

IMPLEMENTAÇÃO DE DIRETRIZES PARA ESTRUTURAÇÃO E GESTÃO DE DADOS EM UM LABORATÓRIO DE PESQUISAS

Ieda Pelógia Martins Damian

Doutora em Administração de Organizações pela Universidade de São Paulo, Brasil. Professora da Universidade de São Paulo, Brasil

E-mail: iedapm@usp.br

Ildeberto Rodello

Doutor em Física Aplicada pela Universidade de São Paulo, Brasil.
Professor da Universidade de São Paulo, Brasil.

E-mail: rodello@usp.br

Marcia Regina da Silva

Doutora em Educação pela Universidade Federal de São Carlos. Professora da Universidade de São Paulo, Brasil.

E-mail: marciaregina@usp.br

Carla Barreto Dória Pinto

Especialista em Planejamento Estratégico e Concepção em *Branded Content* pelo Serviço Nacional de Aprendizagem Comercial, Brasil.

E-mail: carlallc@hotmail.com

Resumo

Este artigo apresenta a implementação de diretrizes para a estruturação e gestão de dados de pesquisa no Laboratório de Estudos e Pesquisas em Educação e Economia Social, da Universidade de São Paulo. O objetivo principal é otimizar a organização, o armazenamento e a recuperação de dados, promovendo a transparência, a reprodutibilidade e a segurança dos dados científicos. O processo de implementação inclui o desenvolvimento de um protocolo para padronização de nomeação de arquivos e organização de pastas compartilhadas, com o uso de uma ferramenta computacional específica, o Praticidados, para automatizar a renomeação dos arquivos. Além disso, estabelece procedimentos para a transferência de dados, o versionamento de arquivos e a criação de pastas temáticas. A estratégia foi dividida em etapas, com treinamento e suporte contínuo às equipes envolvidas. O artigo discute os desafios enfrentados, as soluções adotadas e os impactos positivos na governança dos dados do Laboratório, ressaltando a importância dessas diretrizes para fortalecer a Ciência Aberta e o compartilhamento de dados científicos.

Palavras-chave: gestão de dados; laboratório de pesquisa; ciência aberta; dados científicos.

IMPLEMENTATION OF GUIDELINES FOR DATA STRUCTURING AND MANAGEMENT IN A RESEARCH LABORATORY

Abstract

This article presents the implementation of guidelines for structuring and managing research data at the Laboratory for Studies and Research in Education and Social Economy at the University of São Paulo. The main objective was to optimize the organization, storage, and retrieval of data, promoting transparency, reproducibility, and security of scientific information. The implementation process included the

development of a protocol for naming files and organizing shared folders, using a specific computational tool, Praticidados, to automate file renaming. In addition, procedures were established for data transfer, file versioning, and the creation of thematic folders. The strategy was divided into stages, with training and ongoing support for the teams involved. The article discusses the challenges faced, the solutions adopted, and the positive impacts on data governance at the Laboratory, highlighting the importance of these guidelines for strengthening Open Science and the sharing of scientific data.

Keywords: data management; research laboratory; open science; scientific data.

1 INTRODUÇÃO

O gerenciamento de dados de pesquisa se tornou um desafio crescente para pesquisadores, laboratórios e grupos de pesquisas, exigindo engajamento e mudança de postura quanto às estratégias de administração, organização, armazenamento e proteção desses dados. Esse cenário é impulsionado pelas diretrizes do Movimento da Ciência Aberta, cujo principal objetivo é tornar a pesquisa e os dados científicos mais acessíveis (Borgman, 2015). Além disso, a entrada em vigor da Lei Geral de Proteção de Dados Pessoais (LGPD) - Lei nº 13.709/2018, trouxe novas preocupações especialmente no que diz respeito às especificidades metodológicas da pesquisa científica.

Os dados de pesquisa desempenham um papel fundamental no avanço de investigações científicas e no ensino da ciência. A Gestão de Dados (GD) envolve uma série de procedimentos estratégicos que engloba desde a definição de políticas de governança até a implementação de rotinas voltadas ao uso de ferramentas e tecnologias, como seleção e implementação de banco de dados, integração de sistemas, limpeza e normalização de dados, realização de backups e aplicação de medidas de segurança que garantam a privacidade e integridade dos dados (Wilkinson *et al.*, 2016).

As agências de fomento têm seguido os direcionamentos de escopo mundial para o armazenamento e compartilhamento de dados de pesquisa, orientando pesquisadores e instituições sobre a necessidade de elaborar planos para a gestão de dados. No entanto, a gestão de dados de pesquisa ainda enfrenta desafios decorrentes de limitações na infraestrutura tecnológica e de fatores institucionais, como políticas, apoio organizacional e incentivos às práticas de gestão de dados (Marlina; Hidayanto; Purwandari, 2025). Além disso, a falta de diretrizes padronizadas e de um reconhecimento efetivo da autoria dos dados dificulta sua disseminação e reuso.

Para enfrentar esses desafios, foram estabelecidos os princípios FAIR (*Findable, Accessible, Interoperable and Reusable*), diretrizes essenciais para a governança dos dados científicos. Esses princípios incentivam a adoção de boas práticas que asseguram a rastreabilidade e a interoperabilidade dos dados em diferentes plataformas e domínios científicos (Wilkinson *et al.*, 2016).

Ferguson *et al.* (2014) identificaram uma série de obstáculos que dificultam a integração, o compartilhamento e o reuso dos dados de pesquisa. Entre eles destacam-se a falta de infraestruturas tecnológicas e gerenciais, bem como de políticas institucionais que assegurem a estabilidade, a persistência e a interoperabilidade. Além disso, a ausência de controle de qualidade e a padronização representa um desafio, pois a natureza heterogênea e fragmentada dessas coleções exige estratégias diversificadas para sua gestão. Outros entraves incluem a falta de políticas voltadas para a publicação de dados; a inexistência de esquemas de reconhecimento da autoria e de incentivos para a organização e disseminação. Por fim, há também a falta de interesse por parte dos pesquisadores em divulgá-los, além dos seus círculos profissionais mais próximos, especialmente quando se trata de hipóteses não confirmadas, resultados negativos e dados considerados auxiliares de estudos.

A invisibilidade de um segmento significativo da pesquisa científica gera descontinuidade e enfraquece a ciência enquanto empreendimento social e humanístico, trazendo diversos desdobramentos. Os mais contundentes são: o registro incompleto e tendencioso dos processos de geração de conhecimento, criando lacunas nas memórias acadêmicas das instituições; a duplicação de esforços, prolongando, desnecessariamente, o ciclo de comunicação científica e, sobretudo, a impossibilidade de se garantir os princípios fundamentais, como a reprodutibilidade dos experimentos científicos e a autocorreção da ciência (Ferguson *et al.*, 2014).

Nesse contexto, observa-se uma diferença marcante nos processos de gestão de dados. Enquanto grandes projetos recebem ampla atenção na formação de coleções, preservação e reutilização, os dados gerados por pesquisadores individuais em projetos de menor escala costumam ser negligenciados. No entanto, novas estruturas sociais e avanços técnicos podem aumentar significativamente a disponibilidade e o valor desses dados. O desafio para a política científica e para as instituições de pesquisa é desenvolver infraestruturas e práticas, como repositórios digitais disciplinares, que possibilitem o seu aproveitamento de forma ampla e útil para a sociedade (Van Gend; Zuiderwijk, 2023).

Da mesma maneira, a proteção de dados é um pilar fundamental da governança da informação, especialmente no caso de dados sensíveis provenientes de pesquisas científicas. Sayão e Sales (2014) destacam a importância da implementação de políticas institucionais que garantam a segurança dos dados, abrangendo o uso de criptografia, autenticação de usuários e auditorias periódicas. Essas medidas são essenciais para assegurar que os dados armazenados em repositórios institucionais e disciplinares estejam protegidos contra acessos não autorizados e possíveis vazamentos.

