

CURADORIA DOS DADOS DE PESQUISA NO REPOSITÓRIO NACIONAL DEPOSITA DADOS

Letícia Guarany Bonetti

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

leticiabonetti@ibict.br

 <https://orcid.org/0000-0002-3012-8465>

Tatyane Guedes Martins da Silva

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

tatyanesilva@ibict.br

 <https://orcid.org/0000-0002-1743-0467>

Caterina Groposo Pavão

Universidade Federal do Rio Grande do Sul (UFRGS), Brasil

caterina@ufrgs.br

 <https://orcid.org/0000-0003-3712-7200>

Samile Andrea de Souza Vanz

Universidade Federal do Rio Grande do Sul (UFRGS), Brasil

samile.vanz@ufrgs.br

 <https://orcid.org/0000-0003-0549-4567>

Rene Faustino Gabriel Junior

Universidade Federal do Rio Grande do Sul (UFRGS), Brasil

rene.gabriel@ufrgs.br

 <https://orcid.org/0000-0003-1021-3360>

Marcel Garcia de Souza

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

marcelSouza@ibict.br

 <https://orcid.org/0000-0003-2255-199X>

Washington Luís Ribeiro de Carvalho Segundo

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

washingtonsegundo@ibict.br

 <https://orcid.org/0000-0003-3635-9384>

DOI: 10.22477/xiv.biredial.379

EJE TEMÁTICO: Datos abiertos

RESUMEN

A curadoria dos dados de pesquisa é um processo essencial para garantir a qualidade e a padronização dos registros no repositório. Entretanto, ela não é uma tarefa fácil devido às complexidades intrínsecas aos dados. O objetivo deste trabalho é, portanto, apresentar os casos desafiadores identificados durante o processo de curadoria no repositório Deposita Dados, apontando as soluções adotadas. Ao longo de 2024, a equipe se depa-rou com quatro casos frequentes que não estavam de acordo com as boas práticas e as políticas do repositório: 1) depósito de conjuntos de dados que continham dados pessoais; 2) depósito de dados não estruturados; 3) preenchimento incorreto dos metadados; 4) depósito de arquivos que não eram dados de pesquisa. As soluções adotadas, após o processo de revisão, foram, respectivamente: 1) anonimização dos dados pelo depositante; 2) estruturação adequada dos dados pelo depositante, seguindo os princípios FAIR; 3) correção dos metadados pela própria equipe de curadoria; 4) recusa do depósito por não se enquadrar na tipologia documental do repositório. Este trabalho contribui para o fortalecimento de boas práticas na gestão de repositórios de dados de pesquisa no Brasil, e serve como base para outras equipes de curadoria.

Palabras-clave: Curadoria de dados. Repositório de dados de pesquisa. Repositório Deposita Dados. Dados abertos.



ABSTRACT

Research data curation is an essential process to ensure the quality and standardization of records within the repository. However, it is not an easy task due to the intrinsic complexities of the data. The objective of this work is, therefore, to present the challenging cases identified during the curation process in the Deposita Dados repository, outlining the solutions adopted. Throughout 2024, the team encountered four frequent cases that were not in accordance with good practices and the repository's policies: 1) deposit of datasets containing personal data; 2) deposit of unstructured data; 3) incorrect filling of metadata; 4) deposit of files that were not research data. The solutions adopted, after the review process, were, respectively: 1) anonymization of the data by the depositor; 2) adequate structuring of the data by the depositor, following the FAIR principles; 3) correction of the metadata by the curation team itself; 4) refusal of the deposit as it did not fit the repository's document typology. This work contributes to strengthening good practices in the management of research data repositories in Brazil, and serves as a basis for other curation teams.

Keywords: Data curation. Research data repository. Deposita Dados repository. Open data.

INTRODUÇÃO

Os dados de pesquisa vêm ganhando, em escala global, um papel de destaque. Eles representam um enorme potencial para solucionar problemas e apoiar tomadas de decisão assertivas, deixando de ser meros subprodutos das atividades científicas para tornarem-se uma fonte primária de novos conhecimentos (Sales et al., 2020). Isso porque, se devidamente geridos, os dados de pesquisa trazem benefícios para a ciência e para a sociedade como a transparência e a reprodutibilidade dos resultados das pesquisas, além da economia de recursos financeiros e humanos, uma vez que os dados podem ser reutilizados em novas investigações, poupando os pesquisadores de coletarem dados que já existem. Mas, para isso, é preciso que eles sejam compartilhados.

O compartilhamento livre dos dados ganha força com o movimento conhecido como Dados Abertos (ou Open Data), um dos pilares da Ciência Aberta. De acordo com Balbino et al. (2020), “[...] os dados abertos são compreendidos como dados disponíveis a qualquer pessoa, seja física ou jurídica, para acessar, utilizar, reutilizar, manipular, compartilhar, assim como gerar novos produtos ou serviços” (p. 154). Os dados de pesquisa podem ser definidos, de acordo com Sayão e Sales (2020), como “[...] todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e aceito como evidência pela comunidade científica e necessário para analisar, validar e produzir resultados de pesquisa” (p. 32).

