico para qualquer processo que siga um ciclo no qual se aprimora a prática pela oscilação sistemática entre agir no campo da prática e investigar a respeito dela”. Para o autor, que estabeleceu as quatro fases do Ciclo da Pesquisa-ação,

estas visam o aprimoramento de sua execução e melhora das práticas advindas de sua aplicação, no que resulta em maior nível de aprendizagem durante a processualidade da pesquisa inserida na própria investigação. Tendo em vista a vastidão dos conceitos e características, essas fases pretendem delimitar melhor o objeto de estudo e impulsionar a união do conhecimento com a ação prática, isto é, com a pesquisa empírica ou “pesquisa-ação participativa”, como o autor assim denomina.

Importante ressaltar que as utilizações e desenvolvimento do Ciclo de investigação podem requerer ações distintas em cada etapa como também em diferentes lugares/ambientes. Desta forma e de posse das etapas do Ciclo, descrevem-se o desenvolvimento e a aplicabilidade do PGD em conformidade com as contribuições de Tripp (2005) que ocorreu no período de 23 de agosto a 28 de novembro de 2023.

4 Resultados e discussões: planejar, agir, descrever e avaliar

Nesta seção foram descritas as etapas da pesquisa-ação propostas por Tripp (2005), quais sejam: planejar, agir, descrever e avaliar; conforme discorre-se nas subseções que seguem.

4.1 Planejar

O projeto *DataPB* consiste em um repositório de dados voltado para a área da Biodiversidade e teve seu início com o projeto Laboratório Misto Internacional IDEAL (*Artificial Intelligence, Data Analytics and Earth Observation Applied to Sustainability Lab*)⁷. Envolve “[...] um consórcio de pesquisa socioecológico franco-brasileiro aprovado pelo Instituto Francês de Pesquisa para o Desenvolvimento (IRD) com vigência entre os anos de 2023-2027 (*DataPB*, 2024).

Com o objetivo de reunir tecnologias da informação e ciência de dados para construção de uma rede de pesquisa que possibilite transformações

socioculturais, políticas, econômicas e ambientais, segundo preceitos da sustentabilidade (DataPB, 2024), o projeto *DataPB* visa a criação de um repositório de dados para armazenamento e disseminação de dados produzidos pela área da biodiversidade, podendo expandir-se para outras áreas do conhecimento.

Nesse contexto, surge a oportunidade de colaborar com a construção desse repositório, em especial à gestão dos dados, com a elaboração de um PGD que poderá ser integrado ao *DataPB*, como padrão na submissão de conjunto de dados. Somar a expertise de pesquisadores de diversas áreas – Ciência da Informação, Tecnologia, Administração, Direito, Biodiversidade etc – neste processo, faz-se salutar para o bom desenvolvimento dos recursos e ferramentas, ampliando a rede de troca de saberes. Em acréscimo, reitera-se que esta é a primeira iniciativa da criação de um repositório de dados no âmbito da Universidade Federal da Paraíba, o que destaca o seu caráter desafiador e inovador.

Para dar início à elaboração do PGD, foi necessário, *a priori*, estabelecer um planejamento que orientasse as práticas a serem desenvolvidas, estabelecendo ações, prazos e equipes responsáveis por cada etapa. Após reunião de alinhamento, definiram-se quatro etapas para elaboração do PGD (Quadro 1).

Quadro 1 - Etapas da elaboração do PGD (planejamento)

ETAPA	AÇÃO	RESPONSÁVEL	PRAZO DE EXECUÇÃO
1	Pesquisa e estudos acerca de modelos e orientações para elaboração do PGD	Orientador e orientandos	23/08/2023 a 04/09/2023
2	Elaboração do PGD	Orientandos, sob supervisão do orientador	05/09/2023 a 24/10/2023
3	Teste de aplicação do PGD com pesquisadores da biodiversidade	Orientandos, sob supervisão do orientador	25/10/2023 a 01/11/2023
4	Avaliação do modelo construído conforme percepções da equipe em seus diários de pesquisa	Orientandos e orientador	02/11/2023 a 28/11/2023

Fonte: Dados da pesquisa.

Ressalta-se que as etapas supracitadas estão alinhadas com o que propõe Tripp (2005) ao tratar da pesquisa-ação, pois parte da necessidade de aprimoramento de uma prática (planejar), até as ações empenhadas para implantar as melhorias (agir), acompanhamento da aplicação (descrever) e avaliação das estratégias/ferramentas (avaliar).

4.2 Agir

Com o planejamento definido deu-se início à preparação necessária para a ação. O primeiro passo consistiu na pesquisa e levantamento de exemplos de PGD, assim foram encontrados os modelos e as orientações disponibilizadas pelo *Digital Curation Center* (DCC) e o *DMPTool* utilizados na elaboração do PGD a ser implantado no *DataPB* (Digital Curation Center, 2023). Esta ação iniciou-se no dia 23 de agosto de 2023 e finalizou em quatro de setembro do mesmo ano, momento em que, de posse das orientações e modelos disponíveis, bem como o amparo na literatura pertinente acerca da temática, seguiu-se a elaboração do PGD, no dia cinco de setembro de 2023.

Conforme apontam Sayão e Sales (2022), a elaboração de PGD adequado deve considerar uma junção de funcionalidades e especificações acerca dos dados a serem disponibilizados, de modo que ofereça de maneira eficaz o suporte técnico para pesquisadores que poderão utilizar os dados a posteriori. Entendendo a importância de um recurso como o PGD no contexto da abertura de dados à comunidade geral (Peset; Ferrer-Sapena; Subirats-Coll, 2011) e, ainda, a identificação de pontos imprescindíveis a serem contidos no PGD (Michener, 2015), elaborou-se o PGD teste do *DataPB*.