A disponibilidade e o compartilhamento de dados de pesquisa são fundamentais para favorecer a transparência, a reprodutibilidade e a reutilização dos resultados científicos. Entretanto, a consolidação dessas práticas requer mecanismos de incentivo e reconhecimento que valorizem o compartilhamento de dados no âmbito da avaliação científica e institucional. Nesse contexto, a adoção dos princípios FAIR e o depósito dos dados em repositórios adequados contribuem para ampliar sua encontrabilidade, acessibilidade, interoperabilidade e reutilização, favorecendo novas análises e colaborações científicas (Hughes *et al.*, 2023).

Inserido no contexto das pesquisas e enfrentando os desafios da gestão de dados, o Laboratório de Estudos e Pesquisas em Educação e Economia Social (LEPES), fundado em 2011, na Universidade de São Paulo, desenvolve pesquisas e realiza avaliações de impacto de políticas sociais em todo o Brasil. Ao longo desse período, o LEPES gerou uma grande quantidade de dados de pesquisas, armazenados em um servidor de acesso restrito e em computadores locais. Esse modelo de armazenamento, no entanto, apresenta diversas limitações, como dificuldades de acesso, falta de padronização, fragilidades na segurança, problemas de localização e gerenciamento de versões, entre outros.

O LEPES tem se destacado pela excelência e abrangência de seus projetos, impactando diretamente diversos seguimentos sociais. A decisão de elaboração de um projeto de gestão de dados surgiu dos gestores do Laboratório e da equipe de gestão de dados, estabelecida em 2020. Composta por pesquisadores, estagiários e consultores, incluindo docentes das áreas de Gestão e Ciência da Informação, essa equipe teve como principal objetivo definir diretrizes para a adequação à LGPD e a implementação de um Plano de Gestão de Dados (PGD).

Diante de tal cenário, esse artigo busca responder a seguinte questão: como alinhar o desenvolvimento de um PGD voltado ao campo educacional com as práticas de gestão de dados? Para isso, seu principal objetivo é apresentar as etapas da criação do PGD desenvolvido para o LEPES. Esta pesquisa pode ser considerada como relato de caso, pois reflete e detalha os procedimentos adotados e os resultados obtidos na elaboração desse Plano em um contexto específico.

A maioria das pesquisas sobre gestão de dados está direcionada ao contexto organizacional, deixando uma lacuna de desenvolvimento de pesquisas de GD no âmbito da pesquisa científica. No campo da Ciência da Informação, essa temática ainda carece de uma exploração mais aprofundada, evidenciando a necessidade de novas investigações sobre o assunto.

2 PROCEDIMENTOS METODOLÓGICOS

Uma pesquisa aplicada, de abordagem qualitativa, com objetivos exploratório-descritivos foi desenvolvida por meio de um estudo de caso único. Com o objetivo de produzir conhecimento acerca da gestão de dados voltada ao campo educacional, justifica-se o uso da pesquisa aplicada que, de acordo com Gil (2022), é indicada para gerar conhecimento para a solução de problemas específicos.

A abordagem qualitativa se deve ao intuito de analisar práticas e processos organizacionais em seu contexto, com a possibilidade de considerar percepções, necessidades e experiências dos envolvidos (Creswell; Poth, 2018). O caráter exploratório, por sua vez, advém do levantamento de práticas e necessidades relacionadas à gestão dos dados, enquanto o caráter descritivo se associa à sistematização das etapas de elaboração e implantação das diretrizes (Gil, 2022).

Para a elaboração e implementação de diretrizes para gestão dos dados de pesquisa no LEPES, utilizou-se estudo de caso como estratégia de investigação por se tratar de uma análise detalhada de um fenômeno contemporâneo em seu contexto real (Yin, 2018).

Os dados foram coletados por meio de formulário aplicado aos integrantes do laboratório, documentos e protocolos produzidos pela equipe de gestão de dados, registros das reuniões de acompanhamento, informações decorrentes da aplicação do projeto-piloto e registros gerados durante a implementação das diretrizes. Para a renomeação dos arquivos se fez uso do Praticidados, um protocolo para padronização da nomeação de arquivos e organização de pastas compartilhadas desenvolvido por meio de uma ferramenta computacional específica.

3 DADOS DE PESQUISA: CONTEXTO

Contemporaneamente, os dados de pesquisa desempenham um papel central no cenário científico, pois não apenas sustentam o desenvolvimento de novas investigações, mas conferem maior credibilidade às pesquisas publicadas. Os pressupostos da Ciência Aberta ressignificaram a prática científica ao incentivar o compartilhamento de materiais, códigos, instrumentos, metodologias e outros recursos essenciais. Nesse contexto, o reuso de dados tornou-se fundamental, permitindo o melhor aproveitamento de tempo e recursos empregados pelos pesquisadores. Além disso, a ampliação do acesso aos dados de pesquisa tem promovido maior transparência no processo científico (Wilkinson *et al.*, 2016).

Para compreender como foi alcançado o cenário atual, é fundamental revisitar os paradigmas científicos. O primeiro paradigma corresponde à ciência empírica que estuda a relação entre fenômenos por meio de experimentos. O segundo paradigma refere-se à ciência teórica, que desenvolve modelos para descrever e explicar fenômenos naturais. Já o terceiro paradigma se baseia na simulação computacional, utilizando modelos matemáticos e algoritmos numéricos para testar hipóteses científicas. Esse avanço levou ao surgimento do *eScience*, também conhecido como quarto paradigma científico (Hey *et al.*, 2009), que marca a transição para a chamada Era do Big Data Científico. Esse quarto paradigma se caracteriza pelo uso de novas metodologias e tecnologias para coletar, manipular, analisar e visualizar grandes volumes de dados, ampliando as fronteiras do conhecimento. Mesmo no Brasil, o termo e-

Science ou e-Ciência é amplamente reconhecido e, por isso, adotado neste texto. O termo *e-Science* passou a ser empregado no final da década de 1990 e início dos anos 2000 para caracterizar formas de investigação científica apoiadas por infraestruturas computacionais avançadas, capazes de favorecer a colaboração e o processamento e compartilhamento de grandes volumes de dados. Nas práticas científicas contemporâneas, essa perspectiva está estreitamente relacionada à gestão, ao compartilhamento e à reutilização de dados de pesquisa (Rodrigues *et al.*, 2025).

As tecnologias e ferramentas atuais têm ampliado o acesso à captura, curadoria e análise de dados de pesquisa, tornando esse processo mais acessível para além das grandes instituições científicas. Esse avanço possibilita que grupos de pesquisa desenvolvam PGD abrangentes, contemplando desde a coleta até a disponibilização dos dados em repositórios de acesso público.

Uma parcela significativa do trabalho científico permanece invisível, tanto para a sociedade, em termos de benefícios e impacto de qualidade de vida, quanto para a própria comunidade científica. Conforme Ferguson *et al.* (2014, p. 1443) “estudos recentes indicam que mais de 50% das descobertas científicas não aparecem na literatura publicada, ao invés disso, residem nas gavetas e nos computadores pessoais dos pesquisadores”. Esse cenário evidencia a necessidade de políticas e práticas que incentivem a divulgação e o compartilhamento de dados científicos.

Esse fenômeno se torna ainda mais evidente no segmento conhecido como “cauda longa da ciência”, no qual muitas pequenas equipes de pesquisadores e laboratórios independentes geram, em seu dia a dia, uma ampla diversidade de coleções de dados. Embora individualmente modestas, essas pequenas coleções estão sendo cada vez mais reconhecidas como ativos informacionais de alto valor, pois, quando consideradas em conjunto, podem ter um impacto significativo, superando a mera soma de suas partes (Wyborn; Lehnert, 2016). Os dados da cauda longa representam a maior parcela do volume de dados científicos produzidos e constituem um espaço de constante criatividade e inovação, que precisa ser revelado, integrado e compartilhado. Além disso, os dados são fundamentais para garantir a veracidade dos resultados das pesquisas e dos artigos científicos publicados, servindo como alicerce para o avanço da ciência (Molloy, 2011; Borgman, 2015).