Apesar dos benefícios atrelados ao compartilhamento livre dos dados de pesquisa, ele não é um fim em si mesmo. É preciso, além de compartilhar, garantir que esses dados sejam devidamente geridos em ambientes seguros e confiáveis, e é nesse cenário que entram os repositórios de dados. A criação e a disseminação dessas infraestruturas tecnológicas, segundo Curty (2019), ganharam um impulso expressivo a partir de diretivas governamentais e institu-



cionais voltadas ao compartilhamento de dados de pesquisa financiados com recursos públicos.

Como exemplo pode-se citar a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que reconhece a gestão adequada dos dados como parte essencial das boas práticas de pesquisa. A FAPESP, inclusive, solicita aos pesquisadores que considerem o uso de repositórios para a descrição e acesso aos dados coletados. Outro exemplo a ser citado é o da Conferência Lusófona de Ciência Aberta (ConfOA) que, a partir de 2025, exigirá a disponibilização dos dados de pesquisa utilizados nos trabalhos submetidos. Para apoiar os pesquisadores, a ConfOA também sugere repositórios confiáveis para depósito.

Os repositórios de dados, de acordo com Sanchez, Vidotti e Vechiato (2017), são aqueles que “[...] buscam organizar, estruturar, permitir acesso, disseminar e preservar todos os dados gerados por meio de pesquisas realizadas em sua maioria por Instituições de Ensino e Pesquisa” (p. 3). Sayão e Sales (2016) explicam que os repositórios de dados trazem vários benefícios, tanto para a ciência como para a sociedade. Dentre eles pode-se citar o aumento do acesso e da visibilidade dos dados de pesquisa depositados; a facilidade de identificação e citação adequada dos autores dos dados; a preservação a longo prazo; a criação de uma memória institucional ou nacional, dentre outros.

Com isso em mente, e para apoiar os pesquisadores brasileiros e fortalecer o movimento de Dados Abertos no Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) desenvolveu um repositório nacional de dados de pesquisa em acesso aberto, o Deposita Dados. Ele foi lançado no XXII Seminário Nacional de Bibliotecas Universitárias (SNBU), em 2023, e tem como objetivo integrar e disseminar conjuntos de dados de pesquisa de pesquisadores brasileiros vinculados a instituições científicas que ainda não possuem seus repositórios de dados de pesquisa e/ou de pesquisadores brasileiros que executaram seus conjuntos de dados por meio de colaboração científica em instituições estrangeiras de ensino e pesquisa (Deposita Dados, 2023). O Deposita Dados é, inclusive, um dos repositórios de dados indicados pela ConfOA para o depósito seguro dos dados.

O serviço foi lançado e divulgado no SNBU de 2023, mas passou por um povoamento significativo em 2024, a partir das iniciativas da Rede Brasileira de Repositórios Digitais (RBRD) para incentivar o compartilhamento de dados de pesquisa e o fortalecimento dos repositórios de dados. Pesquisadores de diferentes áreas e regiões do Brasil iniciaram os depósitos no Deposita Dados, na maioria das vezes com auxílio de bibliotecários da RBRD. Após finalizado o processo de depósito por parte dos pesquisadores, inicia a etapa de curadoria, que é realizada pela equipe gestora do repositório.

De acordo com Resende e Bax (2019), “As funções da curadoria digital de dados são: organizar, descrever, filtrar e analisar os recursos, gerenciando-os e buscando por anomalias e melhorias.” (p. 3). A curadoria agrega valor aos dados e transforma dados armazenados para



uso imediato em dados preservados com um propósito de longo prazo, assegurando que os pesquisadores possam acessá-los e utilizá-los após o fim do estudo. Resende e Bax (2019) reforçam isso ao afirmar que “[...] a missão central da curadoria digital de dados é a de garantir o compartilhamento e a reutilização de dados acumulados nos médio e longo prazos” (p. 3) .

A curadoria alimenta o objetivo de criar uma rede de dados que será preservada a longo prazo de forma escalável e interoperável. Mas, de acordo com Resende e Bax (2019), ela traz diversos desafios por exigir uma combinação de habilidades. Na opinião de Triques, Arakaki e Castro (2020), a curadoria digital apresenta-se como uma prática interdisciplinar abrangente que busca estabelecer diretrizes e um conjunto de ações inter-relacionadas que visa a manutenção do material com valor informacional.

A curadoria não é uma tarefa fácil, especialmente porque dados de pesquisa têm um nível de abstração muito maior que publicações tradicionais como artigos, teses e dissertações. Sem a devida contextualização e descrição, sua interpretação e reutilização tornam-se inviáveis. No entanto, o desafio não se limita aos usuários dos dados: as equipes de curadoria dos repositórios enfrentam diariamente a complexidade de avaliar conjuntos de dados provenientes de diversas áreas do conhecimento. Cabe ao curador atuar como mediador, transformando os dados brutos depositados pelo pesquisador em registros acessíveis e úteis para os usuários.

Durante o segundo semestre de 2024, a equipe de curadoria do Deposita Dados se deparou com alguns casos desafiadores e recorrentes. O objetivo deste trabalho é, então, apresentar esses casos junto com as soluções padronizadas que foram adotadas pela equipe depois de reuniões internas. Justifica-se o estudo pela importância do compartilhamento de experiências com outros profissionais e equipes de curadoria, com a finalidade de mostrar as soluções que foram adotadas nos casos mais frequentes e, muitas vezes, complexos, contribuindo para os estudos na área. Assim será possível fortalecer uma cultura de boas práticas em repositórios de dados no Brasil.