Para tanto, a equipe reuniu-se semanalmente com vistas a discutir os itens e pensar estratégias que viabilizassem o preenchimento do PGD por pesquisadores de quaisquer áreas, utilizando uma linguagem mais clara e objetiva, inserindo explicações e sugestões quando necessário para a melhor compreensão dos futuros usuários.

Após os ajustes, realizou-se uma reunião presencial para apresentação do PGD construído pela equipe, com o objetivo de identificar possíveis falhas e necessidades de esclarecimentos. Finalizada a etapa de elaboração do PGD (24/10/2023), partiu-se para sua aplicação, em um pré-teste com pesquisadores da área da biodiversidade, que se voluntariaram em preencher o PGD.

Dessa forma, no dia 25 de outubro de 2023, a equipe responsável pela elaboração do PGD reuniu-se com os pesquisadores e orientou, de forma individual, o preenchimento do PGD por cada um deles. Durante esse processo, foram anotadas as dúvidas mais recorrentes, questões a serem mais bem estruturadas, entre outros pontos necessários para o aperfeiçoamento.

Com as anotações da experiência de teste, percebeu-se a falta de clareza de algumas questões postas no PGD, o que levou a uma nova reunião para troca de experiências e ajustes no PGD, finalizado em 28/11/2023. Cada uma das etapas possibilitou um aprendizado aprofundado acerca das funções de um PGD.

A elaboração de modelo que poderá ser implantado em um repositório de dados, auxiliará outros pesquisadores no gerenciamento e disseminação de seus conjuntos de dados. Dessa forma, por ora cabe o registro/descrição detalhada dessas ações, com vistas a inspirar pesquisadores interessados na temática.

4.3 Descrever

Marques e Sayão (2023, p. 4) ressaltaram a necessidade da “[...] criação de um modelo de informação de dados de pesquisa mais realista, que considere a complexidade de uma nova condição de publicação de dados que seja aberta, contextualizada e em rede”. Os autores afirmaram que “[...] existe um papel fundamental da Ciência da Informação e da Biblioteconomia na revolução da descoberta computacional que está subjacente à *e-Science*” (Marques; Sayão, 2023, p. 17).

Nesse contexto, e com a intenção de instrumentalizar os conhecimentos obtidos no curso da disciplina *E-science no âmbito da Ciência da Informação* do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da UFPB, os

discentes foram estimulados a produzir um PGD, e aplicá-lo junto a um grupo de pesquisadores atuantes no *Projeto Biodiversidade da Paraíba: Status, Ameaças e Oportunidades*. O projeto é financiado pela Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ-PB), em seu edital direcionado ao Programa de Apoio a Núcleos de Excelência (PRONEX) (FAPESQ-PB, 2022). De acordo com o *site* do Centro de Ciências Aplicadas e Educação (CCAIE) da UFPB (2023), trata-se de um grupo formado por mais de 50 pesquisadores de 13 instituições brasileiras e estrangeiras, coordenado por um pesquisador, do Departamento de Sistemática e Ecologia (DES) da UFPB, que compõe o Núcleo de Excelência em Biodiversidade da Paraíba, vinculado ao Programa de Pós-Graduação em Ciências Biológicas da UFPB.

O Núcleo funciona em colaboração com o Laboratório Misto Internacional IDEAL. Para executar o planejamento realizado, os estudantes do PPGCI/UFPB construíram uma proposta de PGD com base nas ferramentas desenvolvidas pelo *Digital Curation Center* (DCC) da *University of Edinburgh* e pelo *California Digital Center* da *University of California* (*DMPTool*) (Digital Curation Center, 2023).

A ação teve como finalidade favorecer o compartilhamento dos conjuntos de dados já existentes e subsidiar a padronização dos novos registros a serem incluídos no repositório *DataPB*, foi utilizado para a criação de um protótipo de PGD a ferramenta de criação de formulários *Google Forms*.

A construção do PGD foi iniciada com a avaliação dos questionários já existentes nas ferramentas do *DCC* e *DMPTool*, e a posterior inclusão de novas questões consideradas necessárias para o contexto da atividade. Os ajustes foram feitos através de reuniões remotas no aplicativo *Google Meet*, e a documentação do processo foi realizada por meio de documento colaborativo na ferramenta *Google Docs*.

O PGD elaborado seguiu a estrutura das ferramentas já existentes. Iniciou-se com a identificação do projeto, finalidades, resumo, autores, área de domínio, cronograma e identificação de possível agente financiador da pesquisa. Em

seguida, as perguntas foram distribuídas em sete categorias: (1) coleta de dados; (2) documentação e metadados; (3) ética e conformidade legal; (4) armazenamento e *backup*; (5) seleção e preservação; (6) compartilhamento de dados; (7) responsabilidade e recursos. Durante a elaboração observou-se que a terminologia utilizada para a construção do plano poderia dificultar a compreensão e a consequente resposta adequada por parte dos pesquisadores. Por isso, adotou-se como estratégia, realizar uma apresentação do PGD para o grupo de pesquisadores antes da etapa de coleta das respostas.

Com a apresentação do PGD, foi possível oferecer um contexto sobre a ciência aberta, *e-Science*, regras para financiamento de pesquisas e gestão dos dados, de forma que os potenciais respondentes pudessem obter um panorama sobre a necessidade e objetivos de um PGD. No segundo momento, cada discente participante da atividade atuou junto a um pesquisador do projeto para coletar as respostas. Nessa etapa, a intenção era identificar possíveis dificuldades na compreensão das questões e colaborar para a construção de um PGD adequado à realidade dos pesquisadores da área da biodiversidade.

