

Pecha Kucha

**Política de curadoria de dados pessoais e sensíveis no
repositório Deposita Dados**

**Policy for curating personal and sensitive data in the Deposita
Dados repository**

**Política de curaduría de datos personales y sensibles en el
repositorio Deposita Datos**

Letícia Guarany Bonetti*

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

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

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

leticiabonetti@ibict.br

Tatyane Guedes Martins da Silva

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

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

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

tatyanesilva@ibict.br

Caterina Groposo Pavão

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

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

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

caterina@ufrgs.br

Samile Andrea de Souza Vanz

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

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

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

samile.vanz@ufrgs.br

Rene Faustino Gabriel Junior

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

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

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

rene.gabriel@ufrgs.br



Marcel Garcia de Souza

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

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

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

marcelsouza@ibict.br

* Autor correspondente.

Resumo

O compartilhamento de dados de pesquisa, alinhado aos princípios da Ciência Aberta, amplia a transparência, a reprodutibilidade, a visibilidade, a reutilização e o impacto social da pesquisa. Nesse contexto, os repositórios de dados desempenham papel estratégico na descrição, preservação, acesso e reutilização segura desses recursos. O Deposita Dados, repositório multidisciplinar gerido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia, atende pesquisadores brasileiros e de países lusófonos vinculados a instituições sem repositórios próprios. Em 2024 e 2025, enfrentou desafios recorrentes na curadoria de conjuntos de dados com dados pessoais e sensíveis protegidos pela Lei Geral de Proteção de Dados Pessoais (LGPD). A equipe de curadoria mapeou esses casos e estabeleceu diretrizes internas baseadas em ações preventivas e na revisão criteriosa dos arquivos submetidos. Como solução, os conjuntos de dados são devolvidos aos autores para correções, como anonimização dos dados pessoais ou restrição de acesso em casos específicos. Essas práticas evidenciam o papel da curadoria na mediação entre abertura científica e proteção legal de dados, reforçando a necessidade de políticas institucionais.

Palavras-chave:

Dados pessoais; Curadoria de dados; Repositórios de dados de pesquisa; Ciência Aberta

Abstract

The sharing of research data, aligned with the principles of Open Science, expands transparency, reproducibility, visibility, reuse, and the social impact of research. In this context, data repositories play a strategic role in the description, preservation, access, and safe reuse of these resources. Deposita Dados, a multidisciplinary repository managed by the Instituto Brasileiro de Informação em Ciência e Tecnologia, serves Brazilian researchers and researchers from Portuguese-speaking countries affiliated with institutions without their own repositories. In 2024 and 2025, it faced recurring challenges in curating datasets containing personal and sensitive data protected by the Lei Geral de Proteção de Dados Pessoais (LGPD). The curation team mapped these cases and established internal guidelines based on preventive actions and careful review of submitted files. As a solution, datasets are returned to authors for correction, such as anonymizing personal data or restricting access in specific cases. These practices demonstrate the role of curation in

mediating between scientific openness and legal data protection, reinforcing the need for institutional policies.

Keywords:

Personal data; Data curation; Research data repositories; Open Science

Resumen

El intercambio de datos de investigación, alineado con los principios de la Ciencia Abierta, amplía la transparencia, la reproducibilidad, la visibilidad, la reutilización y el impacto social de la investigación. En este contexto, los repositorios de datos desempeñan un papel estratégico en la descripción, preservación, acceso y reutilización segura de estos recursos. Deposita Dados, repositorio multidisciplinario gestionado por el Instituto Brasileiro de Informação em Ciência e Tecnologia, atiende a investigadores brasileños y de países lusófonos vinculados a instituciones sin repositorios propios. En 2024 y 2025, enfrentó desafíos recurrentes en la curaduría de conjuntos de datos con datos personales y sensibles protegidos por la Lei Geral de Proteção de Dados Pessoais (LGPD). El equipo de curaduría mapeó estos casos y estableció directrices internas basadas en acciones preventivas y en la revisión rigurosa de los archivos sometidos. Como solución, los conjuntos de datos se devuelven a los autores para correcciones, como la anonimización de los datos personales o la restricción de acceso en casos específicos. Estas prácticas evidencian el papel de la curaduría en la mediación entre la apertura científica y la protección legal de datos, reforzando la necesidad de políticas institucionales.