MÉTODOS

O estudo qualitativo e exploratório foi realizado por meio da coleta de informações durante o processo de curadoria dos conjuntos de dados de pesquisa depositados no repositório Deposita Dados no segundo semestre de 2024. Este estudo identifica e analisa os conjuntos de dados em acesso aberto que não cumpriram com os requisitos e boas práticas para um compartilhamento efetivo no Deposita Dados. O estudo minucioso foi realizado por duas profissionais bibliotecárias que atuam na Coordenação de Tratamento, Análise e Disseminação da Informação Científica (CODIC), à frente de repositórios de dados, buscando criar e aperfeiçoar serviços de gestão de dados de pesquisa. A equipe também conta com três pesquisadores da Universidade Federal do Rio Grande do Sul (UFRGS), que fornecem apoio técnico e consultoria



para a curadoria, principalmente nos casos desafiadores. Todos os três atuam como docentes na UFRGS e são bibliotecários.

O Deposita Dados é um repositório novo e, portanto, encontra-se em constante processo de aperfeiçoamento, tanto no que diz respeito à infraestrutura como em relação às questões administrativas e de gestão. Foi a partir do estudo diário dos depósitos realizados no segundo semestre de 2024 que a equipe de curadoria passou a identificar os problemas mais frequentes e propor soluções padronizadas para garantir a uniformidade nos registros e a implementação de boas práticas no Deposita Dados.

O objetivo deste trabalho é apresentar os casos desafiadores identificados durante o processo de curadoria no repositório Deposita Dados, apontando as soluções adotadas para a padronização dos registros. O estudo se deu em três etapas principais: 1) identificação de casos desafiadores frequentes; 2) reuniões internas para definir soluções padronizadas e 3) sistematização dos casos e soluções em uma planilha para controle e análise. A planilha¹ está disponível para consulta no repositório de dados de pesquisa do Ibict, o Aleia. Ao sistematizar essa experiência, espera-se contribuir para o debate teórico-prático sobre boas práticas e curadoria digital em repositórios de dados de pesquisa, auxiliando outros profissionais e equipes gestoras.

A próxima seção apresenta o contexto histórico do repositório e os desafios encontrados. Estes foram categorizados em quatro grupos: 1) depósito de conjuntos de dados que continham dados pessoais; 2) depósito de dados não estruturados; 3) preenchimento incorreto dos metadados; 4) depósito de arquivos que não eram dados de pesquisa. Em seguida, são apresentadas as soluções adotadas e as conclusões.

RESULTADOS

O Deposita Dados, desenvolvido com o software Dataverse, é um repositório nacional de acesso aberto mantido pelo Ibict. Seu foco é ser um serviço para os pesquisadores brasileiros cujas instituições ainda não possuem um repositório de dados. A ideia é que os pesquisadores tenham um repositório nacional e gratuito para depositar, evitando que os dados fiquem armazenados apenas nos computadores pessoais ou em repositórios estrangeiros como o Zenodo.

O Deposita Dados é dividido em seis comunidades, construídas a partir das regiões brasileiras (Norte, Nordeste, Sul, Sudeste e Centro-Oeste), e mais uma comunidade para pesquisadores brasileiros que coletaram seus conjuntos de dados por meio de colaboração científica em instituições estrangeiras de ensino e pesquisa. Na Figura 1 tem-se a página inicial do serviço.

¹ Disponível em: <https://doi.org/10.48472/aleia/TNAXHU>



Como já dito, o povoamento mais intenso do Deposita Dados iniciou-se no segundo semestre de 2024, quando a RBRD deu apoio para que pesquisadores de várias instituições depositassem seus dados. A RBRD é constituída pelas cinco Redes Regionais de Repositórios Digitais: Rede Norte, Rede Sul, Rede Sudeste, Rede Nordeste e Rede Centro-Oeste. As Redes Regionais são independentes entre si e cabe ao Ibict disseminar as boas práticas e as recomendações internacionais.

Figura 1 - Página inicial do Deposita Dados



Fonte: captura de tela do Deposita Dados (2025).

Neste mesmo período, a equipe do repositório iniciou um workshop² on-line e síncrono voltado exclusivamente para o uso do Deposita Dados (Figura 2). O treinamento foi divulgado dentro da RBRD e reuniu interessados de todas as Redes Regionais, principalmente bibliotecários e pesquisadores.

Com duração de 1 hora, o workshop traz pontos importantes como: 1) definição e conceito de repositório de dados; 2) contextualização do Dataverse, que é o software adotado para a estruturação do Deposita Dados; 3) apresentação do fluxo de depósito e do Guia de Usuário e 4) passo a passo para depositar e editar os conjuntos de dados. O workshop é dividido em duas partes: uma teórica, onde é feita a contextualização do Deposita Dados, e uma prática,

² Disponível em: https://www.canva.com/design/DAGFORcMjQs/IcXcO4WTgy5EuU5EeEpZzQ/edit?utm_content=DAGFORcMjQs&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

onde a equipe demonstra como acessar o repositório, criar uma conta, solicitar a permissão para depositar e, por fim, como fazer o depósito.