Com a aplicação dos questionários, considerando que um dos discentes atendeu simultaneamente a dois pesquisadores, obtiveram-se as respostas necessárias para a construção de PGDs de sete conjuntos de dados, todos no campo das Ciências Biológicas, e correspondentes aos assuntos: (1) flora; (2) cupins; (3) mamíferos; (4) pseudoescorpiões; (5) fauna de solo; e (6) ictiofauna da Paraíba, (7) colêmbolas. Destes seis foram classificados como para fins educacionais e um para prática profissional.

O contato com os pesquisadores indicou que havia a necessidade de adaptações no formulário desenvolvido. A ferramenta precisaria ser mais clara para conseguir ser respondida de forma autônoma por novos pesquisadores. Perguntas que tratavam sobre versionamento, licença de uso e padrão de metadados apresentaram maior dificuldade de compreensão pelos pesquisadores envolvidos na atividade. Para isso, a equipe desenvolveu esclarecimentos para as questões que ensejaram dúvidas, incluindo no formulário conceitos, explicações e exemplos que facilitassem as respostas.

Os pesquisadores também puderam oferecer sugestões de melhoria ao PGD. Foi sugerido pela professora pesquisadora responsável pelo Herbário da UFPB, parametrizar as áreas de domínio com o padrão da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) do Ministério da Educação (MEC); e que as escolhas tanto das áreas como das finalidades não fossem excludentes, já que um conjunto de dados poderia servir a mais de uma área do conhecimento e finalidades.

Nesse sentido, também se observou a necessidade de incluir no formulário um campo para os identificadores digitais dos pesquisadores, tais como *Lattes* e *ORCID*, de modo a facilitar a cooperação efetiva entre os interessados nas pesquisas realizadas por estes sujeitos. O trabalho de desenvolvimento do PGD em conjunto com os pesquisadores da área da biodiversidade possibilitou identificar mais de 40.000 registros já existentes em planilhas do software *Microsoft Excel*, cujos campos ou títulos das colunas (atributos) não seguiam padrões específicos da área alvo da pesquisa. Por isso, optou-se por realizar uma apresentação acerca de um padrão de metadados próprio para a área da biodiversidade, o padrão *DarwinCore*.

A apresentação foi realizada pelos discentes do PPGCI/UFPB para os pesquisadores do PRONEX com o objetivo de subsidiar o planejamento das ações necessárias para a preparação das planilhas, de modo que houvesse uma padronização dos conjuntos de dados antes do depósito a ser realizado no *DataPB*.

O *DataPB* usa o produto de software *Dataverse*, desenvolvido pela universidade de *Harvard* e configura-se como o primeiro repositório de dados da UFPB, no qual podem ser encontrados conjuntos de dados referentes às pesquisas de biodiversidade da Paraíba. Além disso, em uma segunda etapa, o repositório também poderá receber arquivos de outras áreas do conhecimento que possam contribuir com a proposta de compartilhamento de recursos de pesquisa preconizada pela ciência aberta.

4.4 Avaliar

As etapas do ciclo da pesquisa-ação proposta por Tripp (2005) finalizam com a avaliação e a aferição dos resultados da ação, mas não se exaurem em definitivo, dado o caráter contínuo e iterativo do método. No campo da aplicação do PGD junto aos pesquisadores do *Projeto Biodiversidade da Paraíba: Status, Ameaças e Oportunidades*, uma vez atendidos os critérios dispostos nas etapas anteriores, resta nessa seção avaliar e, portanto, contribuir com a melhoria da prática e da própria investigação.

Sob o ponto de vista das oportunidades, preliminarmente, considera-se fundamental o intercâmbio entre o PPGCI/UFPB, o PRONEX e o Programa de Pós-Graduação em Ciências Biológicas da UFPB, para a construção conjunta da proposta de pesquisa relatada. A pesquisa-ação se adequa a esta proposta especialmente por prover uma estreita associação entre a resolução de um problema coletivo por meio da participação dos envolvidos e em formato colaborativo e participativo. Entende-se que este objetivo foi alcançado tanto no processo de elaboração do protótipo do PGD quanto nas interações entre os participantes durante todo o processo. Destaca-se, ainda, a aproximação do método da pesquisa-ação com as fases de desenvolvimento de software das principais metodologias de software ágil disponíveis. O caráter iterativo e incremental de ambos os métodos colaborou, sobremaneira, para a aplicabilidade e a efetividade desta pesquisa.

A etapa final do ciclo da pesquisa-ação se propõe a apresentar os resultados das ações realizadas e, neste caso, a aplicação do PGD elaborado na ferramenta de criação de formulários *Google Forms* foi uma oportunidade identificada para automatizar o processo de aplicação do PGD. A opção pela utilização deste recurso se deu em razão da flexibilidade da ferramenta, tanto para a construção do instrumento de coleta quanto pela possibilidade de alteração das perguntas durante o processo de aplicação com os pesquisadores.