Os dados de pesquisas podem ser definidos como “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (OECD, 2020, p.13). O conceito de dados de pesquisa é amplo, pois abrange informações provenientes de diferentes processos e fontes. Esses dados podem assumir diversas formas, como números, imagens, textos, vídeos, áudio, *software*, algoritmos, equações, animações, modelos e simulações. Dada essa diversidade, o armazenamento e a curadoria das coleções de dados devem refletir essa abrangência, garantindo sua preservação e acessibilidade (Sayão; Sales, 2014).

Os dados de pesquisa podem ser diferenciados de acordo com a maneira como são obtidos ou produzidos. Entre as classificações utilizadas na literatura estão os dados observacionais, computacionais e experimentais, distinção que permanece presente em estudos recentes sobre gestão de dados (Thanos *et al.*, 2023). Outras categorias também vêm sendo incorporadas, como dados derivados, compilados e de referência, acompanhando a diversidade crescente das práticas de pesquisa.

Os dados observacionais resultam do registro de fenômenos, eventos ou comportamentos em determinado contexto. Podem incluir, por exemplo, observações de campo, entrevistas, questionários, imagens e medições realizadas em situações específicas. Como muitos desses registros estão vinculados a condições particulares de tempo e espaço, sua obtenção nem sempre pode ser repetida. Por essa razão, sua preservação e adequada

documentação são especialmente importantes, sobretudo para que o contexto em que foram produzidos possa ser compreendido posteriormente (Thanos *et al.*, 2023).

Os dados computacionais são gerados por modelos, algoritmos, simulações ou outros procedimentos realizados em ambiente computacional. A possibilidade de reproduzi-los está relacionada à disponibilidade das informações necessárias para refazer o processo que lhes deu origem. Isso envolve documentar não apenas os resultados, mas também os dados de entrada, códigos, parâmetros, softwares e demais componentes relevantes do ambiente computacional. Nesse caso, a gestão dos dados está diretamente relacionada à preservação das condições que possibilitam a reprodução do processo de pesquisa (Thanos *et al.*, 2023).

Já os dados experimentais são produzidos em experimentos conduzidos sob condições definidas, geralmente com o emprego de instrumentos, equipamentos, materiais e procedimentos específicos. Embora muitos experimentos possam ser repetidos, isso não significa que os dados possam ser reproduzidos de maneira idêntica. Diferenças nas condições experimentais, nos equipamentos ou nos próprios procedimentos podem interferir nos resultados. A documentação dos métodos, instrumentos e circunstâncias em que os dados foram produzidos torna-se, portanto, parte importante de sua gestão. Estudos recentes sobre dados experimentais também destacam que uma documentação adequada favorece a transparência, a reprodutibilidade e a reutilização dos dados (Thanos *et al.*, 2023; Cunha-Oliveira; Ioannidis; Oliveira, 2024).

Essa classificação não deve ser entendida como rígida. As práticas contemporâneas de pesquisa produzem conjuntos de dados cada vez mais diversificados e, muitas vezes, combinam diferentes formas de obtenção e processamento. Por isso, além da origem dos dados, seu planejamento e sua gestão precisam considerar as condições de produção, as possibilidades de reprodução, os metadados necessários e o potencial de reutilização e preservação ao longo do tempo.

Borgman (2015) introduz os conceitos de *big data* e *little data* fazendo uma analogia com os termos *little science* e *big science*, propostos por Price (1986). Price não diferenciou os termos com base no tamanho dos projetos, mas sim na maturidade da ciência. Enquanto *big science* é caracterizada como internacional e colaborativa, a *little science* envolve métodos e dados heterogêneos, com análises e controle local (Reichmann *et al.*, 2021).

Segundo Borgman (2015), a distinção entre *big data* e *little data* não se baseia apenas no volume de informações, mas no que pode ser feito com esses dados. Características como volume, variedade e velocidade, ou uma combinação desses fatores, influenciam essa classificação. O avanço da tecnologia móvel, como os celulares, aumentou exponencialmente o fluxo de dados rastreáveis, criando modelos ricos de atividade social. No entanto, a autora ainda enfatiza que o uso desses dados está se proliferando mais rapidamente do que as legislações de privacidade e as políticas de informação conseguem acompanhar.

O PGD requer um entendimento aprofundado sobre a natureza e os processos de criação dos dados de pesquisa. Esse conhecimento é essencial para desenvolver estratégias eficazes de armazenamento e curadoria, garantindo que o sistema seja robusto, confiável e alinhado às melhores práticas de GD.

4 DADOS DE PESQUISAS DO LEPES

O LEPES se constitui em uma rede de especialistas voltada para a transformação social no Brasil, com foco na produção de conhecimento aplicado para subsidiar políticas públicas orientadas ao desenvolvimento de crianças e jovens. Suas principais áreas de atuação abrangem educação, violência e saúde mental.

Dentre os princípios norteadores do LEPES, tem-se a ampliação do papel da escola na sociedade contemporânea, exigindo sua preparação para lidar com questões como

desenvolvimento socioemocional, inclusão, diversidade e bem-estar estudantil. Além disso, enfatiza-se a formulação de estratégias pedagógicas que promovam o engajamento dos estudantes e a conciliação da diversidade e desigualdade entre os indivíduos no ambiente escolar.

O Laboratório atua em diversas frentes, dentre as quais se destacam:

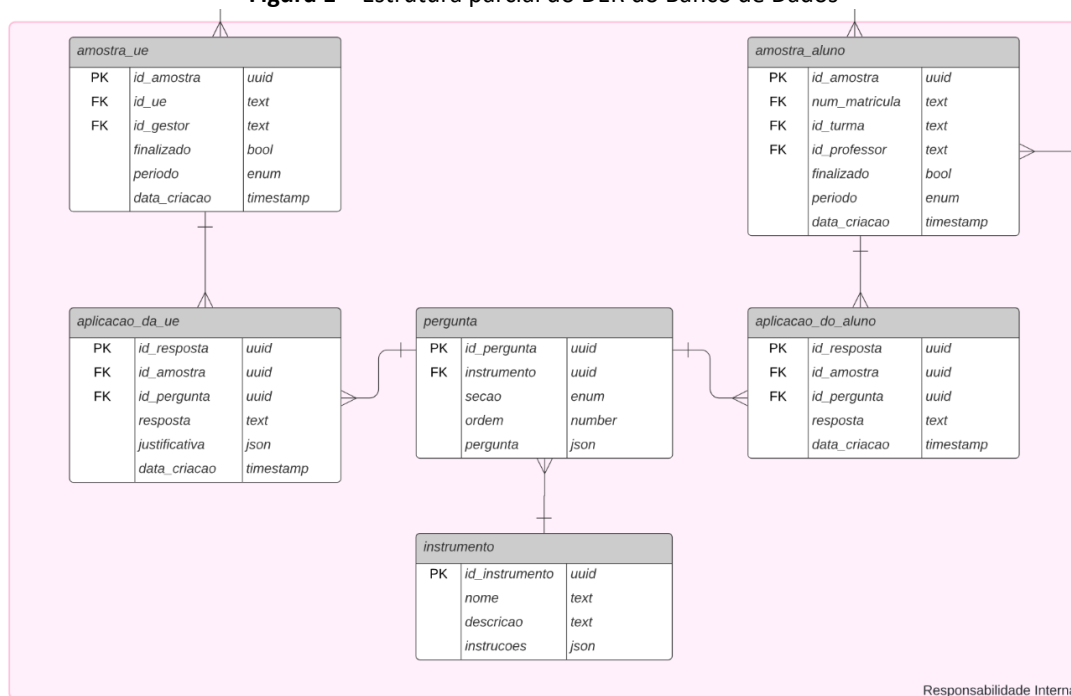
- **Organização Institucional de Sistemas de Ensino:** Investiga mecanismos para otimizar recursos e melhorar a qualidade da educação, com ênfase na governança educacional e na valorização dos profissionais da docência. Nesse contexto, destacam-se iniciativas como o Observatório da Qualidade da Educação Infantil, que monitora e avalia redes de ensino, e o Instrumento das Aprendizagens na Primeira Infância, voltado para a avaliação de desigualdades e acompanhamento da aprendizagem na pré-escola.
- **Saúde Mental e Prevenção às Vulnerabilidades:** Pesquisa fatores de risco associados à violência e às fragilidades relacionadas à saúde mental de crianças e jovens, buscando estratégias para enfrentamento desses desafios no ambiente escolar e familiar.
- **Desenvolvimento Infantil, Educação Infantil e Alfabetização:** Trabalha na elaboração de instrumentos avaliativos alinhados à Base Nacional Comum Curricular (BNCC), visando a qualificação da gestão e da formação continuada dos profissionais de Educação Infantil, bem como a redução das desigualdades na primeira infância.
- **Desenvolvimento Integral e Transição Escola-Trabalho:** Propõe um modelo de formação integral que vá além do ensino tradicional, promovendo um desenvolvimento holístico dos estudantes e preparando-os para a transição para a vida adulta. Um exemplo dessa abordagem é o programa "Eu Posso Te Ouvir", uma iniciativa intersetorial que busca oferecer suporte psicossocial a estudantes em situação de vulnerabilidade, promovendo acolhimento e reorganização do ambiente escolar para garantir proteção social.

