

Pecha kucha

Desafios na curadoria de dados de pesquisa no repositório nacional Deposita Dados

Challenges in curating research data in the national repository Deposita Dados

Desafíos en la curación de datos de investigación en el repositorio nacional Deposita Dados

Letícia Guarany Bonetti*

Mestre em Ciência da Informação

Instituto Brasileiro de Informação em Ciência e Tecnologia

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

Lattes: <http://lattes.cnpq.br/1895977717955732>

E-mail: leticiabonetti@ibict.br

Tatyane Guedes Martins da Silva

Bacharel em Biblioteconomia

Instituto Brasileiro de Informação em Ciência e Tecnologia

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

Lattes: <http://lattes.cnpq.br/7310861285054095>

E-mail: tatyanesilva@ibict.br

Caterina Groposo Pavão

Doutora em Comunicação e Informação

Universidade Federal do Rio Grande do Sul

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

Lattes: <http://lattes.cnpq.br/4834791532698069>

E-mail: caterina@ufrgs.br

Samile Andrea de Souza Vanz

Doutora em Comunicação e Informação

Universidade Federal do Rio Grande do Sul

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

Lattes: <http://lattes.cnpq.br/5243732207004083>

E-mail: samile.vanz@ufrgs.br

Rene Faustino Gabriel Junior

Doutor em Ciência da Informação

Universidade Federal do Rio Grande do Sul

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

Lattes: <http://lattes.cnpq.br/5900345665779424>

E-mail: rene.gabriel@ufrgs.br

Marcel Garcia de Souza

Mestre em Educação em Ciências Química da Vida e Saúde

Instituto Brasileiro de Informação em Ciência e Tecnologia

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

Lattes: <http://lattes.cnpq.br/9517728665816047>

E-mail: marcelsouza@ibict.br

Washington Segundo

Doutor em Informática

Instituto Brasileiro de Informação em Ciência e Tecnologia

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

Lattes: <http://lattes.cnpq.br/9453481318889500>

E-mail: washingtonsegundo@ibict.br

Resumo

O repositório Deposita Dados, lançado em 2023 e mantido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia, recebe conjuntos de dados de pesquisa provenientes de diversas regiões do Brasil e áreas do conhecimento, tornando essencial a implementação de um processo de curadoria eficiente. A curadoria envolve diversos processos, entre eles, a revisão dos metadados e dos arquivos depositados, garantindo a padronização e aderência aos Princípios FAIR (Findable, Accessible, Interoperable, Reusable). Neste artigo, apresentamos a experiência da curadoria no Deposita Dados, destacando os métodos adotados, os desafios encontrados e as soluções implementadas. A discussão contribui para o aprimoramento das boas práticas na gestão de repositórios de dados de pesquisa, fornecendo subsídios para aprimorar a infraestrutura de dados abertos no Brasil.

Palavras-chave: Curadoria de dados; Dados de pesquisa; Repositório Deposita Dados; Repositório de dados de pesquisa.

Abstract

The Deposita Dados repository, launched in 2023 and maintained by the Instituto Brasileiro de Informação em Ciência e Tecnologia, receives research datasets from various regions of Brazil and fields of knowledge, making the implementation of an efficient curation process essential. Curation involves several processes, including the review of metadata and deposited files, ensuring standardization and adherence to the FAIR Principles (Findable, Accessible, Interoperable, Reusable). In this article, we present the curation experience at Deposita Dados, highlighting the methods adopted, the challenges encountered, and the solutions implemented. The discussion contributes to the improvement of best practices in the management of research data repositories, providing insights for improving the open data infrastructure in Brazil.

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

Resumen

El repositorio Deposita Dados, lanzado en 2023 y mantenido por el Instituto Brasileiro de Informação em Ciência e Tecnologia, recibe conjuntos de datos de investigación de diversas regiones de Brasil y áreas de conocimiento, lo que hace esencial la implementación de un proceso de curación eficiente. La curación implica varios procesos, incluida la revisión de metadatos y archivos depositados, garantizando la estandarización y el cumplimiento de los Principios FAIR (Findable, Accessible, Interoperable, Reusable). En este artículo, presentamos la experiencia de curación en Deposita Dados, destacando los métodos adoptados, los desafíos encontrados y las soluciones implementadas. El debate contribuye a la mejora de las buenas prácticas en la gestión de repositorios de datos de investigación, aportando perspectivas para optimizar la infraestructura de datos abiertos en Brasil.