Figura 2 - Slide de apresentação do workshop “Repositório Deposita Dados”

O slide de apresentação do workshop "Repositório Deposita Dados" apresenta o logo do repositório no canto superior esquerdo. O título principal "Workshop: Repositório Deposita Dados" está em uma fonte grande e preta. Abaixo do título, o texto "Treinamento" e a data "21 de novembro de 2024 às 15h" são exibidos. À direita, há dois círculos com fotos das palestrantes, Leticia Bonetti e Tatyane Silva, acompanhadas por seus nomes e e-mails. No canto inferior esquerdo, há logos do ibict 70 anos, do Ministério da Ciência, Tecnologia e Inovação, e do Governo Federal. No canto inferior direito, há um grande gráfico decorativo composto por formas abstratas em tons de azul, marrom e bege.

Fonte: captura de tela da apresentação no Canva (2025).

O workshop foi criado para auxiliar os pesquisadores na utilização do Deposita Dados, uma vez que o povoamento é feito por autoarquivamento, ou seja, o próprio pesquisador deposita seus dados. Mas, mesmo após o workshop, muitos pesquisadores receberam apoio da RBRD para o depósito, já que essa prática é ainda recente no Brasil e gera diversas dúvidas.

Pensando nisso, a equipe do repositório também elaborou uma série de materiais de capacitação, como o Guia de Usuário³ e o tutorial⁴ em vídeo para depósito. Ambos estão disponíveis publicamente para consulta e são enviados diretamente para os pesquisadores no momento em que eles solicitam a permissão para depositar conjuntos de dados. Esses materiais trazem instruções sobre como nomear os arquivos, como preencher os metadados, como identificar dados pessoais nos arquivos, qual formato de arquivo escolher, dentre outros. Isso evita que alguns erros simples sejam cometidos pelos depositantes.

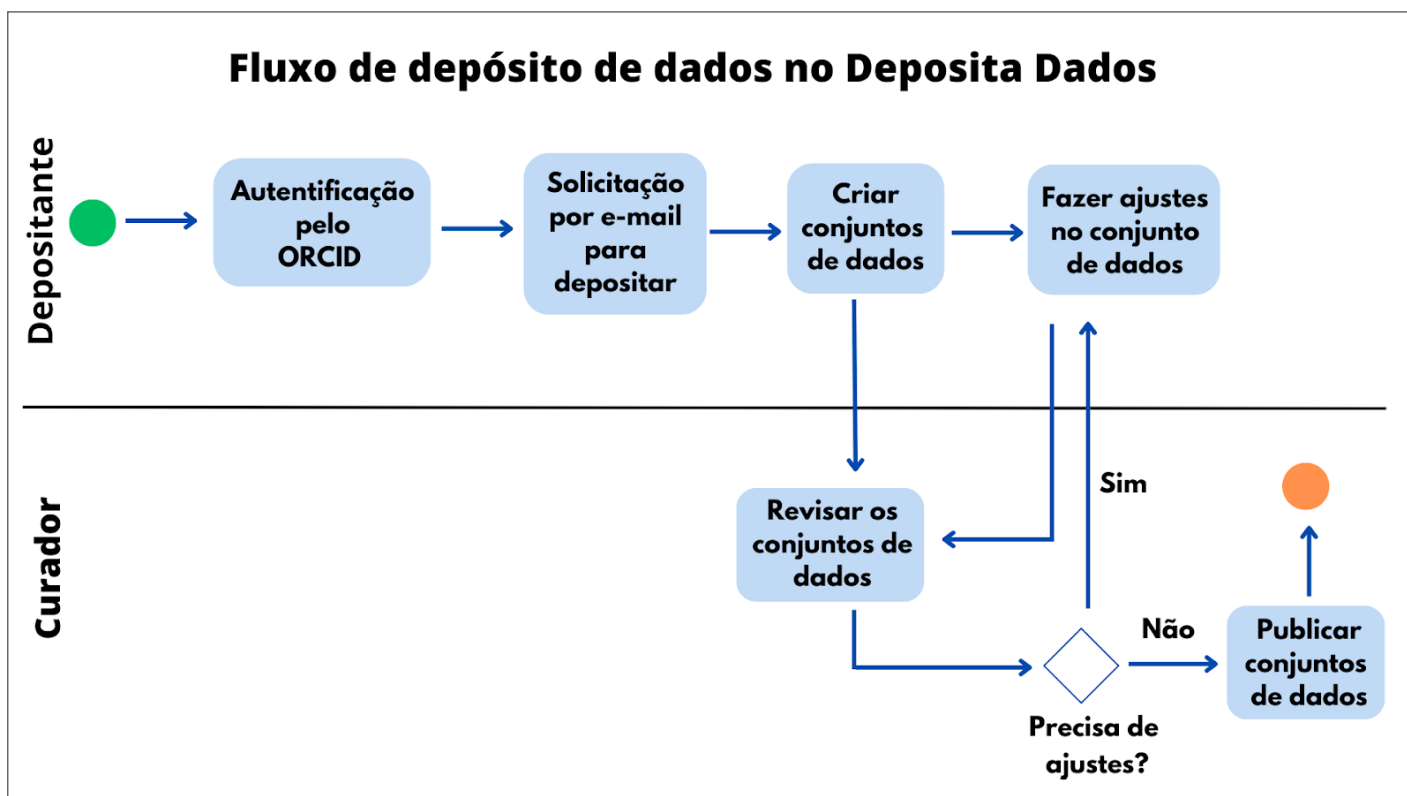
Entretanto, mesmo com as informações disponíveis nos manuais, os usuários ainda podem cometer alguns erros, e por isso nenhum conjunto de dados é publicado automaticamente no Deposita Dados. Todos os dados depositados precisam ser enviados para revisão da

³ Disponível em: <https://depositadados.ibict.br/dvn/guide/guides.html>

⁴ Disponível em: https://www.youtube.com/watch?v=fHHD_obDO2Q&t=22s.

equipe de curadoria, que os avalia individualmente e toma a decisão: 1) publicar, 2) devolver ao autor para correções ou 3) recusar (em casos onde o arquivo submetido não é um dado de pesquisa). Na Figura 3 abaixo é possível visualizar o fluxo de depósito.