Palabras clave:

Datos personales; Curación de datos; Repositorios de datos de investigación; Ciencia abierta

Introdução

Os dados de pesquisa são insumos valiosos para a ciência e trazem benefícios à sociedade e aos pesquisadores quando compartilhados conforme os princípios da Ciência Aberta. Entre as vantagens estão a promoção de colaborações internacionais, a validação de pesquisas, o aumento do impacto acadêmico, a economia de recursos públicos e a aceleração de novos resultados (Silva et al., 2024; Chaves et al., 2025). Nas Ciências da Saúde, por exemplo, a reutilização de dados tem se tornado comum por reduzir custos de novas coletas e permitir a combinação de diferentes tipos de dados (Quinn, 2017).

Os repositórios são tecnologias amplamente usadas para o armazenamento e o compartilhamento confiável de dados de pesquisa, pois apoiam sua descrição, preservação, acesso e reutilização. Segundo Curty (2019), sua criação e disseminação foram impulsionadas por

políticas governamentais e institucionais voltadas ao compartilhamento de dados financiados com recursos públicos.

No Brasil, há diversas iniciativas de repositórios (Gabriel Junior et al., 2019) e demandas pela gestão e depósito dos dados, como na Fundação de Amparo à Pesquisa do Estado de São Paulo (Monteiro & Lucas, 2019). Internacionalmente, cresce a exigência de instituições de pesquisa e agências de fomento, como National Science Foundation, National Institutes of Health e Wellcome Trust (Silva et al., 2024).

O incentivo ao compartilhamento de dados, somado às políticas mandatórias de revistas, eventos e agências de fomento, ampliou as exigências sobre pesquisadores financiados com recursos públicos. Nesse contexto, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) criou o Deposita Dados, repositório de dados gratuito voltado a pesquisadores brasileiros e de países lusófonos vinculados a instituições sem repositório próprio. A diversidade de áreas dos depositantes, contudo, impõe desafios à equipe do repositório (Bonetti et al., 2025).

Um dos principais desafios é a curadoria de *datasets* com dados pessoais e sensíveis, protegidos pela Lei nº 13.709/2018, a Lei Geral de Proteção de Dados Pessoais (LGPD). Os casos identificados estão ligados sobretudo às áreas da Saúde e das Ciências Sociais. A partir de 2024, a equipe de curadoria adotou o procedimento de identificar as ocorrências, registrá-las em planilha (para controle e análise) e então aplicar soluções padronizadas para impedir a disponibilização desses dados em acesso aberto no Deposita Dados.

Com o objetivo de contribuir para as boas práticas na gestão de repositórios de dados de pesquisa, este estudo descritivo apresenta os casos identificados e as soluções adotadas na curadoria de *datasets* com dados pessoais e sensíveis no Deposita Dados em 2024 e 2025. As diretrizes apresentadas podem orientar a curadoria em outros repositórios.

Desenvolvimento

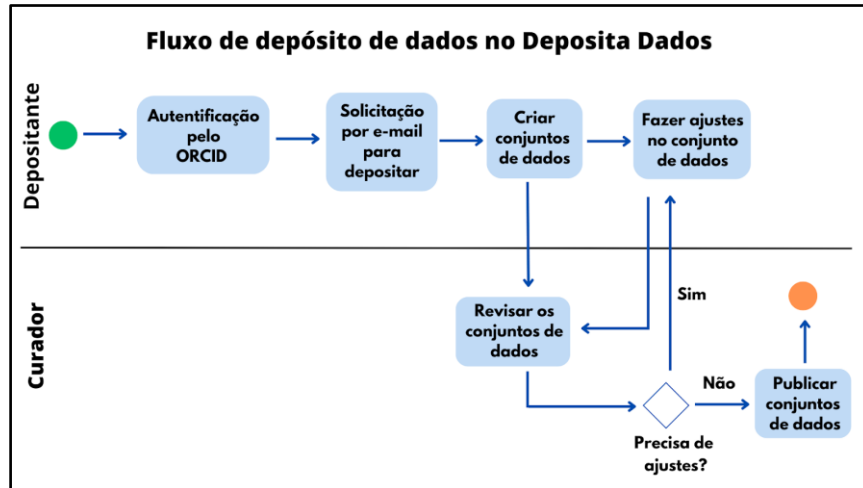
Dados de pesquisa são, em termos simples, aqueles gerados e coletados durante uma investigação. Suas tipologias e definições variam conforme a área, pois são produzidos com diferentes finalidades, por distintas comunidades acadêmicas e científicas e por diversos processos (Sayão & Sales, 2015). Essa heterogeneidade impacta a curadoria em repositórios digitais, infraestruturas voltadas à organização, ao acesso, à disseminação e à preservação dos dados de pesquisa (Sanchez et al., 2017).

