

Pecha kucha

A presença brasileira no registro de repositórios de dados de pesquisa

The Brazilian presence in the registry of research data repositories

La presencia brasileña en el registro de repositorios de datos de investigación

Gisela Fernanda Monteiro Danin
Mestrado em Ciência da Informação
Universidade Federal do Pará
ORCID: <https://orcid.org/0000-0002-4068-4318>
Lattes: <http://lattes.cnpq.br/7338198971491367>
E-mail: giselamdanin@gmail.com

Hamilton Vieira de Oliveira
Doutorado em Ciência da Informação
Universidade de Brasília
ORCID: <https://orcid.org/0000-0002-6439-0058>
Lattes: <http://lattes.cnpq.br/3172995771315859>
E-mail: hamilton@ufpa.br

Alegria Benchimol
Doutorado em Ciência da Informação
Ibict/Universidade Federal do Rio de Janeiro
ORCID: <https://orcid.org/0000-0003-0920-992X>
Lattes: <http://lattes.cnpq.br/9550647395377220>
E-mail: alegria.benchimol@gmail.com

Resumo

Este estudo tem como objetivo identificar e analisar a presença de repositórios brasileiros de dados de pesquisa no diretório re3data.org, reconhecido por mapear e descrever repositórios de dados científicos em escala global. A pesquisa é bibliográfica, descritiva e quantitativa. Inicialmente, realizou-se uma

revisão de literatura sobre repositórios de dados de pesquisa, seguida pela coleta de dados na fonte e pela organização destes utilizando planilhas do Excel. Os resultados revelam a presença de 12 repositórios de dados de pesquisas genuinamente brasileiros no re3data.org. Isso mostra que as instituições brasileiras, assim como seus pesquisadores, estão dando um passo importante rumo à proposta da ciência aberta. No entanto, ainda são poucos os repositórios de dados de pesquisas brasileiros no diretório internacional analisado. Percebeu-se, adicionalmente, que a maioria dos repositórios de dados de pesquisa encontrados pertencem a instituições de ensino superior e à área das humanidades, que também é a área com mais dados disponibilizados.

Palavras-chave: Acesso aberto; Ciência aberta; Repositório de dados de pesquisa; Registro de Re却itórios de Dados de Pesquisa.

Abstract

This study aims to identify and analyze the presence of Brazilian research data repositories in the re3data.org directory, which is recognized for mapping and describing scientific data repositories on a global scale. The research is bibliographic, descriptive, and quantitative. Initially, a literature review on a research data repositories was conducted, followed by data collection from the source and its organization using Excel spreadsheets. The results reveal the presence of 12 genuinely Brazilian research data repositories in re3data.org. This shows that Brazilian institutions, as well as their researchers, are taking an important step towards the open science proposal. However, the number of Brazilian research data repositories in the analyzed international directory is still small. Additionally, it was noted that most of the repositories of the research data repositories found belong to higher education institutions and the humanities field, which is also the area with the most data made available.

Keywords: Open Access; Open Science; Research Data Repository; Registry of Research Data Repositories.

Resumen

Este estudio tiene como objetivo identificar y analizar la presencia de repositorios brasileños de datos de investigación en el directorio re3data.org, reconocido por mapear y describir repositorios de datos científicos a escala global. La investigación es bibliográfica, descriptiva y cuantitativa. Inicialmente, se realizó una revisión de literatura sobre repositorios de datos de investigación, seguida de la recolección de datos en la fuente y su organización utilizando hojas de cálculo de Excel. Los resultados revelan la presencia de 12 repositorios de datos de investigación genuinamente brasileños en re3data.org. Esto muestra que las instituciones brasileñas, así como sus investigadores, están dando un paso importante hacia la propuesta de ciencia abierta. Sin embargo, aún son pocos los repositorios de datos de investigación brasileños en el directorio internacional analizado. Adicionalmente, se percibió que la mayoría de los repositorios de datos de investigación encontrados pertenecen a instituciones de educación superior y al área de humanidades, que es también el área con más datos disponibles.

Palabras clave: Acceso Abierto; Ciencia Abierta; Repositorio de Datos de Investigación; Registro de Re却itórios de Datos de Investigación.

Repositório de dados de pesquisa e sua importância no contexto da ciência aberta

Para Silva e Silveira (2019), a proposta da ciência aberta vai desde a concepção da investigação até a disponibilização em software livre. O propósito é elaborar metodologias para a gestão de dados brutos coletados durante todo o percurso da pesquisa, para que possam ser disponibilizados e reutilizados, sem custos, por todos os níveis da sociedade (Danin, 2021).

Os dados de pesquisa incluem qualquer registro coletado, observado ou gerado por pesquisas científicas e aceito pela comunidade científica para validar resultados. Eles variam conforme os pesquisadores e as áreas do conhecimento e precisam ser geridos com metadados para facilitar sua recuperação. Podem ser números, vídeos, softwares, equações, animações, entre outros, e se dividem em três tipos: experimentais (laboratório), computacionais (modelos) e observacionais (fenômenos irrepetíveis) (Sayão; Sales, 2016).

Os Repositórios de Dados de Pesquisa (RDPs) são essenciais para armazenar, organizar e disponibilizar esses dados gratuitamente, garantindo visibilidade, segurança, reuso e interoperabilidade. Existem quatro tipos de RDPs: institucionais, disciplinares, multidisciplinares e orientados por projetos. A implementação de um RDP exige planejamento estratégico e conhecimento dos fluxos de dados da instituição (Sayão; Sales, 2016).