O LEPES, por meio dessas iniciativas, busca fomentar políticas públicas e práticas educacionais inovadoras que contribuam para a melhoria da qualidade da educação e o desenvolvimento integral de crianças e jovens no Brasil.

Em consonância com a diversidade de temas abordados pelas pesquisas, há uma ampla variedade de dados gerados, tanto em formato quanto em estruturação e conteúdo. Iniciativas como o Observatório da Qualidade da Educação Infantil e o Instrumento das Aprendizagens na Primeira Infância contam com suporte de plataformas digitais, possibilitando a produção de dados estruturados, armazenados em bancos de dados relacional.

A Figura 1 apresenta um recorte do Diagrama de Entidade-Relacionamento (DER) de uma das bases de dados. Em fevereiro de 2025, o banco de dados do Observatório da Qualidade da Educação Infantil contava com 41 tabelas, responsáveis pelo armazenamento de um total de 221.640 registros.

Figura 1 – Estrutura parcial do DER do Banco de Dados



Fonte: os autores (2025)

Por outro lado, projetos desenvolvidos no eixo Saúde Mental e Prevenção às Vulnerabilidades geram, por exemplo, dados não estruturados, armazenados em diversos formatos de arquivo. Além disso, o Laboratório também trabalha com dados provenientes de fontes externas, como o Censo Escolar do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), a Pesquisa Nacional por Amostra de Domicílios (PNAD) e outras bases do Instituto Brasileiro de Geografia e Estatística (IBGE).

Os *datasets* que contêm esses dados podem estar armazenados em formatos como XLSX, CSV ou JSON. No entanto, é comum que esses arquivos sejam duplicados em diferentes pastas das diversas equipes do laboratório. Diante desse cenário, torna-se essencial adotar práticas organizacionais eficazes para evitar redundâncias, otimizar o uso do espaço de armazenamento e garantir a identificação correta das diferentes versões dos arquivos.

Além disso, grande parte desses dados é classificada como sensível conforme a LGPD, o que exige um planejamento minucioso para sua gestão e segurança. Nesse sentido, destaca-se a necessidade de um plano de gestão de dados abrangente, que estabeleça diretrizes claras para a coleta, armazenamento, processamento e compartilhamento das informações. Esse plano deve assegurar a conformidade com as normativas legais vigentes, além de promover a ética na pesquisa e contribuir para a elaboração de políticas públicas baseadas em evidências.

5 PREMISSAS PARA O PROJETO DE DIRETRIZES PARA GESTÃO DE DADOS DE PESQUISA

Para que os dados de pesquisas sejam verdadeiramente úteis e passíveis de reuso, é fundamental estabelecer uma estrutura clara e consistente. Essa estrutura deve abranger a padronização dos metadados, metadados detalhados e documentação completa, além de contar com uma infraestrutura confiável, que confira a segurança e a privacidade dos dados. Também são essenciais a curadoria de dados, o incentivo à cultura de compartilhamento e a avaliação e aprimoramento contínuo. A Figura 2 ilustra essa estrutura.

Figura 2 – Estrutura de um projeto de gestão de dados de pesquisa

Fonte: os autores (2025)

A organização cuidadosa e a padronização dos metadados são essenciais para facilitar a busca, a recuperação e o uso eficiente dos dados. O desenvolvimento de ontologias específicas ou de vocabulários controlados contribui para a categorização e uniformização dos dados, assegurando que estejam alinhados às necessidades da área de pesquisa (W3C, 2014).

Para garantir que os dados de pesquisa possam ser reaproveitados e agreguem valor ao conhecimento científico, é fundamental que cada conjunto de dados contenha metadados detalhados e documentação completa. Esses elementos não apenas descrevem o contexto de coleta e os métodos empregados, mas especificam autoria, versões e outras informações relevantes, promovendo a transparência e a reprodutibilidade (Tenopir *et al.*, 2011). Dessa forma, os dados podem ser consultados, validados e utilizados por outros pesquisadores, servindo uma base sólida para novas investigações.

Devido ao desenvolvimento simultâneo e compartilhado de diversos projetos, o LEPES identificou a necessidade de elaborar um protocolo que estabelecesse requisitos essenciais, como o contexto da coleta, métodos utilizados, as autorias, versões e outras informações específicas de cada projeto. Esses elementos complementam os metadados e contribuem para a documentação detalhada dos projetos, garantindo maior organização e rastreabilidade dos dados.

Outro aspecto fundamental para a gestão eficaz de dados é o desenvolvimento de uma infraestrutura confiável que assegure tanto o armazenamento quanto a disponibilização dos dados. Nesse contexto, repositórios institucionais e disciplinares desempenham um papel essencial, pois garantem a acessibilidade e o compartilhamento seguro dos dados. A adoção de repositórios alinhados às diretrizes FAIR permite que os dados sejam gerenciados de forma

a maximizar seu valor científico, facilitando seu reuso em diferentes estudos e contextos de pesquisa (Sayão; Sales, 2014; Wilkinson *et al.*, 2016).

A segurança e a privacidade dos dados devem ser prioridade em qualquer projeto de GD, especialmente quando envolvem dados sensíveis. A adoção de práticas como criptografia e controles de acesso, garante a integridade dos dados e a proteção da privacidade dos participantes, em conformidade com a LGPD. Além disso, o monitoramento contínuo dos sistemas de segurança é fundamental para prevenir vazamentos e acessos não autorizados, reforçando a confiança na infraestrutura de dados (Freund *et al.*, 2023).

Além disso, a curadoria de dados deve ser um processo contínuo, envolvendo verificações regulares para corrigir erros e assegurar que os dados sejam precisos e estejam sempre atualizados. Essa prática reforça a confiabilidade dos dados e amplia sua utilidade para a comunidade científica, especialmente em estudos de longo prazo (Sayão; Sales, 2012).