A Figura 1 apresenta o fluxo de depósito no Deposita Dados. Por autoarquivamento, o pesquisador cria a conta, preenche os metadados e faz o upload dos arquivos desejados. Antes da publicação, porém, o conjunto de dados passa pela revisão da equipe, que verifica todos os metadados e arquivos enviados. Segundo Pennock (2007), a curadoria busca “manter e agregar valor a um corpo confiável de informação digital para uso atual e futuro” (p. 1, tradução nossa),

por meio do gerenciamento ativo ao longo de seu ciclo de vida. No Deposita Dados, essa etapa envolve a análise dos metadados e dos arquivos depositados, sendo a verificação do conteúdo dos arquivos o principal desafio.

Figura 1

Fluxo de depósito de dados no Deposita Dados



Nota: elaboração própria.

Em 2024 e 2025, a equipe do Deposita Dados identificou, entre as 78 submissões aceitas, cinco casos com dados pessoais e/ou sensíveis. Pela LGPD, dado pessoal é a informação relacionada a pessoa natural identificada ou identificável, e dado sensível inclui, entre outros, informações sobre saúde, origem racial ou étnica, convicção religiosa, opinião política, vida sexual, dados genéticos ou biométricos (Brasil, 2018, Cap. I, art. 5, inc. I, II).

Segundo a Autoridade Nacional de Proteção de Dados (ANPD), a LGPD busca equilibrar a proteção de dados pessoais, a privacidade e a autodeterminação informativa com a liberdade acadêmica e a circulação de informações necessárias à pesquisa. O artigo 13 da LGPD reafirma o uso de dados pessoais para fins científicos, com requisitos de segurança e prevenção, especialmente em pesquisas em saúde pública (Vargas, 2023).

Nesse contexto, dados pessoais e sensíveis devem ser tratados no âmbito da instituição de pesquisa, conforme protocolo ou consentimento, armazenados em ambientes seguros e, sempre que possível, submetidos à anonimização ou pseudonimização, vedada a divulgação de informações pessoais nos resultados (Vargas, 2023). Para evitar a disponibilização pública desses dados, a equipe envia ao depositante uma *checklist* preventivo com orientações para identificar

dados pessoais, escolher o método de anonimização e reforçar a necessidade de irreversibilidade, além do Guia de Usuário¹, que contém uma seção sobre anonimização.

Para apoiar os pesquisadores, a equipe elaborou o Guia de Boas Práticas para Anonimização de Dados de Pesquisa (Pavão et al., 2026), com conceitos, técnicas e exemplos práticos. Publicado pela Editora Ibict em abril de 2026, o Guia passou a ser enviado junto com o Guia de Usuário e foi disponibilizado para consulta e download gratuito. Mesmo com essas medidas, ainda há submissões com dados pessoais ou sensíveis. A identificação é totalmente manual e envolve a verificação de todos os arquivos em busca de dados que identifiquem indivíduos, como CPF, RG, nome completo, telefone e endereço, além de dados sensíveis previstos na LGPD.

Atualmente, a curadoria é realizada por duas bibliotecárias, e o tempo de análise varia conforme a quantidade e a complexidade dos arquivos. Nesse sentido, a equipe está avaliando o uso de ferramentas de inteligência artificial para automatizar parte do processo e, em casos complexos, é feita uma consulta ao comitê gestor do repositório, composto por profissionais especializados em gestão de dados de pesquisa. O autor não tem um prazo para realizar as alterações nos arquivos, mas o conjunto de dados só é publicado após a correção solicitada.

Quando há ocorrências, o conjunto de dados é devolvido para correções por meio da funcionalidade “retornar ao autor” do Dataverse, que permite ao curador solicitar alterações, e ao pesquisador editar novamente o registro. Essa funcionalidade é essencial porque a política interna do Deposita Dados impede que curadores alterem o conteúdo dos arquivos depositados. A equipe orienta o pesquisador por e-mail, recomendando anonimização ou, excepcionalmente, restrição de acesso, já que o repositório é de acesso aberto. A restrição deve ser acordada e justificada pelo pesquisador e, até o momento, não foi adotada nos casos identificados.