Palabras clave: Curación de datos; Datos de investigación; Repositorio Deposita Dados; Repositorio de datos de investigación.

Introdução

O Deposita Dados, gerenciado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), desempenha um papel importante na promoção do compartilhamento de dados de pesquisa no cenário brasileiro e no fortalecimento do movimento dos dados abertos. O repositório Deposita Dados foi lançado no XXII Seminário Nacional de Bibliotecas Universitárias (SNBU), em 2023, e tem como objetivo armazenar, disseminar e preservar conjuntos de dados coletados por pesquisadores brasileiros vinculados a instituições científicas que ainda não possuem seus repositórios de dados de pesquisa; e/ou conjuntos de dados coletados por pesquisadores brasileiros em projetos que envolvam colaboração científica internacional (Deposita Dados, 2025).

Por ser um repositório nacional e multidisciplinar, o Deposita Dados recebe conjuntos de dados de diversas regiões e áreas do conhecimento, o que amplia os desafios para a etapa de curadoria. A curadoria digital é compreendida por Pennock (2007, p. 1, tradução nossa) como a atividade de “[...] manter e agregar valor a um

corpo confiável de informação digital para uso atual e futuro: em outras palavras, é o gerenciamento ativo e a avaliação das informações digitais durante todo o seu ciclo de vida”. De acordo com Triques et al. (2020):

A curadoria digital apresenta-se como uma prática interdisciplinar abrangente que busca estabelecer diretrizes e um conjunto de ações inter-relacionadas para a manutenção do material com valor informacional, se valendo para isso, dos metadados para garantir todo o ciclo de gerenciamento dos dados, visando o acesso, o uso e o reuso a longo prazo. (p. 2)

A curadoria é um componente essencial na gestão dos dados de pesquisa e é considerada um dos maiores desafios para instituições e equipes gestoras de repositórios (Vanz et al., 2021). A execução de um fluxo de trabalho de curadoria bem articulado pode tornar os dados ainda melhores ao descrever seu conteúdo, criar uma estrutura coerente, fornecer documentação significativa, permitir a automação por meio de código e sintaxe, e vincular os dados a outros dados (Dataverse Curation Guide, 2021).

O Deposita Dados utiliza o software Dataverse e funciona por autoarquivamento. Todos os dados de pesquisa submetidos pelos pesquisadores passam por uma etapa de revisão pela equipe gestora, que avalia desde o preenchimento dos metadados até o formato e conteúdo dos arquivos depositados. Essa etapa é essencial para garantir a qualidade e a padronização, permitindo que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, conforme estabelecido pelos Princípios FAIR.

Com a finalidade de contribuir para a discussão das boas práticas na gestão de repositórios de dados de pesquisa, este trabalho tem como objetivo apresentar os métodos aplicados, os desafios enfrentados e as soluções adotadas na curadoria dos dados de pesquisa do Deposita Dados. A próxima seção apresenta os quatro principais desafios do processo de curadoria do Deposita Dados, e discute as soluções encontradas. A última seção apresenta as conclusões.

Desenvolvimento

O repositório Deposita Dados possui em seu fluxo de trabalho a etapa de curadoria. Essa etapa 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 repositório e a legislação brasileira.

Ao longo do processo de curadoria, etapa que antecede a rejeição ou a publicação dos conjuntos de dados, a equipe gestora do repositório identificou quatro desafios principais: 1) depósito de conjuntos de dados que continham dados pessoais ou sensíveis; 2) depósito de conjuntos 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, relatórios e artigos acadêmicos. Os

quatro desafios são detalhados a seguir, e estão disponíveis para consulta no conjunto de dados¹ depositado pelos autores no repositório Aleia.

No Brasil, a Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018) apresenta diretrizes para a coleta, o armazenamento, o tratamento e a disseminação de dados pessoais (Brasil, 2018). Esses princípios devem ser seguidos com a finalidade de garantir a privacidade, a segurança e a transparência no uso das informações pessoais. Como mencionado, o Deposita Dados, por ser um repositório multidisciplinar, recebe conjuntos de dados provenientes de diversas áreas, incluindo aqueles que podem conter dados pessoais ou passíveis de identificação, com especial atenção aos dados da área da Saúde. Diante desse contexto, torna-se fundamental garantir que o que é depositado no repositório esteja em conformidade com a LGPD.