Incentivar a cultura de compartilhamento é outra dimensão importante da GD, devendo incluir estratégias que motivem os pesquisadores a disponibilizarem seus dados de forma aberta sempre que possível. Políticas institucionais de reconhecimento e recompensa podem estimular a contribuição de dados de alta qualidade, fortalecendo o ambiente colaborativo da pesquisa científica. Monteiro e Sant'Ana (2024) destacam anomalias relacionadas a essa cultura do compartilhamento de dados. Uma dessas anomalias ocorre quando os pesquisadores enfrentam dificuldades devido “à confusão ou ao pouco entendimento sobre as normas institucionais, morais e sociais, a inobservância das orientações institucionais, as disparidades entre as metas e os caminhos para atingi-las e uma pequena competência para a gestão dos dados” (Monteiro; Sant'Ana, 2024, p. web).

Por fim, um PGD bem-sucedido deve incorporar práticas de monitoramento e avaliação para mensurar o impacto dos dados compartilhados. Além disso, o planejamento do repositório deve incluir a implementação de ferramentas que permitam a rastreabilidade de citações e *downloads*.

O Quadro 1 sintetiza as premissas adotadas para o projeto.

Quadro 1 - Premissas adotadas para o projeto

Dimensão	Descrição
Padronização e organização dos metadados	Estruturação dos metadados para facilitar a busca, recuperação e uso eficiente dos dados. Inclui ontologias e vocabulários controlados.
Documentação e Transparência	Garantia de que os conjuntos de dados contenham metadados detalhados e documentação completa, especificando contexto, métodos, autoria e versões.
Infraestrutura confiável	Desenvolvimento de repositórios institucionais e disciplinares para armazenamento seguro e acessível dos dados, alinhados às diretrizes FAIR.
Segurança e privacidade	Aplicação de criptografia, controles de acesso e monitoramento contínuo para proteção dos dados, em conformidade com a LGPD.
Curadoria de dados	Processo contínuo de verificação, correção de erros e atualização dos dados, garantindo confiabilidade e usabilidade em longo prazo.
Cultura de compartilhamento	Estratégias para incentivar pesquisadores a compartilhar dados, incluindo políticas de reconhecimento e recompensa.
Avaliação e aprimoramento contínuo	Monitoramento do impacto dos dados compartilhados, rastreabilidade de citações e <i>downloads</i> para melhoria das práticas de gestão.

Fonte: os autores (2025)

As premissas apresentadas no Quadro 1 são essenciais para compreender como os dados estão sendo utilizados, permitindo a coleta de informações que podem melhorar futuras práticas de gestão e maximizar o valor científico dos dados disponibilizados.

5 PROJETO E IMPLANTAÇÃO DAS DIRETRIZES

5.1 Caracterização da Equipe Envolvida

A equipe de gestão de dados era composta por nove membros: sete estagiários (seis do curso de Economia e um do curso de Biblioteconomia e Ciência da Informação) e duas lideranças, ambas alunas de mestrado em Economia. Devido ao aumento das demandas, em maio de 2021, a equipe foi ampliada, passando a contar com mais dois integrantes: uma liderança com formação em Psicologia e atuação em Psicometria, além de um novo estagiário de Economia.

A equipe era responsável pelo tratamento das bases de dados das pesquisas do Laboratório, desenvolvimento e normatização de instrumentos e elaboração de protocolos de organização. Com o volume crescente de trabalho e a ausência de um padrão estruturado para a organização dos dados, a necessidade de um PGD tornou-se não apenas relevante, mas urgente.

É importante ressaltar que o projeto do Plano de Gestão de Dados do Laboratório teve a supervisão e orientação direta de três Docentes, sendo dois da área de Ciência da Informação e um da área de Administração com ênfase em Tecnologia e Sistemas de Informação.

5.2 Levantamento dos Requisitos e Implantação de um Piloto

Como estratégia inicial, optou-se por elaborar e aplicar um formulário interno, com o objetivo de mapear o comportamento e as percepções dos integrantes do Laboratório em relação à GD. Os dados obtidos permitiram a construção de um perfil detalhado dos usuários, considerando seus hábitos e preferências de trabalho, o que forneceu uma base sólida para o desenvolvimento das etapas subsequentes.

A partir dos resultados, iniciou-se o desenvolvimento das estruturas de pastas e do protocolo de nomeação de arquivos. O ponto de partida foi a proposição de modelos de organização de pastas, alinhados às preferências identificadas no questionário. Durante esse processo, houve um diálogo constante com todas as equipes, garantindo que as soluções adotadas atendessem às suas demandas específicas, promovendo uma escuta ativa dos usuários. Todas as questões levantadas eram discutidas nas reuniões semanais da equipe de dados, o que permitiu ajustes e aprimoramentos contínuos.

Durante a elaboração dos protocolos de organização, houve uma troca constante de informações entre os membros de todas as equipes do laboratório. Como parte desse processo, um modelo piloto foi aplicado na unidade de Sobral, no Estado do Ceará. As reuniões com a equipe de Sobral aconteciam semanalmente e as alterações no modelo eram feitas ao longo da semana. As atualizações eram repassadas aos integrantes que acompanhavam a aplicação piloto do projeto, que testavam o modelo e, na semana seguinte, compartilhavam suas impressões, possibilitando ajustes e refinamentos contínuos.

O resultado desse processo foi a criação do protocolo de nomeação de arquivos e do protocolo de organização de pastas. Para garantir a padronização do trabalho dos membros do Laboratório, foi elaborado um documento detalhado, descrevendo os procedimentos corretos para documentar, armazenar e organizar os arquivos. Esse documento serve como referência

para orientar ações futuras e foi publicado *online* no RPubS lepes_dados, perfil que reúne publicações do laboratório relacionadas à gestão de dados.

5.3 Protocolo de Nomeação de Arquivos

A adoção de um modelo único de nomenclatura facilita a recuperação de dados entre as equipes e o acesso geral aos arquivos. Além de evitar o desaparecimento de arquivos, o protocolo visa resolver problemas relacionados ao versionamento, minimizando as dificuldades encontradas pelos usuários durante os processos de busca (Sayão *et al.*, 2015).

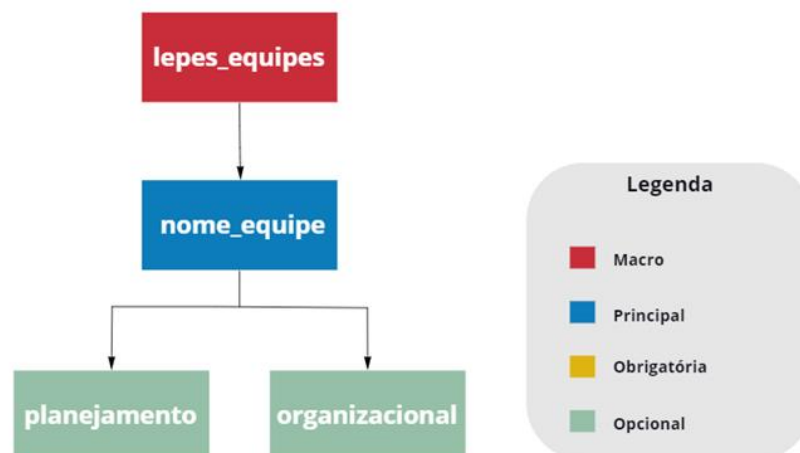
As regras gerais para a padronização da nomeação dos arquivos elaboradas pela equipe foram:

- Não utilizar acentuação, "ç" (cedilha), símbolos (ª ou º), ou caracteres especiais/pontuação (como ^~*!#%\$&);
- Não utilizar espaços em branco ou hífen (-). Os campos devem ser separados por *underscores* (_) sem espaço de digitação;
- Utilizar apenas letras minúsculas;
- Evitar preposições, conjunções e artigos (como o, a, e, de, do, da...);
- Os nomes dos arquivos devem facilitar a sua busca, sendo orientados pelo conteúdo. Exemplos: "relatório_de_..." ou "apresentação_X_proposta_Y";
- A data da última edição do arquivo será utilizada como versão, seguindo o padrão ISO 8601-2:2019, no formato ano-mês-dia (AAAAMMDD);
- O versionamento pode ser inserido antes ou depois do nome do arquivo. Em um mesmo projeto, deve ser consistente, ou seja, sempre antes ou sempre depois do nome.