Dos cinco casos citados, quatro envolviam dados pessoais em planilhas, como telefones, e-mails e nomes completos de bebês internados. Em um deles, o autor não realizou a anonimização, o que impossibilitou a publicação do conjunto de dados, em conformidade com a LGPD. O quinto caso envolvia transcrições com nomes de participantes. Após consultar o trabalho final associado, a equipe verificou que os nomes não haviam sido citados e recomendou a padronização entre a tese e o conjunto de dados. A alteração foi realizada.

A LGPD define anonimização como o uso de meios técnicos razoáveis para impedir a associação, direta ou indireta, de um dado a um indivíduo (Brasil, 2018, Cap. I, art. 5, inc. XI). Entre os métodos mais comuns estão generalização, supressão e mascaramento. A supressão é recomendada para identificadores únicos, como nome completo, CPF, RG, e-mails ou telefones pessoais. Nesses casos, o pesquisador remove os campos correspondentes e envia nova versão ao repositório. O mascaramento, segundo El Emam e Luk Arbuckle (2013), reduz o risco de identificação por técnicas de transformação, mas pode comprometer a análise. Por isso, é indicado quando os

¹ Disponível em: <https://depositadados.ibict.br/guide.xhtml>

campos mascarados não serão necessários à pesquisa, como na substituição de números do CPF por símbolos.

Quando é necessário preservar a utilidade dos dados, deve-se equilibrá-la com a privacidade do sujeito, pois “não é relevante disponibilizar um conjunto de dados se estes não apresentarem utilidade para o usuário” (Affonso et al., 2017, p. 82). Nesses casos, a generalização permite manter parte da utilidade dos dados. Para Affonso, Oliveira e Sant'Ana (2017), sua vantagem é preservar a representação semântica e ampliar registros com valores similares, dificultando a identificação. Exemplos incluem transformar idade em faixa etária ou localização precisa em região, com atenção a intervalos pequenos que facilitem a reidentificação.

Embora a equipe recomende soluções, cabe ao depositante escolher as técnicas e garantir a privacidade dos titulares dos dados. Após as correções, o conjunto de dados passa por nova revisão e, se estiver conforme as diretrizes do repositório, é publicado. Além da verificação dos arquivos, recomenda-se informar, na descrição do conjunto de dados, o parecer do Comitê de Ética em Pesquisa e o Certificado de Apresentação para Apreciação Ética (CAAE), o que amplia a confiabilidade dos dados e permite consultar as informações éticas da pesquisa. Com essas medidas, o Deposita Dados busca evitar a disponibilização indevida de dados pessoais e sensíveis em acesso aberto.

Considerações finais

Os repositórios de dados de pesquisa são infraestruturas essenciais para acesso, preservação e reutilização de dados em consonância com a Ciência Aberta. Contudo, dados pessoais e sensíveis não podem ser compartilhados abertamente. No Deposita Dados, a equipe de curadoria identificou cinco casos de arquivos com esses dados e, como resposta, estabeleceu diretrizes internas, incluindo guias e o envio preventivo de *checklist* para identificação de dados pessoais.

Quando há identificação de dados pessoais ou sensíveis, o conjunto é devolvido ao autor para correção, com orientações por e-mail. As alterações podem envolver anonimização adequada ou, em casos específicos, restrição de acesso. Após as correções do autor, os curadores revisam novamente os arquivos e, se não houver dados pessoais ou sensíveis, o conjunto de dados é publicado. Essa dupla verificação garante a conformidade das correções.