A medida adotada para tentar evitar o depósito desses dados é o envio, por e-mail, de um documento chamado “Passo a passo para identificar se há dados pessoais”. É uma medida preventiva, já que todos os usuários que criam uma conta no Deposita Dados precisam enviar um e-mail solicitando permissão para depositar. Ao conceder a permissão, a equipe já envia o documento com as orientações sobre dados pessoais e o Guia do Usuário², que tem uma seção destinada à anonimização dos dados. Se, apesar dessa medida, arquivos com dados pessoais forem submetidos, a equipe deverá recusar o conjunto de dados e devolvê-lo ao autor. É nesse momento que a equipe aponta para o autor a presença dos dados pessoais e solicita que eles sejam excluídos ou anonimizados com base na Lei nº 13.709/2018 (Brasil, 2018). Assim que o depositante faz as correções, o conjunto de dados volta para a equipe, que o revisa novamente, e, então, o publica.

Já para o caso de dados não estruturados – outro desafio encontrado -, a equipe gestora também recusa o conjunto de dados e devolve para alterações do autor. Para auxiliar o depositante, é enviado um e-mail com recomendações para estruturar os dados, como conversão do formato, organização dos dados na planilha etc. Dados não estruturados dificultam a interpretação tanto por humanos quanto por máquinas, que não conseguem processar automaticamente os dados. Como o objetivo é que os conjuntos de dados do Deposita Dados estejam alinhados aos Princípios FAIR, existe essa preocupação com os arquivos depositados.

Outro desafio encontrado durante a curadoria é o preenchimento incorreto dos metadados. O repositório possui seis metadados obrigatórios, sendo eles: título, autor, e-mail para contato, descrição, assunto e palavras-chave; e 29 metadados opcionais, entre eles: título alternativo, publicação relacionada, colaborador e data de coleta. O preenchimento adequado dos metadados afeta diretamente a aplicabilidade dos Princípios FAIR e a facilidade de interpretação dos dados depositados.

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

² Disponível em: <https://depositadados.ibict.br/dvn/guide/guides.html>

O processo de curadoria envolve, entre outros, a verificação da completude dos metadados. Com isso, a criação do manual que orienta os pesquisadores no preenchimento dos metadados é essencial. O Deposita Dados disponibiliza o Guia do Usuário, que tem como objetivo fornecer instruções claras sobre o processo de submissão de dados no repositório. O documento é divulgado por e-mail quando a equipe fornece a permissão para o depósito dos dados, e está disponível para consulta no site do repositório.

O Guia do Usuário lista todos os campos e a forma correta de preenchê-los. Além disso, durante o preenchimento dos metadados, o usuário pode consultar uma explicação rápida de cada campo no próprio formulário de depósito. Mas, ainda assim, ocorrem casos de preenchimento incorreto. Um deles é quando o depositante registra a informação no campo errado, a exemplo da indicação da publicação relacionada (um artigo, uma tese, uma dissertação derivada do conjunto de dados) fora do campo “Publicação relacionada”. Outro erro é colocar a informação no campo certo, mas preencher de forma incorreta, como não colocar o nome dos autores na ordem indireta, ou seja, não fazer a entrada pelo sobrenome. Em ambos os casos, por serem correções rápidas e ligadas exclusivamente a um erro de preenchimento, a própria equipe gestora faz as alterações necessárias e publica o dataset.

Por fim, tem-se o caso dos pesquisadores que tentam depositar arquivos que não são conjuntos de dados. Alguns exemplos são arquivos contendo documentos como artigos, apresentações de powerpoint, relatórios, dissertações, entre outros. Esses documentos não se enquadram na definição de dado de pesquisa, qual seja: “[...] registros factuais usados como fontes primárias na pesquisa científica, e que são geralmente aceitos na comunidade científica como necessários para validar os resultados de pesquisa” (OECD, 2007, p. 13, tradução nossa). Essa definição está presente e exemplificada no Guia do Usuário. Nesses casos, o arquivo é recusado, e o autor é notificado por e-mail. De forma sucinta, a equipe do Deposita Dados explica que os arquivos não se enquadram na tipologia documental aceita no repositório, e sugere que o pesquisador tente um novo depósito, dessa vez, das fontes primárias que ele utilizou na pesquisa. São enviados por e-mail alguns exemplos para que ele consiga compreender a diferença entre os produtos científicos, sendo uma escolha do pesquisador tentar um novo depósito.