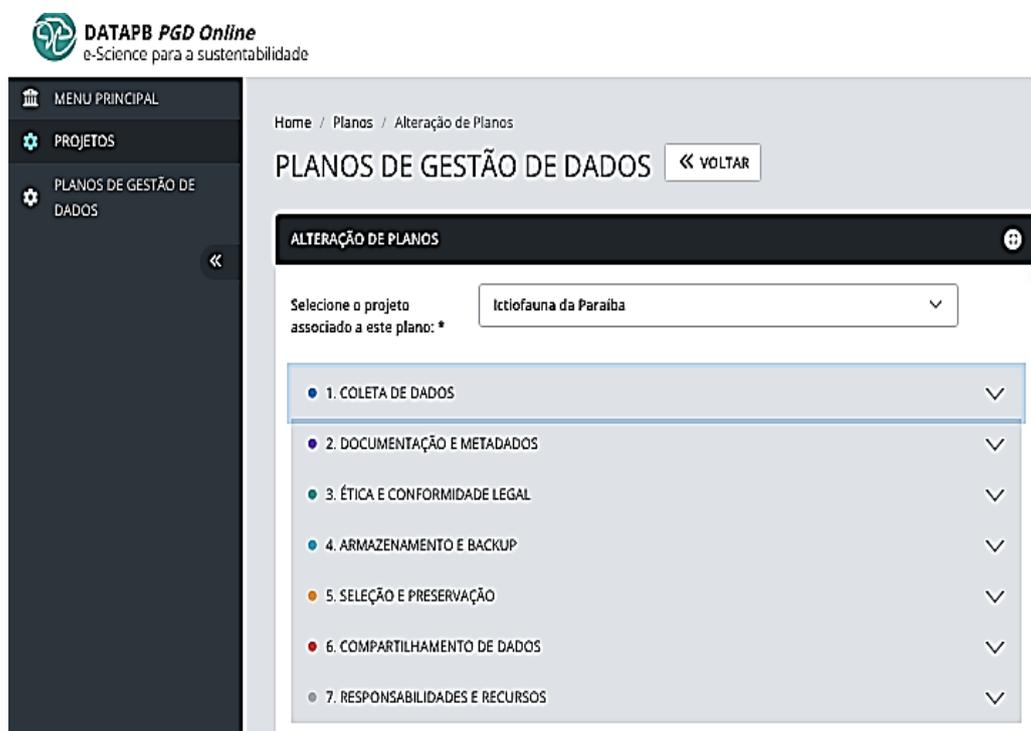
Por outro lado, utilizar formulário desenvolvido com o *Google Forms* foi a alternativa adotada para subsidiar a prototipação de uma ferramenta de software

(em construção) (Figura 2) que permitirá a geração dos mais variados PGDs no âmbito do *DataPB*, e para o que se destaca o compartilhamento das informações e consequente amadurecimento dos questionamentos propostos para o preenchimento dos PGDs durante a aplicação entre os participantes.

O software em desenvolvimento propõe a inserção dos metadados do PGD identificado pelo projeto ao qual está vinculado. O protótipo apresentado utilizou a linguagem Java e tecnologia baseada em *Servlets/Java Server Pages (JSP)* como base em seu desenvolvimento, com a persistência de dados em banco de dados relacional *PostgreSQL*.

O formulário apresenta campos pré-formatados obedecendo a uma sequência de preenchimento (não taxativa), com orientações e textos explicativos para facilitar a construção do PGD (Araújo *et al.*, 2023) para que a comunidade científica possa utilizar e participar da construção da solução com a realização de testes e sugestões de melhorias no sistema.

Figura 2 - Prévia do software em construção



Fonte: Elaborado pelos autores.

O *download* do PGD em formato PDF (*Portable Document Format*) é possível através da utilização de uma implementação sobre a estrutura do *JasperReports/IRreport*.

Figura 3 - Prévia do PGD em formato PDF



DATA PB
e-Science para a sustentabilidade
PGD Online
PLANO DE GESTÃO DE DADOS

PROJETO Ictiofauna da Paraíba
Finalidade: Fins educacionais **Domínio:** Ciências Biológicas
Integrantes:

Resumo: Síntese do conhecimento da ictiofauna marinha e de água doce do estado da Paraíba

Data de início: 01/08/2023 **até** 08/01/2025
Possui agente financiador? SIM - FAPESQ, CNPq

1. COLETA DE DADOS

1.1 Quais dados você coletará ou criará?
Dados referentes às espécies de peixes capturadas na Paraíba.

1.1.1 Que tipo, formato e volume de dados?
Planilhas excel no formato Darwin Core, com aproximadamente 9000 registros.

1.1.2 Os formatos e software escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?
Estamos usando o formato de Excel e que permite o compartilhamento e acesso a depender de compatibilidade das versões futuras do Excel. As informações coletadas em planilha são exportadas para a plataforma de Specify.

1.1.3 Existem dados existentes que você pode reutilizar?
Existe uma planilha elaborada e homologada pela coleção de ictiologia da UFPB, sobre a qual continuaremos ingressando dados dos diferentes espécimes coletados.

Fonte: Elaborado pelos autores.

O protótipo desenvolvido permite ainda o *download* do PGD em formato de dados aberto e legível por máquina (XML), mediante a seleção de opção no menu de ações específico para esse fim.