Figura 3 - Fluxo de depósito de dados no Deposita Dados



Fonte: elaborado pelos autores (2025).

É na etapa de curadoria que os erros são identificados e corrigidos. Ela é essencial para garantir a qualidade e a padronização dos registros, permitindo que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, conforme estabelecidos pelos princípios FAIR.

No Deposita Dados, a etapa de curadoria tem como principal característica a centralização da revisão dos conjuntos de dados depositados pelos pesquisadores, com o objetivo de verificar se os metadados foram preenchidos corretamente e se os arquivos estão em conformidade com a política do repositório⁵ e a legislação brasileira de proteção de dados.

Durante o estudo realizado no processo de curadoria, etapa que antecede a rejeição, a devolução ou a publicação do conjunto de dados, a equipe do repositório identificou quatro casos desafiadores recorrentes: 1) depósito de conjuntos de dados que continham dados pessoais; 2) depósito de dados não estruturados; 3) preenchimento incorreto dos metadados obrigatórios ou opcionais; e 4) depósito de arquivos que não eram dados de pesquisa, como apresentações de powerpoint, relatórios, teses e artigos acadêmicos. Esses casos serão comentados em mais detalhes a seguir, assim como o procedimento adotado para solucionar o

⁵ Disponível em: <https://depositadados.ibict.br/about.xhtml>

problema identificado.

O primeiro caso diz respeito ao depósito de conjuntos de dados de pesquisa com dados pessoais. No Brasil a Lei nº 13.709/2018, mais conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD), dispõe sobre o tratamento desses dados, inclusive nos meios digitais, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. O Deposita Dados, sendo um repositório multidisciplinar, recebe dados de pesquisa de diversas áreas, incluindo aqueles que podem conter dados pessoais ou passíveis de identificação, ligados principalmente à área da Saúde. Com isso em mente, durante o processo de curadoria, a equipe precisa se atentar aos arquivos depositados, para assim garantir que os dados de pesquisa estejam de acordo com a LGPD.

Como medida preventiva, a equipe envia, por e-mail, um checklist de dados pessoais contendo instruções. Ainda assim, o repositório recebe submissões com dados pessoais e, nesses casos, a equipe de curadoria precisa devolver o conjunto de dados para o depositante fazer as alterações.

O pesquisador recebe então um e-mail explicando que o depósito foi devolvido devido à presença de dados pessoais, que não podem estar disponíveis publicamente no repositório. A equipe de curadoria também sugere soluções, como a supressão das colunas que contêm esses dados nos arquivos. Esse procedimento foi realizado, por exemplo, para um conjunto de dados que continha telefones e e-mails dos participantes da pesquisa. Com a supressão dessas colunas, o problema foi solucionado sem prejuízos à interpretação dos dados.

Após as devidas alterações por parte do depositante, o conjunto de dados é submetido novamente para revisão e, caso esteja tudo certo, ele é publicado no repositório. A sugestão de exclusão dos dados pessoais nos arquivos tornou-se a medida padrão para esses casos por ser uma solução eficaz e simples.

O segundo caso diz respeito ao depósito de dados não estruturados. São aqueles que não estão organizados em uma estrutura predefinida, o que dificulta sua interpretação e uso. É um cenário comum, provavelmente ligado ao contexto ainda recente de compartilhamento de dados de pesquisa no Brasil. Muitos pesquisadores não estão habituados à prática de estruturar dados para depósito.

O Deposita Dados costuma receber tabelas e quadros em arquivos PDF ou JPG, o que dificulta a navegação e a extração de dados. O ideal é que as informações estejam bem organizadas em planilhas, de tal forma que sejam legíveis não só por humanos, mas também por máquinas. Essa questão é importante quando se pensa em boas práticas como os princípios FAIR, que buscam justamente otimizar o processamento automático. Além disso, sem a contextualização adequada das tabelas e dos quadros, não é possível interpretar os dados disponíveis. O mesmo vale para o depósito de gráficos isolados em arquivos Excel, sem a presença dos dados que foram utilizados para gerar os gráficos. Outros pesquisadores não conseguem



interpretar e, conseqüentemente, os dados não são reutilizáveis.

Nesses casos, o conjunto de dados é devolvido para as devidas alterações do autor. No e-mail a equipe explica o motivo e sugere melhorias práticas. No caso dos gráficos, por exemplo, a equipe pede que o pesquisador forneça os dados primários, aqueles que foram utilizados como base para a construção dos gráficos submetidos. Já para as tabelas e quadros em PDF ou JPG, a equipe solicita a organização dos dados disponíveis em uma planilha e a alteração do formato (para CSV ou ODS, por exemplo). Os curadores também podem solicitar a disponibilização de documentos complementares contextualizando os dados depositados (como um arquivo textual readme/leiam).