A equipe de gestão de dados também definiu um protocolo de organização de pastas com base nas respostas ao questionário dos membros. O modelo proposto utiliza o Google Drive como plataforma de trabalho, devido à familiaridade dos membros com a ferramenta. Isso facilita a adaptação dos usuários. Foi elaborada uma estrutura composta por seis pastas principais.

A pasta "Equipes" contém arquivos pertinentes a cada equipe, como cronogramas, fluxos de trabalho e outros documentos relativos à organização interna da equipe. A Figura 3 apresenta a estruturação proposta a pasta Equipes.

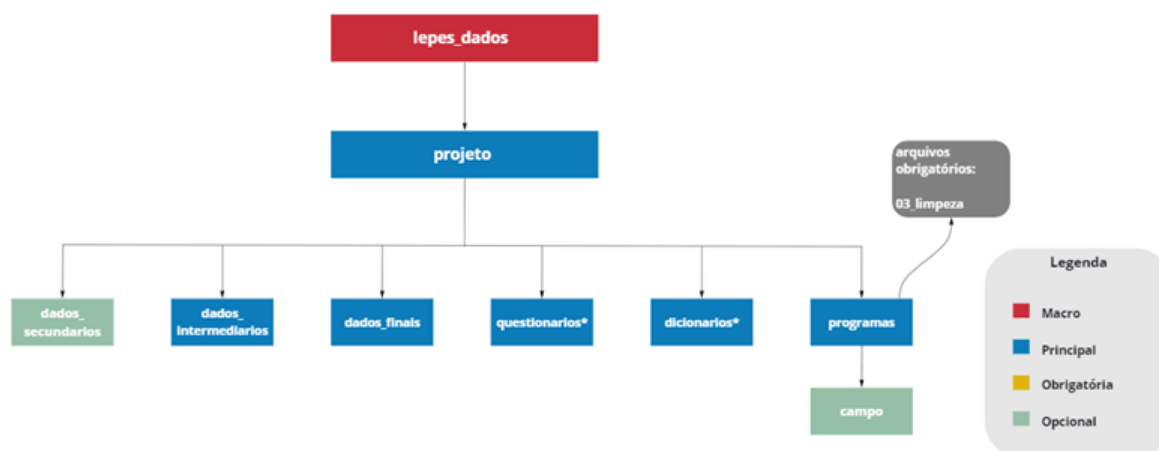
Figura 3 - Estrutura pasta Equipes



Fonte: os autores (2025)

A pasta “Dados” visa armazenar arquivos relacionados aos dados já anonimizados de cada projeto (Figura 4).

Figura 4 - Estrutura pasta dados



Fonte: os autores (2025)

Dentro da estrutura da pasta "Dados", foram definidas algumas subpastas, cada uma com uma função específica para garantir uma organização eficiente e adequada dos dados. O Quadro 2 apresenta a descrição de cada pasta proposta.

Quadro 2 - Descrição das pastas propostas

Pasta	Descrição
Dados Secundários	Armazena dados externos ao projeto, como Censo ou Censo Escolar. Evita duplicações referenciando arquivos existentes no LEPEs.
Dados Intermediários	Contém dados em processo de finalização, permitindo checagens ou ajustes futuros.
Dados Finais	Armazena apenas dados tratados, limpos e anonimizados, prontos para análise e uso no projeto.
Questionários	Guarda os questionários utilizados na pesquisa, geralmente em formatos .doc ou .pdf.
Dicionários	Contém dicionários de dados finais, detalhando variáveis e campos para melhor compreensão.
Programas	Armazena scripts utilizados no processamento dos dados após a anonimização, organizados por ordem de execução.
Campo	Guarda programas específicos utilizados na coleta de dados, sem conter informações pessoais.

Fonte: os autores (2025)

A pasta “Dados Pessoais” contém os arquivos com dados ainda não anonimizados do projeto. A estrutura de subpastas proposta é apresentada no Quadro 3.

Quadro 3 - Estrutura de subpastas proposta

Pasta	Descrição
Resultado Sorteio	Contém o resultado dos sorteios da amostra, incluindo dados pessoais.
Dados Secundários	Pasta opcional para armazenar dados externos utilizados no projeto, como dados do Censo ou Censo Escolar.
Dados Brutos	Armazena dados originais da coleta, sem alteração, geralmente contendo

	informações pessoais.
Programas	Contém programas utilizados até a anonimização, organizados conforme o processo de tratamento dos dados.
Campo	Armazena programas usados na coleta de dados, como scripts de checagem de qualidade.
Macroprojetos	Pasta consolidada para cada projeto, contendo um arquivo LEIA-ME explicando como o projeto se relaciona com outras agendas do LEPES.
Apresentações	Armazena as versões finais das apresentações realizadas no projeto.
Documentos Jurídicos	Guarda materiais burocráticos, como contratos, termos de consentimento e acordos de cooperação.
Planejamento	Contém conteúdo orientadores do projeto, como cronogramas, orçamentos e atas de reuniões.
Instrumentos	Armazena os instrumentos de pesquisa e informações sobre sua construção e tradução, como questionários.
Insumos	Pasta para elementos consumidos no projeto, como materiais literários utilizados.
Resultados	Contém todos os resultados obtidos no projeto, organizados em imagens, tabelas, textos e programas.

Fonte: os autores (2025)

Na pasta “Projetos” são guardados os arquivos do trabalho consolidado, com uma pasta dedicada a cada projeto específico. É recomendável que cada pasta de projeto contenha um arquivo LEIA-ME, elaborado no formato .txt, para explicar como o projeto se relaciona com as demais agendas do LEPES. Dentro dessa pasta, estão organizadas as subpastas descritas conforme o Quadro 4.

Quadro 4 - Descrição das subpastas da pasta Projetos

Pasta	Descrição
Apresentações	Armazena as versões finais de todas as apresentações realizadas no projeto.
Acompanhamento Campo	Contém arquivos relacionados ao acompanhamento do campo de pesquisa. Caso haja necessidade de compartilhamento externo, deve-se utilizar a pasta "Acesso Externo".
Documentos Jurídicos	Destinada a materiais burocráticos, como contratos, documentos de comitês de ética, acordos de cooperação, termos de responsabilidade e consentimento. Pode incluir subpastas para Comitê de Ética e Pesquisa (CEP), acordos e termos, plano de trabalho e outros conforme necessário.
Planejamento	Contém materiais essenciais para o planejamento do projeto, como <i>outlines</i> , cronogramas, orçamentos e atas de reuniões.
Instrumentos	Armazena instrumentos utilizados no projeto, como versões originais e traduzidas, artigos de validação, questionários e demais materiais qualitativos ou quantitativos. Deve conter subpastas para versões intermediárias e finais.
Insumos	Contém materiais de referência utilizados no desenvolvimento do projeto, como artigos e literatura. Não devem ser armazenados arquivos intermediários, apenas aqueles efetivamente utilizados.
Resultados	Reúne todos os resultados do projeto, organizados em subpastas para imagens, tabelas, textos e programas. A pasta "Programas" não precisa de subdivisões, mas as demais devem incluir pastas intermediárias e finais para melhor organização.

Fonte: os autores (2025)

A pasta “Utilitários” é proposta para conter os modelos de apresentações, textos e outras informações úteis para todos os membros do LEPEs. As subpastas estão descritas no Quadro 5.