Figura 4 - Prévia do PGD em formato XML

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<pgd id="6">
  <projeto>
    ...
  </projeto>
  <dados>Dados referentes às espécies de peixes capturadas na Paraíba.</dados>
  <tipoforमतovolume>Planilhas excel no formato Darwin Core, com aproximadamente 9000 registros.</tipoforमतovolume>
  <formatocompartilhamento>Estamos usando o formato de Excel e que permite o compartilhamento e acesso a depender de compatibilidade das versoes futuras do Excel. As informacoes coletadas em planilha sao exportadas para a plataforma de Specify.</formatocompartilhamento>
  <dadosparareutilizar>Existe uma planilha elaborada e homologada pela colecao de icitologia da UFPB, sobre a qual continuaremos ingressando dados dos diferentes espécies coletadas.</dadosparareutilizar>
  <coleta>Individuos das diferentes espécies presentes no estado, sao coletados com diferentes técnicas, dependendo da localidade de coleta.</coleta>
  <metodologia>Sao usados diferentes aparelhos de pesca, como tarrafa, redes de arrastros, espínhel, dependendo do habitat amostrado. Também sao usados dados provenientes da pesca artesanal. Depois da captura dos individuos, eles sao fixados no formol, etiquetados com as informacoes pertinentes como localidade de captura, coletor, data de coleta. Posteriormente no laboratorio, os individuos sao identificados taxonomicamente por especialistas e levados para alcool, para o posterior ingresso a colecao. Cada lote de captura leva uma etiqueta com toda esta informacao.</metodologia>
</pgd>
```

Fonte: Elaborado pelos autores.

Com vistas à evolução do sistema, novas contribuições também podem apresentar requisitos capazes de ampliar o porte da ferramenta, sobretudo ao se comparar com outras aplicações consolidadas globalmente. Nesse sentido, vale a pena destacar algumas ferramentas amplamente utilizadas que podem servir de alicerce para tais melhoramentos.

O *DMP OPIDoR* é uma ferramenta para criação de planos de gestão de dados desenvolvida e mantida pelo Inist-CNRS (*Institut de l'Information Scientifique et Technique*), e se baseia na estrutura *DMPRoadmap* de código-fonte aberto que foi implementada sob o framework *Ruby on Rails* e entrega funcionalidades como: criar e manter diferentes versões de PGD, fornecer orientações sobre gerenciamento de dados e financiamento das pesquisas, exportar planos de gestão de dados e permitir o trabalho colaborativo. Instituições de ensino e pesquisa francesas, a exemplo da *Université Savoier Mont Blanc* e da *Université de Montpellier* tem adotado soluções com base na *DMP OPIDoR* para o fornecimento dos dados administrativos dos seus projetos de pesquisa em prol da Ciência Aberta (Institut de l'Information Scientifique et Technique, 2025).

O *Data Stewardship Wizard* é uma plataforma de código aberto que se propõe a incrementar as pesquisas por meio do planejamento e gerenciamento de dados colaborativos e dinâmicos. Sua utilização em forma de *wizard* oferece as orientações de preenchimento dependendo do contexto do usuário, indicando para cada resposta a adesão aos princípios FAIR para que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis por máquinas e humanos. Foi desenvolvido pela *Czech Technical University* em Praga, em linguagem *Python*. Combina a estrutura de mapas mentais com perguntas aninhadas e *links* cruzados, conectando diferentes nós. Gera PGD textual nos formatos HTML, PDF, MS Word e LaTeX (CTU, 2025).

O *NSD DMP, framework* construído pelo *Norwegian Centre for Research Data*, inova em relação ao compartilhamento de dados de pesquisa. Ao combinar elementos acionáveis por máquina com conceitos da modelagem de processos BPMN (*Business Process Model and Notation*), incrementa o gerenciamento e planejamento de dados para suportar a automação e integração de sistemas num contexto institucional. Novos requisitos incluem princípios FAIR, Ciência Aberta e regulações a respeito da proteção de dados pessoais conforme a GDPR Europeia (*General Data Protection Regulation*) (Kvamme, 2024).

Argos é uma ferramenta desenvolvida pela OpenAIRE para apoio às atividades de gerenciamento de dados de pesquisa e elaboração de planos de gestão de dados. Também adota melhores práticas voltadas aos princípios FAIR e código aberto e aplica padrões para DMPs acionáveis por máquina. Sua estrutura se baseia no OpenDMP e utiliza Java e Node como linguagens de programação de base. Os dados gerados são persistidos em banco de dados Postgresql (OpenAIRE, 2025).

Já a ferramenta TU Wien DMP Tool, disponível aos membros da Universidade Técnica de Viena (TU Wien), fornece um guia de preenchimento do plano de gestão de dados em dez etapas, com informações de apoio e sugestões de texto para as respostas e permite download do plano de gestão de dados em formato MS Word e JSON. Se ancora na implementação base do DMP Tool (TU Wien, 2025).

A partir da perspectiva prática e das pesquisas realizadas, por meio da metodologia empregada, se constatou que a abordagem interdisciplinar que norteou o desenvolvimento da disciplina ampliou o conhecimento entre as áreas e oportunizou momentos de trocas de experiências e apropriação dos conceitos apresentados.

No que se refere à possíveis ameaças que possam interferir no processo de proposição de um modelo de PGD, foi possível observar que mudanças nas exigências dos órgãos que demandam os PGDs, o volume dos dados e a sua especificidade de descrição, a forma como os diversos pesquisadores sistematizam os seus dados, armazenando-os em aplicações sem medidas de segurança da informação poderão impactar no uso do modelo.