Para auxiliar o pesquisador nessas questões mais técnicas, no Guia de Usuário existe uma lista de formatos preferidos que ajudam a preservar e garantir o acesso a longo prazo aos arquivos depositados. A recomendação é sempre optar por formatos abertos e não proprietários. Após as devidas correções por parte do pesquisador, o conjunto de dados passa por uma nova revisão da equipe e, caso esteja tudo correto, é publicado.

O terceiro caso frequente é o preenchimento incorreto dos metadados. No Guia de Usuário do Deposita Dados existe uma seção detalhando os campos e a forma de preenchimento correto (Figura 4), uma medida para auxiliar o depositante. Além disso, durante o preenchimento, o usuário pode consultar uma explicação rápida de cada campo no próprio formulário de depósito. Mas, ainda assim, a equipe se depara com casos de preenchimento incorreto.

Figura 4 - Seção para preenchimento dos metadados no Guia de Usuário do repositório

Abaixo serão apresentadas as orientações detalhadas para o preenchimento dos metadados obrigatórios e opcionais.

Metadado obrigatórios	Preenchimento
Título*	Preencha de forma objetiva e sucinta o nome pelo qual o conjunto de dados será conhecido. Somente a letra inicial, substantivos próprios e nomes científicos devem ser grafados em letra maiúscula. Não finalizar com ponto.
Autor*	O nome do autor e a afiliação são preenchidos de forma automática via ORCID, se alguma informação estiver equivocada, é necessário alterar na própria conta do ORCID. No subcampo Esquema de Identificadores, selecione a plataforma e insira o seu número de identificação.
Entre em contato*	Nome: Sobrenome (s), Prenome (sem ponto final) Afiliação: Informe o nome, por extenso, da instituição a qual o (s) autor (es) está (ão) filiado (s) e entre parênteses acrescente a sigla da instituição. Ex.: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Fonte: captura de tela do Guia de Usuário do Deposita Dados (2025).

Até o momento do estudo foram identificadas duas situações recorrentes: 1) dado fornecido no campo errado, ou seja, o pesquisador fornece uma informação correta, mas no metadado incorreto. Um exemplo é indicar a publicação relacionada (um artigo, uma tese, uma dissertação derivada do conjunto de dados) no campo “Notas” ao invés do campo de metadado correto, qual seja, “Publicação relacionada”; e 2) preenchimento incorreto do dado, mas no campo correto. Um exemplo é quando o depositante não coloca o nome dos autores na ordem indicada, ou seja, não faz a entrada pelo sobrenome. Ou quando o usuário coloca o título todo em caixa alta. Nesses casos, por serem correções rápidas dos metadados, a própria equipe gestora faz as alterações necessárias e publica o conjunto de dados. Isso agiliza o processo e garante que os metadados estejam de acordo com o padrão estabelecido pelo repositório.

O quarto caso é o depósito de arquivos que não são dados de pesquisa. Como já mencionado, a prática do compartilhamento de dados ainda é recente no Brasil, e levanta muitas dúvidas nos pesquisadores. Uma delas diz respeito ao que são dados de pesquisa. Isso foi visto principalmente durante os workshops que foram ministrados pela equipe para a RBRD. O Deposita Dados costuma receber depósitos de artigos, relatórios, teses e dissertações, mas



como eles não se enquadram na tipologia documental aceita no repositório, o depósito é recusado.

Os depositantes recebem então um e-mail que explica que os arquivos não se enquadram na tipologia documental aceita no repositório e a equipe sugere que o pesquisador realize um novo depósito, desta vez das fontes primárias que ele utilizou na pesquisa. Ao mesmo tempo são apresentados alguns exemplos para que ele consiga compreender a diferença, mas é uma escolha do pesquisador tentar um novo depósito.

A equipe também indica que o local adequado para o depósito de publicações como artigos, teses e dissertações são os repositórios de publicações. A própria instituição do pesquisador costuma oferecer esse serviço, mas o Ibict também gerencia um repositório de publicações nacional: o Deposita⁶. Caso o pesquisador opte por tentar uma nova submissão, os arquivos são novamente revisados para garantir que se tratam de dados de pesquisa e assim procede-se com a publicação no repositório.

⁶ Disponível em: <http://deposita.ibict.br/>

Figura 5 - Definição de dados de pesquisa e exemplos disponíveis no Guia de Usuário

Dados de pesquisa

Dados de pesquisa 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” (OCDE, 2007, p. 13). Eles podem ser, por exemplo, números, imagens, textos, vídeos, áudio, software, algoritmos, equações, animações, modelos, simulações, variando de acordo com cada área do conhecimento.

Exemplos do que não são dados de pesquisa:

- 1) **Artigos, teses, dissertações, capítulos de livros, livros, anais de congressos:** esses documentos configuram-se como publicações tradicionais. Essas tipologias documentais são os resultados alcançados a partir dos dados coletados na pesquisa. Para depositar esses documentos, consulte se sua instituição possui um repositório de publicações. Caso não, é possível depositar no Repositório Comum do Brasil (Deposita), também gerido pelo Ibict;
- 2) **Apresentações de slides:** as apresentações devem ser depositadas em ambientes para recursos educacionais abertos, como o caso da plataforma Slideshare;
- 3) **Prints de tabelas ou quadros da publicação:** imagens de quadros e tabelas não são documentos pesquisáveis e reutilizáveis. Caso o depositante deseje depositar dados tabulares, deve enviar como uma planilha, em formatos como csv ou ods, além de fornecer uma descrição rica dos valores contidos na planilha.