Quadro 5 - Descrição das subpastas da pasta Utilitários

Pasta	Descrição
Protocolos	Armazena todos os protocolos vigentes no LEPEs.
Cursos	Contém vídeos e materiais dos cursos realizados e oferecidos pelo LEPEs.
Templates	Armazena modelos de slides e apresentações, manual de comunicação visual, além de outros recursos gráficos e orientações.
Arquivos da Administração e RH	Contém estruturas de contratos e relatórios disponibilizados pelo setor administrativo. Caso o modelo desejado não esteja nesta pasta, é necessário consultar o líder de equipe para mais orientações, pois o arquivo pode não estar disponível para acesso geral.

Fonte: os autores (2025)

Por fim, a pasta “Instrumentos” é recomendada quando o projeto envolva o desenvolvimento ou criação de um instrumento. As subpastas são estão descritas no Quadro 6.

Quadro 6 - Descrição das subpastas da pasta Instrumentos

Pasta	Descrição
Campo Validação	Contém arquivos de acompanhamento do campo realizado para validação do instrumento.
Apresentações	Armazena as versões finais de todas as apresentações feitas no projeto.
Documentos Jurídicos	Destinada a armazenar materiais burocráticos, como contratos, comitês de ética, acordos de cooperação, termos de responsabilidade e documentos jurídicos relacionados à pesquisa. Inclui subpastas como CEP (Comitê de Ética e Pesquisa), acordos e plano de trabalho.
Planejamento	Contém documentos que auxiliam na organização do projeto, como <i>outline</i> , cronogramas, orçamentos e atas de reuniões.
Insumos	Destinada a armazenar materiais de referência utilizados no projeto, como literatura. Não devem ser incluídos arquivos intermediários, apenas aqueles efetivamente utilizados no desenvolvimento do trabalho.
Feedback Avaliação	Pasta obrigatória que contém feedbacks de especialistas recebidos na etapa de validação de conteúdo do instrumento, utilizados para sua melhoria.
Resultados	Contém subpastas como artigos, relatórios, validação, normatização, certificação, treinamento, taxonomia e instrumento, que organizam os resultados das análises feitas na criação do instrumento. Dentro da pasta de validação, por exemplo, os arquivos são organizados em programas, textos, imagens, tabelas e audiovisual (vídeos e áudios) relacionados à validação do instrumento.

Fonte: os autores (2025)

De maneira geral, a nomeação das pastas segue o mesmo padrão adotado para os arquivos, garantindo uniformidade e organização. No entanto, o versionamento não deve ser aplicado às pastas. Além disso, estabelece-se um limite de 10 níveis para pastas e subpastas, a fim de evitar estruturas excessivamente complexas e facilitar a navegação e a recuperação de arquivos.

5.4 Processo de Implementação

Como mencionado, a nova estrutura de organização foi desenvolvida levando em consideração o fluxo de trabalho das equipes. Para garantir uma transição eficiente, a implementação seguiu um processo estruturado em etapas.

O primeiro passo foi a realização de um seminário para todos os integrantes do laboratório, apresentando desde a motivação e as principais justificativas até as diretrizes práticas para a implementação. Após o seminário, uma planilha de agendamentos foi disponibilizada para os líderes de cada equipe, com orientações detalhadas sobre o preenchimento, permitindo que cada um escolhesse uma data para realizar a transferência e a renomeação dos arquivos.

A etapa inicial consistiu na transferência dos arquivos para o novo Drive. O líder de equipe indicava a data desejada e informava o caminho dos arquivos a serem transferidos. Caso contivessem dados sensíveis, os arquivos eram direcionados para uma pasta de acesso restrito.

No mesmo dia, ocorria a renomeação dos arquivos utilizando o aplicativo PRATICIDADOS, que será apresentado a seguir. Esse processo foi planejado para ser realizado presencialmente, permitindo que as equipes esclarecessem dúvidas e compreendessem as diretrizes de nomeação estabelecidas no protocolo. Ao final da renomeação automática, o aplicativo gerava um inventário, contendo uma planilha com o caminho e o nome original dos arquivos, bem como os novos caminhos e nomes.

A terceira etapa consistia na transferência manual dos arquivos não contemplados pelo algoritmo para a nova estrutura de pastas no Drive do laboratório. Essa atividade era agendada e contava com a participação de pelo menos dois membros da equipe atendida e dois membros da equipe de gestão de dados. Nesse momento, os usuários podiam esclarecer dúvidas sobre armazenamento e organização. Quando as dúvidas não surgiam espontaneamente, a equipe de dados compartilhava questionamentos comuns de outras equipes para tornar o processo mais didático.

Após a finalização dessa etapa, o aplicativo PRATICIDADOS era novamente utilizado para atualizar o inventário com os novos caminhos dos arquivos, gerando, assim, um importante artefato na gestão do conhecimento.

Uma recomendação essencial foi evitar o armazenamento de arquivos em locais aleatórios, pois isso poderia dificultar sua recuperação futura. Durante os testes, alguns arquivos inicialmente considerados "inclassificáveis" foram discutidos em conjunto com a equipe para determinar o local mais adequado para seu armazenamento. Caso as equipes se deparassem com essa situação no futuro e não chegassem a um consenso, a orientação era entrar em contato com a equipe de dados para suporte.

5.5 A Ferramenta PRATICIDADOS

A ferramenta computacional PRATICIDADOS foi desenvolvida pela equipe de gestão de dados para automatizar a renomeação dos arquivos conforme os parâmetros estabelecidos no protocolo. Ao concluir o processo, o sistema gerava automaticamente uma planilha dentro da pasta informada, contendo as seguintes informações: diretório original, nome original do arquivo, novo diretório e novo nome do arquivo, conforme ilustrado na Figura 5.

Figura 5 - PRATICIDADOS - tela inicial

Fonte: Dados da pesquisa (2025)

A ferramenta permitiu a implementação rápida e eficaz das diretrizes e do protocolo, uma vez que a quantidade de pastas e arquivos a serem renomeados era muito grande.

6 CONSIDERAÇÕES FINAIS

Para responder à questão de como alinhar o desenvolvimento de um PGD voltado ao campo educacional com as práticas de gestão de dados, apresentou-se um relato de pesquisa que apresentou as etapas da criação do Plano de Gestão de Dados desenvolvido para o LEPES, o que permitiu que o objetivo desta pesquisa fosse alcançado.

A implementação da nova estrutura de organização de dados no LEPES representa um avanço significativo na gestão de dados científicos, garantindo maior transparência, acessibilidade e segurança dos arquivos. Ao padronizar a organização e a nomenclatura dos documentos, o novo modelo facilita a rastreabilidade e reutilização dos dados, contribuindo para a integridade e a reprodutibilidade das pesquisas desenvolvidas no laboratório.

Além disso, a adoção do PRATICIDADOS como ferramenta de apoio na renomeação automatizada dos arquivos reforça a eficiência do processo, reduzindo a possibilidade de erros e otimizando o tempo das equipes envolvidas. A abordagem adotada para a implementação, baseada em treinamentos, suporte contínuo e participação ativa dos pesquisadores, garantiu uma transição estruturada e adaptável às necessidades das equipes.

Dessa forma, a estrutura estabelecida não apenas organiza os dados de forma sistemática, mas se alinha às boas práticas de governança da informação, promovendo um ambiente de pesquisa mais colaborativo e sustentável. A iniciativa fortalece o compromisso do LEPES com a ciência aberta e a gestão responsável dos dados, permitindo que os resultados das pesquisas tenham um impacto mais amplo e contribuam para o avanço do conhecimento científico.

Como limitações da pesquisa, tem-se o desenvolvimento e aplicação de um plano de gestão de dados em apenas um laboratório de pesquisa. Para o desenvolvimento de pesquisas futuras sugere-se a aplicação de tal plano em contextos diversos.