O levantamento sistemático dos casos subsidia ações formativas alinhadas às necessidades dos pesquisadores, permite identificar erros recorrentes e lacunas sobre anonimização e manejo de dados pessoais ou sensíveis, além de apoiar a criação de materiais como o Guia de Boas Práticas para Anonimização de Dados de Pesquisa. Os resultados também podem orientar outras equipes de curadoria, oferecendo base empírica para padronizar procedimentos e fortalecer diretrizes sobre tratamento e publicação de conjuntos de dados em repositórios.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Disponíveis no repositório Aleia: <https://doi.org/10.48472/aleia/FBWT7F>

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, Escrita – redação original, edição

Samile Andrea de Souza Vanz | Concetualização, Escrita – redação original, edição

Rene Faustino Gabriel Junior | Concetualização, Escrita – redação original, edição

Marcel Garcia de Souza | Escrita – revisão e edição, Supervisão

Referências

Affonso, E. P., Oliveira, S. C. de, & Sant'Ana, R. C. G. (2017). Análise do equilíbrio entre privacidade e utilidade no acesso a dados. *Informação & Sociedade*, 27(1), 81-92. <http://hdl.handle.net/11449/228317>

Bonetti, L. G., Silva, T. G. M., Pavão, C. M. G., Vanz, S. A. S., Gabriel Junior, R. F., Souza, M. G., & Segundo, W. L. R. (2025). Desafios na curadoria de dados de pesquisa no repositório nacional Deposita Dados. *Revista Ciência Aberta Lusófona*, 1, 179-186. <https://doi.org/10.82524/recal.2025.38>

Brasil. (2018). Lei 13.709 de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União* (agosto). www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm

Chaves, I. T., Fernandes, M. da C., Bertoleza, N. L. C., Dias, G. A., & Sousa, M. R. F. (2025). Percepções e práticas no âmbito do compartilhamento e reuso de dados: análise dos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação do Brasil. *Em Questão*, 31, e146102. <https://doi.org/10.1590/1808-5245.31.146102>

Curty, R. (2019). Abordagens de reuso e a questão da reusabilidade dos dados. *Liinc em Revista*, 15(2), 177-193. <http://revista.ibict.br/liinc/article/view/4777>

El Emam, K., & Arbuckle, L. (2013). *Anonymizing health data: Case studies and methods to get you started*. O'Reilly Media.

Gabriel Junior, R. F., Rocha, R. P., Caregnato, S. E., Pavão, C. M. G., Passos, P. C. S. J., Borges, E. N., Vanz, S. A. S., & Azambuja, L. A. B. (2019). Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. *Ciência da Informação*, 48(3), 87-101. <https://lume.ufrgs.br/handle/10183/212266>

Monteiro, G., & Lucas, E. R. O. (2019). Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. *AtoZ: Novas práticas em informação e Conhecimento*, 8(1), 13–20. <https://doi.org/10.5380/atoz.v8i1.67253>

Pavão, C. M. G., Bonetti, L. G., Souza, M. G., Gabriel Junior, R. F., Vanz, S. A. S., Silva, T. G. M., & Carvalho Segundo, W. L. R. (2026). *Guia de Boas Práticas para Anonimização de Dados de Pesquisa*. Editora do IBICT. <https://repositorio.upf.br/handle/123456789/10196>

Pennock, M. (2007). *Digital Curation: a Life-Cycle Approach to Managing and Preserving Usable Digital Information*. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf

Quinn, P. (2017). The Anonymisation of Research Data — A Pyric Victory for Privacy that Should Not Be Pushed Too Hard by the eu Data Protection Framework? *European Journal of Health Law*, 24(4), 347-367. <https://doi.org/10.1163/15718093-12341416>

Sanchez, F. A., Vidotti, S. A. B. G., & Vechiato, F. L. (2019). A contribuição da curadoria digital em repositórios digitais. *Revista Informação na Sociedade Contemporânea*, 1, 1-17. <https://periodicos.ufrn.br/informacao/article/view/12280>

Sayão, L. F., & Sales, L. F. (2015). *Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores*. CNEN/Instituto de Engenharia Nuclear.

Silva, L. C., Prado, C. B., & Araújo, E. A. (2024). Fatores que influenciam o processo de compartilhamento e reuso de dados de pesquisa: revisão sistemática de literatura a partir da base de dados BRAPCI. *AtoZ: Novas Práticas em Informação e Conhecimento*, 13, 1-15. <https://doi.org/10.5380/atoz.v13i0.91352>

Vargas, A. G., Rabelo, A. H. A., Costa, D. V., Maciel, F. M., Gonçalves, G., Carvalho, L. B., & Favero, S. F. M. (2023). *Tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas: guia orientativo*. ANPD. <https://www.gov.br/anpd/pt-br/centrais-de-conteudo/materiais-educativos-e-publicacoes/web-guia-anpd-tratamento-de-dados-para-fins-academicos.pdf>