Fonte: captura de tela do Guia de Usuário do Deposita Dados (2025).

Esse último caso levou a equipe a revisar o Guia de Usuário para atualização. Na nova versão, que já está disponível no site, o guia oferece uma definição de dados de pesquisa juntos com exemplos, conforme mostra a Figura 5. A definição foi retirada de um documento da Organisation for Economic Co-operation and Development (OCDE, 2007).

Além disso, por sugestão dos usuários do repositório, uma nova seção foi inserida para apresentar as licenças Creative Commons. A ideia é que o Guia passe por constantes atualizações para auxiliar os pesquisadores, já que algumas dúvidas só surgem quando o serviço já está disponível para o público.

A curadoria dos conjuntos de dados no repositório é um processo contínuo e, por isso, novos desafios podem surgir ao longo do tempo. Entretanto, o objetivo deste trabalho era



trazer os principais casos desafiadores identificados ao longo do segundo semestre de 2024, junto com as soluções adotadas, contribuindo para os debates na área. As soluções citadas foram tomadas em conjunto dentro da equipe e padronizadas, garantindo a uniformidade dos registros. A ideia é que o Deposita Dados esteja em conformidade com boas práticas nacionais e internacionais, como os princípios FAIR.

CONCLUSÕES

A curadoria dos dados de pesquisa é um processo essencial para garantir a qualidade e a padronização dos registros em um repositório. Entretanto, não é uma tarefa fácil, justamente devido às complexidades intrínsecas aos dados de pesquisa. Em virtude do Deposita Dados ser um repositório multidisciplinar, a equipe de curadoria precisa analisar dados de diferentes áreas do conhecimento e em diferentes formatos, o que amplia ainda mais os desafios.

Ao longo do processo de curadoria em 2024, a equipe se deparou com alguns casos frequentes que levaram à definição de soluções padronizadas em prol de boas práticas no depósito de dados. Os quatro casos foram: 1) depósito de conjuntos de dados que continham dados pessoais; 2) depósito de dados não estruturados; 3) preenchimento incorreto dos metadados; 4) depósito de arquivos que não eram dados de pesquisa.

Após a avaliação das recomendações encontradas na literatura, das boas práticas para compartilhamento de dados e das discussões dentro da equipe, optou-se por adotar as seguintes soluções, respectivamente: 1) retorno do conjunto de dados ao pesquisador com a solicitação da anonimização dos dados; 2) retorno do conjunto de dados ao pesquisador com a solicitação da estruturação adequada dos dados, de tal forma que eles sejam legíveis por humanos e máquinas, como preconiza os princípios FAIR; 3) alteração manual no preenchimento dos metadados pela própria equipe de curadoria, agilizando o processo de publicação do conjunto de dados e 4) rejeição do depósito por não se enquadrar na tipologia documental aceita no repositório. Nesse caso, o pesquisador é instruído a tentar um novo depósito, dessa vez dos dados coletados durante a pesquisa.

Além desses encaminhamentos, sempre realizados por e-mail, a equipe também fornece, no site do repositório, um Guia de Usuário, que traz em detalhes como fazer um depósito e como preencher corretamente os metadados. Esse Guia é enviado para o pesquisador junto com o tutorial de depósito em vídeo, de tal forma que ele tenha contato com o material de capacitação antes de realizar seu depósito.

A equipe também realizou uma série de treinamentos junto com a RBRD, buscando auxiliar o público-alvo do Deposita Dados. Mas sabe-se que, devido ao compartilhamento de dados ser uma prática recente no Brasil, dúvidas ainda vão surgir, assim como erros no depósito.



Por isso a equipe realiza a curadoria de cada conjunto de dados individualmente, garantindo assim que os erros sejam corrigidos antes do conjunto de dados ser publicado no repositório. A curadoria, além de otimizar o benefício que pode ser extraído dos dados, também fortalece a credibilidade do repositório, uma vez que estabelece boas práticas.

Como metas para 2025, a equipe pretende expandir os treinamentos on-line sobre o Deposita Dados, alcançando um público maior, e também aperfeiçoar os materiais de capacitação disponíveis no repositório, como o Guia de Usuário. Conforme o repositório vai sendo utilizado, novas dúvidas surgem, assim como novos desafios, o que leva a uma necessidade frequente de atualização dos materiais. Além disso, pretende-se criar um Guia de curadoria, onde o processo será descrito em detalhes. Isso vai auxiliar tanto os novos integrantes da equipe quanto os usuários, que poderão consultar o documento e entender como o processo é realizado no Deposita Dados. Ao mesmo tempo, os autores pretendem realizar estudos sobre os depósitos propriamente ditos, incluindo as áreas que mais depositam, as autorias e co-autoria dos dados, as instituições depositantes, etc, levantando um perfil do repositório.