Pelo fato do PGD se caracterizar como uma exigência recente dos órgãos de fomento à pesquisa, foi possível observar a falta de padronização na estruturação dos arquivos de dados e em sua salvaguarda. Tais ameaças reforçam o valor que o PGD possuirá para os pesquisadores, uniformizando suas práticas de sistematização de dados, ao assumir um compromisso de contribuir com a ciência aberta ao publicar os seus dados de forma estruturada em repositórios de dados.

Por fim, a própria característica do PGD enquanto condição dinâmica e passível de aprimoramentos torna-o mais preciso e sólido ao longo do tempo, proporciona ganhos em maturidade e alcance de boas práticas, aprimora a eficiência do pesquisador e permite que os seus dados sejam compreensíveis e reutilizáveis. Desse modo, aliar teoria e prática de forma criteriosa demonstrou ser uma opção viável, tendo em vista que conseguiu documentar e equilibrar os objetivos da ação realizada e da pesquisa, com contribuições práticas assim como para a geração de conhecimentos científicos.

5 Considerações finais

Agências de fomento e Instituições de Ensino e Pesquisa têm solicitado de seus pesquisadores a elaboração de PGDs, a fim de registrá-los e disponibilizá-los para

a sociedade em geral, democratizando seu acesso e ampliando as possibilidades de investigação acerca dos objetos de estudo. Trata-se de uma tendência cada vez mais comum, assim, é preciso discutir para além dos conceitos, compartilhando técnicas e estratégias de elaboração.

Dessa forma, mais pesquisadores poderão compreender sobre a construção deste documento. A partir desse entendimento, surge esta pesquisa que objetivou registrar os processos de elaboração, aplicação e avaliação de um PGD desenvolvido para a área da Biodiversidade, no âmbito da UFPB.

Como resultados, foram elencadas quatro etapas de desenvolvimento do Plano, respectivamente: (a) Planejar; (b) Agir; (c) Descrever; e (d) Avaliar. A primeira delas foi realizada a partir de reuniões de alinhamento com a equipe responsável, ocasião em que foram definidas as seguintes ações: pesquisa e estudo acerca de modelos e orientações para elaboração do PGD; elaboração do PGD; teste de aplicação com pesquisadores da biodiversidade; avaliação do modelo construído. Além da descrição das ações, também foram determinados os pesquisadores responsáveis por cada uma delas e os prazos.

Cabe ressaltar a importância de uma equipe diversificada, neste caso, formada por pesquisadores das áreas da Arquivologia, Biblioteconomia, Tecnologia, Comunicação, Biodiversidade, entre outros. Essa interdisciplinaridade é primordial para o intercâmbio de saberes que contribuam significativamente para a construção de um produto final eficiente. Além disso, foi considerada quando na distribuição de responsabilidades.

Após o planejamento, deu-se início a etapa “agir” com a elaboração do PGD. Para tanto foram adotados os modelos do DCC e *DMPTools* como base para o documento a ser criado. Isto em razão da consolidação dessas organizações no contexto da gestão de dados, bem como sua aplicação frequente através de produtos de *software* voltados para a gestão eletrônica de PGDs. Ainda nesta etapa, após as devidas revisões e ajustes no plano, partiu-se para a aplicação teste, com a participação de pesquisadores do Núcleo de Excelência em Biodiversidade da Paraíba, vinculado ao Programa de Pós-Graduação em Ciências Biológicas da

UFPB. O teste serviu para a identificação de inconsistências no formulário, bem como termos e tópicos a serem melhor esclarecidos, quesitos que foram discutidos na etapa de avaliação.

Para que o processo ficasse registrado, a fim de facilitar o acompanhamento dos avanços e planejamento, orientou-se a redação de um diário de pesquisa, com anotações das percepções da equipe sobre ações e resultados obtidos. É a partir dele que seguiu-se para a etapa seguinte: descrever.

A descrição é primordial para que a equipe, e demais investigadores interessados, pudessem visualizar o processo desde sua gênese, observando a sistemática adotada e os avanços alcançados. Esse registro auxiliou, sobretudo, a etapa final de avaliação, momento em que foram discutidas melhorias a serem aplicadas e definidos os próximos passos.

Salienta-se que a criação do PGD visa sua aplicação no Repositório de dados DataPB, vinculado à UFPB e voltado para a área da Biodiversidade, pretendendo-se sua ampliação, ainda, para outras áreas do conhecimento. Para otimizar sua eficácia, encontra-se em construção o *software* que permitirá a criação eletrônica do PGD no Repositório, apresentando os campos orientados no documento desenvolvido.

Cabe destacar que esta é uma iniciativa inovadora, tendo em vista que poucas universidades no Brasil não dispõem de repositório de dados, e, na UFPB essa é a primeira, fato que reitera o caráter desafiador desta proposta. Compartilhar experiências relacionadas desde a elaboração à análise de PGD é imprescindível para que outros pesquisadores, de quaisquer áreas do conhecimento consigam compreender sua importância e, principalmente, a elaboração, estimulando boas práticas no contexto da gestão de dados científicos.

Recomenda-se, por fim, o desenvolvimento de outras pesquisas acerca desta temática. Além disso, reforça-se o intuito de dar continuidade ao processo de desenvolvimento do software e sua aplicação no Repositório de Dados, percurso que, decerto, consistirá em motriz geradora de futuras pesquisas.

Financiamento

À Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ/PB), Termo no 1969/2022, referente a pesquisa de Pós-Doutorado desenvolvida no Programa de Pós-Graduação em Ciência da Informação na Universidade Federal da Paraíba (PPGCI/UFPB), assim como ao Edital FAPESQ/PB 09/2021, demanda universal. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo 311563/2018-0.

Agradecimentos