Conclusão

Sabe-se que o compartilhamento dos dados de pesquisa ainda é uma iniciativa recente no Brasil, diferentemente do que ocorre com publicações científicas, como artigos, teses e dissertações. Logo, é esperado que os pesquisadores encontrem dificuldades na hora de identificar, selecionar, estruturar e depositar seus dados de pesquisa.

A equipe do Deposita Dados vem tentando auxiliar esses processos não só por meio dos manuais do próprio repositório e das instruções por e-mail, mas também por meio

de capacitação gratuita. A equipe gestora, durante o ano de 2024, ministrou workshops aos pesquisadores para auxiliá-los em todo o processo de depósito dos dados. Os workshops envolveram desde a apresentação do software utilizado pelo repositório (Dataverse), até um passo a passo de como preencher os metadados, subir os arquivos e depositar dados no Deposita Dados. Ele foi ofertado para as equipes das cinco regionais que integram a Rede Brasileira de Repositórios Digitais (RBRD), que é coordenada pelo Ibict. Pretende-se ampliar a oferta de workshops em 2025.

O levantamento dos desafios na curadoria é essencial para que a equipe possa propor novos workshops e elaborar novos materiais de capacitação, como um Guia de Anonimização, para incentivar o compartilhamento dos dados de pesquisa no Brasil. Além disso, este trabalho pode servir como base para outras equipes de curadoria, padronizando soluções em repositórios de dados de pesquisa no país.

Agradecimentos

Agradecemos ao Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) pelo apoio institucional para o desenvolvimento desta pesquisa e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio por meio da Bolsa de Produtividade em Pesquisa (Processo nº 312975/2022-8), que tem contribuído significativamente para o desenvolvimento desta linha de investigação.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Repositório de Dados do Instituto Brasileiro de Informação em Ciência e Tecnologia (ALEIA): <https://doi.org/10.48472/aleia/TNAXHU>

CRedit – Contribuições dos Autores

Letícia Guarany Bonetti | Concetualização, Curadoria de dados, Escrita – redação original, recolha de dados, edição
Tatyane Guedes Martins da Silva | Concetualização, Curadoria de dados, Escrita – redação original, recolha de dados, edição
Caterina Groposo Pavão | Concetualização, Curadoria de dados, Escrita – redação original, recolha de dados, edição
Samile Andrea de Souza Vanz | Concetualização, Curadoria de dados, Escrita – redação original, recolha de dados, edição
Rene Faustino Gabriel Junior | Concetualização, Curadoria de dados, Escrita – redação original, recolha de dados, edição
Marcel Garcia de Souza | Escrita – revisão e edição, Supervisão
Washington Segundo | Escrita – revisão e edição, Supervisão

Referências

- Brasil. (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm
- Dataverse Curation Guide. (2021). <https://zenodo.org/records/5579820#.YZaCAdDMJaQ>
- Deposita Dados. (2025). Sobre o Deposita Dados. <https://depositadados.ibict.br/about.xhtml>
- Pennock, M. (2007). Digital Curation: a Life-Cycle Approach to Managing and Preserving Usable Digital Information. *Library & Archives*, 1–3. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf
- OECD. (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD. <https://web-archive.oecd.org/2018-04-10/137520-38500813.pdf>
- Triques, M. L., Arakaki, A. C. S., & Castro, F. F. de. (2020). Aspectos da representação da informação na curadoria digital. *Encontros Bibli*, 25, 1–21. <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e69898/43140>
- Vanz, S. A. S., Pavão, C. M. G., Caregnato, S. E., Passos, P. C. S. J., Moura, A. M. M., Borges, E. N., Gabriel Junior, R. F., & Rocha, R. P. (2021). Diretrizes para o estabelecimento de um checklist para curadoria de dados de pesquisa. *Informação em Pauta*, 6, 1–18. DOI: <https://doi.org/10.36517/2525-3468.ip.v6i00.2021.68088.1-18>