REFERÊNCIAS

- Balbino, J.N., Silva, H.F.N., Fernandes, F.R., & Mischiatti, J.A.W. (2020). Conceitos de dados abertos: uma exploração da produção científica no contexto brasileiro. *Ciência da Informação*, 49(1), 151–164. <https://revista.ibict.br/ciinf/article/view/4813>
- Curty, R. (2019). Abordagens de reuso e a questão da reusabilidade dos dados científicos. *Liinc em Revista*, 15(2), 177-193. <http://revista.ibict.br/liinc/article/view/4777>.
- Deposita Dados. (2023). Sobre. <https://depositadados.ibict.br/about.xhtml>
- Organisation for Economic Co-operation and Development. (2007). Principles and Guidelines for Access to Research Data. https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr
- Resende, L.C., & Bax, M.P. (2019). Curadoria digital de dados científicos: o cenário brasileiro na ciência da informação. *Encontro nacional de pesquisa e pós-graduação em ciência da informação*; 20(20), 1-10. <https://brapci.inf.br/v/122929>
- Sales, L., Henning, P., Veiga, V., Costa, M. M., Sayão, L. F., da Silva Santos, L. O. B., & Pires, L. F. (2020). GO FAIR Brazil: A Challenge for Brazilian Data Science. *Data Intelligence*, 2(1-2), 238–245. https://doi.org/10.1162/dint_a_00046
- Sanchez, F. A., Vidotti, S. A. B. G., & Vechiato, F. L. (2017). A contribuição da curadoria digital em repositórios digitais. *Revista Informação Na Sociedade Contemporânea*, 1, 1–17. <https://doi.org/10.21680/2447-0198.2017v1n0ID12280>



Sayão, L.F., & Sales, L.F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, 21(2), 90-115. <https://doi.org/10.5433/1981-8920.2016v21n2p90>

Sayão, L.F., & Sales, L.F. (2020). Afinal, o que é dado de pesquisa?. *BIBLOS - Revista do Instituto de Ciências Humanas e da Informação*, 34. <http://hdl.handle.net/20.500.11959/brap-ci/162776>.

Triques, M. L.; Arakaki, A. C. S., & Castro, F. F. (2020). Aspectos da representação da informação na curadoria digital. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 25, 1-21. <https://doi.org/10.5007/1518-2924.2020.e69898>.

ANEXO 1

RESUMEN BIOGRÁFICO DE LOS AUTORES

Me. Letícia Guarany Bonetti

Bibliotecária pela Universidade de Brasília (2019) e Mestre em Ciência da Informação pela Universidade Federal de São Carlos (2023). Atua como pesquisadora bolsista no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) desenvolvendo serviços na área de Ciência Aberta e repositórios de dados.

Tatyane Guedes Martins da Silva

Bacharel em Biblioteconomia pela Universidade de Brasília (2019). Tem experiência na área de Ciência da Informação, com ênfase em Biblioteconomia. Atua como bolsista no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) desenvolvendo serviços na área de Ciência Aberta e repositórios de dados.

Dra. Caterina Groposo Pavão

Doutora em Comunicação e Informação e Bacharel em Biblioteconomia pela Universidade Federal do Rio Grande do Sul (UFRGS). Atua como professora do curso de graduação em Biblioteconomia da Faculdade de Biblioteconomia e Comunicação da UFRGS e no Pós-graduação em Ciência da Informação da mesma Universidade.



Dra. Samile Andréa de Souza Vanz

Professora associada do Departamento de Ciências da Informação e do Programa de Pós-graduação em Comunicação da Universidade Federal do Rio Grande do Sul (PPGCOM UFRGS). Graduada em Biblioteconomia pela Universidade Federal do Rio Grande do Sul (1999), mestre e doutora em Comunicação e Informação pelo PPGCOM UFRGS (2004 e 2009), com estágio sanduíche na Dalian University of Technology (China, 2007-2008). Pós-doutorado pela Universidad Carlos III de Madrid (Madrid, 2016). Editora da revista Em Questão (2014 – 2023).

Dr. Rene Faustino Gabriel Junior

Graduado em Biblioteconomia e Documentação pela Pontifícia Universidade Católica do Paraná (2008), com mestrado em Ciência, Gestão e Tecnologia da Informação pela Universidade Federal do Paraná (2011) e doutorado em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (2014). Atualmente é professor adjunto da Universidade Federal do Rio Grande do Sul e do Programa de Pós-Graduação em Ciência da Informação (PPGCIN) da mesma universidade e chefe do Departamento de Ciências da Informação (DCI).

Me. Marcel Garcia de Souza

Doutorando em Ciência da Informação pela Universidade de Brasília. Mestre em Educação em Ciências pela Universidade Federal do Rio Grande do Sul (2016). Graduado em Psicologia pela Universidade Católica de Brasília (2005). Analista em Ciência e Tecnologia no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) atuando como Coordenador de Tratamento, Análise e Disseminação da Informação Científica, além de coordenar pesquisas aplicadas voltadas à Ciência da Informação, Informação para Sustentabilidade, Avaliação do Ciclo de Vida e Informação Tecnológica.

Dr. Washington Luís Ribeiro de Carvalho Segundo

Doutor e Mestre em Informática pela Universidade de Brasília, com Estágio de Doutorado Sanduíche no Kings College London. Possui graduação em Matemática (Bacharelado e Licenciatura) também pela Universidade de Brasília. É Coordenador-geral de Informação Científica e Técnica no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict / MCTI). É membro e coordena projetos, comitês nas áreas de Ciência Aberta e Ciência de Dados.



ANEXO 2

REQUERIMIENTOS DE EQUIPO TÉCNICO PARA LA PRESENTACIÓN DE LA PONENCIA

Para a apresentação deste trabalho, será necessário um computador e um projetor para exibir a apresentação em powerpoint.