REFERÊNCIAS

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

CRESWELL, John W.; POTH, Cheryl N. **Qualitative inquiry and research design: choosing among five approaches**. 4. ed. Thousand Oaks: SAGE, 2018.

CUNHA-OLIVEIRA, T.; IOANNIDIS, J. P. A.; OLIVEIRA, P. J. Best practices for data management and sharing in experimental biomedical research. **Physiological Reviews**, v. 104, n. 3, p. 1387-1408. Disponível em: <https://journals.physiology.org/doi/full/10.1152/physrev.00043.2023>. Acesso em: 22 ago. 2026.

FERGUSON, Adam; NIELSON, Jessica; CRAGIN, Melissa; BANDROWSKI, Anita; MARTONE, Maryann. Big data from small data: data-sharing in the 'long-tail' of neuroscience. **Nature Neuroscience**, v. 17, n. 11, p. 1442-1447, 2014. Disponível em: <https://doi.org/10.1038/nn.3838>. Acesso em: 12 Dez. 2024.

FREUND, Gislaine Parra; MACEDO, Douglas Dyllon Jeronimo de; FAGUNDES, Priscila Basto. Proteção e privacidade de dados: um modelo para o gerenciamento de evidências. **Em Questão**, v. 29, p. e-128009, 2023. Disponível em: <https://doi.org/10.1590/1808-5245.29.128009>. Acesso em: 20 jan. 2025.

GIL, Antonio Carlos. **Como fazer pesquisa qualitativa**. São Paulo: Atlas, 2022

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. **The Fourth Paradigm**: data-intensive scientific discovery. 2009.

HUGHES, Laura D.; TSUENG, Ginger; DIGIOVANNA, Jack; HORVATH, Thomas D.; RASMUSSEN, Luke V.; SAVIDGE, Tor C.; STOEGER, Thomas; TURKARSLAN, Serdar; WU, Qinglong; WU, Chunlei; SU, Andrew I.; PACHE, Lars. Addressing barriers in FAIR data practices for biomedical data. **Scientific Data**, v. 10, art. 98, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36823198/>. Acesso em: 23 ago. 2026.

LEPES. **Laboratório de Estudos e Pesquisas em Economia Social (LEPES)**. 2023. Disponível em: <http://lepes.fearp.usp.br/>. Acesso em: 02 nov. 2023.

MARLINA, E.; HIDAYANTO, A. N.; PURWANDARI, B. A qualitative exploration of challenges and factors in research data management across Indonesian research institutes. **Journal of Librarianship and Information Science**, v. 57, n. 3, p. 849-861, 2025. Disponível em: <https://doi.org/10.1177/09610006241252662>. Acesso em: 23 ago. 2026.

MOLLOY, Jennifer C. The open knowledge foundation: open data means better science. **PLoS Biology**, v. 9, n. 12, 2011. Disponível em: <http://dx.doi.org/10.1371/journal.pbio.1001195>. Acesso em: 25 set. 2024.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT'ANA, Ricardo César Gonçalves. Cultura e anomia no compartilhamento de dados: uma perspectiva a partir dos pesquisadores. **Transinformação**, v. 36, p. e247095, 2024. Disponível em: <https://www.scielo.br/j/tinf/a/hG8jbBcshXzj7BpTLhHJQdc/> 12 dez. 2024.

OECD. **Enhanced access to publicly funded data for science, technology and innovation**. Paris: OECD Publishing, 2020. Disponível em: <https://doi.org/10.1787/947717bc-en>. Acesso em: 23 ago. 2026

PRICE, Derek S. **Little science, big science... and beyond**. New York: Columbia Press, 1986.

REICHMANN, Stefan et al. Between administration and research: Understanding data management practices in an institutional context. **Journal of the Association for Information Science and Technology**, v. 72, n. 11, p. 1415-1431, 2021. Disponível em: <https://doi.org/10.1002/asi.24492>. Acesso em: 22 ago. 2026.

RODRIGUES, Adriana Alves; Araujo, Joana F. de; Silva, Pedro F. C.; et al. Elaboração, aplicação e avaliação de um plano de gestão de dados: o caso DataPB. **Em Questão**, v. 31, p. e-139676, 2025. Disponível em: <https://www.scielo.br/j/emquestao/a/CbMBMFJZCVr9d3sZ4TvVqyw/?lang=pt>. Acesso em: 23 ago. 2026

SAYÃO, Luis Fernando; SALES, Lilian. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179-191, 2012. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/12224>. Acesso em: 10 set. 2024.

SAYÃO, Luís Fernando; SALES, Lilian. Gestão e curadoria de dados científicos: desafios e iniciativas emergentes. **Data Science Journal**, v. 13, p. 81–86, 2014. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2014-009/>. Acesso em: 10 jan. 2025.

THANOS, C.; MEGHINI, C.; BARTALESI, V. et al. An exploratory approach to data driven knowledge creation. **Journal of Big Data**, v. 10, 29, 2023. Disponível em: <https://doi.org/10.1186/s40537-023-00702-x>. Acesso em: 22 ago. 2026.

TENOPIR, Carol; ALLARD, Suzie; DOUGLASS, Kimberly et al. Data sharing by scientists: practices and perceptions. **PLOS ONE**, v. 6, n. 6, e21101, 2011. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>. Acesso em: 10 jan. 2025.

VAN GEND, Thijmen; ZUIDERWIJK, Anneke. Open research data: A case study into institutional and infrastructural arrangements to stimulate open research data sharing and reuse. **Journal of Librarianship and Information Science**, v. 55, n. 3, p. 782-797, 2023. Disponível em: <https://doi.org/10.1177/0961000622110120>. Acesso em: 22 ago. 2026

WILKINSON, Mark; DUMONTIER, Michel; AALBERSBERG, I. J. et al. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 10 jan. 2025.

WORLD WIDE WEB CONSORTIUM (W3C). **Best practices for publishing linked data**. 2014. Disponível em: <https://www.w3.org/TR/ld-bp/>. Acesso em: 10 jan. 2025.

WYBORN, Leslie; LEHNERT, Kerstin. Exploiting the long tail of scientific data: making small data BIG. In: ERESEARCH AUSTRALASIA CONFERENCE, 10-14 out. 2016, Melbourne, Austrália. **Anais...** Melbourne, Austrália, 2016.

YIN, Robert K. **Case study research and applications: design and methods**. 6. ed. Thousand Oaks: SAGE, 2018.

Recebido em/Received: 07/03/2025 | Aprovado em/Approved: 24/08/2026

Declaração de Autoria e Responsabilidade

Concepção e elaboração do estudo: Ieda Pelógia Martins Damian; Ildeberto Aparecido Rodello; Márcia Regina da Silva; Carla Barreto Dória Pinto

Coleta de dados: Ieda Pelógia Martins Damian; Ildeberto Aparecido Rodello; Márcia Regina da Silva; Carla Barreto Dória Pinto

Análise e interpretação de dados: Ieda Pelógia Martins Damian; Ildeberto Aparecido Rodello; Márcia Regina da Silva; Carla Barreto Dória Pinto

Redação: Ieda Pelógia Martins Damian; Ildeberto Aparecido Rodello; Márcia Regina da Silva

Revisão crítica do manuscrito: Ieda Pelógia Martins Damian; Ildeberto Aparecido Rodello; Márcia Regina da Silva

Declaração de uso da Inteligência Artificial

Durante a preparação deste artigo, a autoria, declara que:

(X) utilizou a seguinte ferramenta de IA ChatGPT para fins de buscas para atualização bibliográfica na fase de correções. Após o uso desta ferramenta, a autoria revisou e editou o conteúdo em conformidade com o método científico e assume total responsabilidade pelo conteúdo da publicação.

() não utilizou ferramenta de IA