Aos pesquisadores do Núcleo de Excelência em Biodiversidade da Paraíba, vinculado ao Programa de Pós-Graduação em Ciências Biológicas da UFPB, pela parceria e troca de experiências ao longo de todo processo.

Referências

ALMEIDA; J. L.; PERUCCHI, V; FREIRE; G. H. A. A Pesquisa-Ação como estratégia metodológica na Ciência da Informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 9, n. 3, p. 130-146, 2019. Disponível em: <http://dx.doi.org/10.21714/2236-417X2019v9n3p130>. Acesso em: 21 nov. 2023.

ARAÚJO, J. F. *et al.* **Plano de Gestão de Dados**. João Pessoa: [s. n.], 2023.

AZAM, A. The first rule of data science. **Berkeley Science Review**, Berkeley, 27 Apr. 2014.

BRASIL. Ministério da Ciência e Tecnologia. Portaria n. 693, de 20 de agosto de 2009. Institui, no âmbito do Programa de Pesquisa em Biodiversidade - PPBio, a Política de Dados. **Diário da Justiça**: seção 1, Brasília, ano 146, n. 160, p. 8, 21 ago. 2009.

COSTA, M. M.; CUNHA, M. B. A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 287-309, 2019. Disponível em: <https://doi.org/10.18617/liinc.v15i2.4763>. Acesso em: 21 fev. 2024.

CURTY, R. G; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais [...]**. Marília: Unesp, 2017.

CTU. **Data Stewardship Wizard**. Praga: Czech Technical University, 2025.

DATAPB. **DataPB: e-Science para a sustentabilidade**. João Pessoa: Universidade Federal da Paraíba, 2024.

DIGITAL CURATION CENTER. **DMP online**. [S. l.: s.n.], 2023.

DIJCK, J. V.; POELL, T.; WAAL, M. **The Platform Society: public values in a connective world**. Oxford: Oxford University Press, 2018.

DUDZIAK, E. A. Políticas de competência em informação: leitura sobre os primórdios e a visão dos pioneiros da information literacy. *In*: ALVES, F. M. M.; CORRÊA, E. C. D.; LUCAS, E. R. O. (org.). **Competência em informação: políticas públicas, teoria e prática**. Salvador: EDUFBA, 2016.

ENGEL, G. I. Pesquisa-ação. **Educar**, Curitiba, n. 16, p. 181-191, 2000. Disponível em: <https://doi.org/10.1590/0104-4060.214>. Acesso em: 15 abr. 2024.

FAPESP. **Gestão de dados**. São Paulo: Fundação de Amparo à Pesquisa do Estado de São Paulo, 2023.

FERRER-SAPENA, A.; SÁNCHEZ-PÉREZ, E. A. Open data, big data: ¿hacia dónde nos dirigimos? **Anuario ThinkEPI**, Valencia, v. 7, n. 1, p. 150-156, 2013.

FERNANDES, H. D. H.; OLIVEIRA, A. F. Gestão da Preservação digital em repositórios de dados de pesquisa. **Revista Ibero-americana de Ciência da Informação**, Brasília, v. 11, n. 1, p. 255-273, 2018. Disponível em: <https://doi.org/10.26512/rici.v11.n1.2018.8541>. Acesso em: 10 ago. 2020.

FIOCRUZ. **Sobre o FioDMP**. Rio de Janeiro: FIOCRUZ, 2025.

FUNDAÇÃO ALFRED SLOAN. **Alfred P. Sloan Foundation**. New York: Fundação Alfred Sloan, 2025.

FAPESQ-PB. **Edital n. 18/2022 - SEECT-FAPESQ/PB**. João Pessoa: Fundação de Apoio à Pesquisa do Estado da Paraíba, 2022.

GALVÃO, M. C. B.; RICARTE, I. L. M.; DAURA, A. P. Tecnologia e informação em saúde: modelo de ensino-aprendizagem transdisciplinar. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 4, p. 73-94, 2011. Disponível em: <https://doi.org/10.1590/S1413-99362011000400007>. Acesso em: 3 nov. 2023.

GREEN, A.; MACDONALD, S.; RICE, R. **Policy-making for research data in repositories: a guide**. Edinburgh: DISC-UK/ DataShare Project, 2009.

HIGGINS, S. The DCC curation lifecycle model. **International Journal of Digital Curation**, Edinburgh, v. 3, n. 1, p. 134-140, 2008. Disponível em: <https://doi.org/10.2218/ijdc.v3i1.48>. Acesso em: 21 jun. 2024.

HIMANEN, L. *et al.* Data-driven materials science: status, challenges, and perspectives. **Advanced Science**, New Jersey, p. 1-23, 2019. Disponível em: <https://doi.org/10.1002/advs.201900808>. Acesso em: 12 nov. 2023.

IBICT. Fundação Alfred **BR**. Rio de Janeiro: IBICT, 2025.

INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE. **DMP OPIDoR**. [S. l.: s.n.], 2025.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH. **Sample data management plan for depositing data with ICPSR**. [S. l.: s.n.], 2025.

KOERICH, M. S. *et al.* Pesquisa-ação: ferramenta metodológica para a pesquisa qualitativa. **Revista Eletrônica de Enfermagem**, Goiânia, v. 11, n. 3, p. 717-23, 2009. Disponível em: <https://doi.org/10.5216/ree.v11.47234>. Acesso em: 20 nov. 2023.

KVAMME, T. **NSD DMP: enabling long-term preservation and sharing of Research Data**. [S. l.]: Norwegian Centre for Research Data, 2024.

MARQUES, L. F. S; SAYÃO, L. F. Conectando a eScience à Ciência da Informação: o big metadado científico e suas funcionalidades. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 21, p. 1-21, 2023. Disponível em: <https://doi.org/10.20396/rdbci.v21i00.8673740>. Acesso em: 6 fev. 2024.

MAYER-SCHONBERGER, V.; CUKIER, K. **Big data: a revolution that will transform how we live, work, and think**. New York: Houghton Mifflin Harcourt, 2013.

MICHENER, W. K. Ten simple rules for creating a good data management plan. **PLOS Computational Biology**, California, v. 11, n. 10, p. e1004525, 2015. Disponível em: <https://doi.org/10.1371/journal.pcbi.1004525>. Acesso em: 21 dez. 2023.

NATIONAL ENDOWMENT FOR THE HUMANITIES. **Notice of funding opportunity**. [S. l.: s.n.], 30 jun. 2021.

NATIONAL INSTITUTES OF HEALTH. **Final NIH policy for data management and sharing.** [S. l.: s.n.], 25 jan. 2023.

OPENAIRE. **Argos.** [S. l.: s.n.], 2025.

PESET, F.; FERRER-SAPENA, A.; SUBIRATS-COLL, I. Open data y linked open data: su impacto en el área de bibliotecas y documentación. **El profesional de la información**, Spain, v. 20, n. 2, p. 165-173, 2011. Disponível em: <https://doi.org/10.3145/epi.2011.mar.06>. Acesso em: 4 fev. 2025.

SALES, L. F.; SAYÃO, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores.** Rio de Janeiro: CNEN/IEN, 2015.

SALES, L. F.; SAYÃO, L. F. Proposta de modelo de serviço de gestão de dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 11, p. 1-13, 2022. Disponível em: <https://doi.org/10.5380/atoz.v11i0.85765>. Acesso em: 12 fev. 2025.

SALES; L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, p. 31-48, 2019. Disponível em: <https://doi.org/10.47681/rca.v4i1.26337>. Acesso em: 30 out. 2020.

SCIENCE EUROPE. **Practical guide to the international alignment of research data management:** extended edition. Brussels: Science Europe, 2021.

THIOLLENT, M. **Metodologia da pesquisa-ação.** São Paulo: Cortez; 1986.

TRIPP, D. Action research: a methodological introduction. **Educação e Pesquisa**, São Paulo, v. 31, n. 3, p. 443-466, 2005. Disponível em: <https://doi.org/10.1590/S1517-97022005000300009>. Acesso em: 12 fev. 2025.

UFPB. **Com aprovação de novo PRONEX pela FAPESQ, mapeamento da biodiversidade da Paraíba auxiliará na construção de políticas voltadas à sustentabilidade.** João Pessoa: Universidade Federal da Paraíba, Centro de Ciências Aplicadas e Educação, 2023.

TU WIEN. **DMP TU Wien:** data management plans. Vienna: Universidade Técnica de Viena, 2025.

VEIGA, V. *et al.* **Plano de gestão de dados de pesquisa:** guia de elaboração. Rio de Janeiro: Fiocruz/Icict, 2022.

VEIGA, V. *et al.* Plano de gestão de dados FAIR: uma proposta para a Fiocruz. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 275-286, 2019. Disponível em: <https://doi.org/10.18617/liinc.v15i2.5030>. Acesso em: 12 fev. 2025.

Preparation, application and evaluation of a data management plan: the DataPB case

Abstract: In contemporary science, Higher Education Institutions, Funding Agencies, Government Bodies, and Research Institutions have been demanding the development of a Data Management Plan, a document that establishes mandatory procedures for research data management intended to aid throughout the entire academic research process. Against this backdrop, this research aimed to analyze the process of creation and applicability of the Data Management Plan in the field of Biodiversity within the Federal University of Paraíba. The investigation was qualitative, applied, and interventionist in nature. The action-research method was utilized to describe and evaluate the implementation of the Plan. The findings delineate a process that included stages of: (a) planning, involving preliminary research on existing Data Management Plans, goal setting, task organization, assigning responsibilities, and establishing deadlines; (b) action, via a trial application of the developed Plan; (c) description, the phase of systematizing the planning process, drafting, and proposing enhancements; and (d) evaluation, based on analyzing the Plan to correct inconsistencies and refine the final outcome. It is concluded that documenting and sharing the process of the DMP's development, application, and evaluation is beneficial for the exchange of ideas and experiences. It also aids in encouraging other researchers and Higher Education Institutions to adopt this tool for managing the data of their inquiries, thus broadening the spectrum of research possibilities and data transparency through publication.

Keywords: data management plan; e-science; search data; big data

Declaração de autoria

Concepção e elaboração do estudo: Adriana Alves Rodrigues, Joana Araújo, Pedro Felipy e Vivianne de Queiroz Leal

Coleta de dados: Joana Araújo, Pedro Felipy e Vivianne de Queiroz Leal

Análise e interpretação de dados: Joana Araújo, Pedro Felipy e Vivianne de Queiroz Leal

Redação: Adriana Alves Rodrigues, Joana Araújo, Pedro Felipy e Vivianne de Queiroz Leal

Revisão crítica do manuscrito: Adriana Alves Rodrigues, Guilherme Ataíde Dias, Alzira Karla Araújo da Silva, Wagner Junqueira de Araújo

Declaração de disponibilidade de dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo.

Autoria para correspondência

Adriana Alves Rodrigues

profadrianaalves@gmail.com

Editor-chefe

Thiago Henrique Bragato Barros

Como citar

RODRIGUES, Adriana Alves; ARAÚJO, Joana Ferreira; SILVA, Pedro Felipy Cunha; LEAL, Vivianne de Queiroz; DIAS, Guilherme Ataíde; SILVA, Alzira Karla Araújo; ARAÚJO, Wagner Junqueira. Elaboração, aplicação e avaliação de um Plano de Gestão de Dados: o caso DataPB. **Em Questão**, Porto Alegre, v. 31, e-139676, 2025. <https://doi.org/10.1590/1808-5245.31.139676>

Parecer(es) aberto(s):

<https://doi.org/10.1590/1808-5245.31.139676A>

Recebido: 15/04/2024

Aceito: 12/03/2025



¹ No original: “*Ahora queremos que los datos también sean accesibles, libre y gratuitamente*”.

² Proposto pela *DataONE* e desenvolvido no contexto da *National Science Foundation* (NSF), inclui oito etapas: planejar, coletar, garantir, descrever, preservar, descobrir, integrar e analisar. Estas são largamente utilizadas no campo da Ciência da Informação.

³ Dados encontráveis, acessíveis, interoperáveis e reutilizáveis.

⁴ University of California Curation Center (UC3), California Digital Library, DataONE, Digital Curation Centre (DCC-UK), Instituto Smithsonian, University of California, Los Angeles

Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library.

⁵No original: “1) Determine the research sponsor’s requirements; 2) Identify the data to be collected; 3) Define how the data will be organized; 4) Explain how the data will be documented; 5) Describe how data quality will be ensured; 6) Present a sound data storage and preservation strategy; 7) Define the project’s data policies; 8) Describe how the data will be disseminated; 9) Assign roles and responsibilities; 10) Prepare a realistic budget.”

⁶Têm como objetivo tornar os dados encontráveis, acessíveis, interoperáveis e reutilizáveis.

⁷Liderado pelo professor Rafael Raimundo (UFPB) e professora Laure Berti-Equille (IRD SPACE DEV).