O re3data.org – registro de dados de pesquisa – é um diretório criado para orientar os pesquisadores, enquanto produtores e usuários de dados, já que proporciona uma descrição dos repositórios digitais de dados de pesquisa existentes no cenário mundial em uma única base de dados. Possui parceria com o Departamento de Biblioteca e Serviços de Informação (LIS) do Centro Alemão de Pesquisa em Geociências (GFZ), a Escola de Biblioteconomia e Ciência da Informação de Berlim, da Humboldt-Universität zu Berlin, e a Biblioteca KIT, do Instituto de Tecnologia de Karlsruhe (KIT) (Panpel et al, 2013).

O re3data.org apresenta em sua estrutura sete ícones desenvolvidos com as principais características de um repositório – cada um possui seu significado – para auxiliar os usuários a identificar um repositório adequado para o armazenamento de seus dados, além de permitir comparar repositórios, identificando pontos fortes e fracos de suas próprias infraestruturas. Para que um repositório seja indexado no re3data.org, são indispensáveis detalhes sobre acesso e licenciamento dos dados da pesquisa (Panpel, et al, 2013).

As versões do vocabulário da infraestrutura para a descrição dos dados no re3data.org foram desenvolvidas de forma colaborativa por vários projetos – que contribuíram com sugestões que foram levadas em consideração –, como a reBIND¹, DataCite² e OpenAIREplus³, entre outros. Os aspectos apontados foram informações gerais, tipos

¹ <http://rebind.bgbm.org>.

² <http://www.datacite.org>.

³ <http://www.openaire.eu>.

de conteúdo, palavras-chave, responsabilidades, políticas, aspectos legais e padrões técnicos e de qualidade (Panpel, et al, 2013).

Considerando esse contexto e essas referências, esta pesquisa objetiva identificar e analisar a presença de RDPs brasileiros no registro de repositórios de dados de pesquisa re3data.org. Trata-se de uma pesquisa bibliográfica com abordagem quantitativa e que se caracteriza como descritiva. Inicialmente, realizou-se uma breve revisão de literatura sobre repositórios de dados de pesquisa, seguida pela coleta de dados na fonte re3data.org e pela organização desses dados em planilhas Excel. Metodologia similar foi utilizada por Bonetti et al. (2024), que analisaram as características dos repositórios no re3data. O propósito principal deste estudo é avaliar a presença brasileira no diretório internacional de registros de dados de pesquisa.

Repositórios brasileiros de dados de pesquisa no re3data

Com base no levantamento realizado no re3data.org, selecionando o Brasil, foram encontrados 22 repositórios. No entanto, ao analisar cada um deles, observou-se que nove são na verdade internacionais, mas indexam os dados de instituição ou pesquisador brasileiro. Além disso, um outro, apesar de brasileiro, apresenta-se indisponível. Nesse sentido, foram analisados os 12 repositórios brasileiros restantes, a saber: o Domus Dados (Unifesp)⁴; a Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR)⁵; o COVID-19 Data Sharing/BR da FAPESP⁶; o Redape (Embrapa)⁷; o da Unesp⁸; o SciELO Data (Multidisciplinar)⁹; o REDU (Unicamp)¹⁰; a CEDAP Research Data Repository (Centro de Documentação e Acervo Digital da Pesquisa)¹¹; o Open Research Data (PUC-Rio)¹²; o Deposita Dados (Ibict)¹³; o Maenduar (Ibict)¹⁴; e o ALEIA (Ibict)¹⁵.

Esses 12 repositórios de dados são genuinamente brasileiros, utilizando como linguagem padrão o português, além do inglês, com exceção do gerido pela FAPESP, que se apresenta apenas em língua portuguesa. Em sua maioria, são de natureza institucional – administrados por uma instituição –, com três de natureza disciplinar e um multidisciplinar. Cabe destacar que todos os RDPs estão disponíveis, abertamente, ao público.

⁴ <https://domusdados.unifesp.br/>.

⁵ <https://bdc.c3sl.ufpr.br>.

⁶ <https://repositoriodatasharingfapesp.uspdigital.usp.br>.

⁷ <https://www.redape.dados.embrapa.br>.

⁸ <https://repositorio.unesp.br/home>.

⁹ <https://data.scielo.org>.

¹⁰ <https://redu.unicamp.br>.

¹¹ <https://cedap.ufrgs.br/jspui/handle/2050011959/90>.

¹² <https://www.maxwell.vrac.puc-rio.br/projetosEspeciais/ResearchData/index.php?b=1>.

¹³ <https://depositadados.ibict.br>.

¹⁴ <https://zenodo.org/communities/larhud/records?q=&l=list&p=1&s=10&sort=newest>.

¹⁵ <https://aleia.ibict.br>.

Os tipos de conteúdo e os assuntos armazenados em cada repositório variam de acordo com a sua finalidade. Os repositórios institucionais pertencentes às instituições de educação armazenam os dados de pesquisa de sua comunidade de servidores e alunos, levando em consideração as áreas de conhecimento das pós-graduações existentes nas instituições. Como exceção, o RDP da FAPESP contém também dados de pacientes que foram testados para a covid-19 (positivos ou negativos) nas instituições de saúde participantes no Brasil. Essa iniciativa disponibiliza três tipos de dados pseudonimizados: dados demográficos (sexo, ano de nascimento e região de residência), exames clínicos e exames laboratoriais. Informações adicionais sobre hospitalização – como dados sobre transferências e resultados – são fornecidas quando disponíveis.

O RDP SciELO Data disponibiliza os dados de pesquisas na área das humanidades, provenientes de artigos submetidos e aprovados para a publicação, já publicados em periódicos da Rede SciELO ou depositados no SciELO Preprints (RE3DATA.ORG, 2024).

Por fim, existem os três RDPs pertencentes ao Ibict, cada um criado para uma finalidade específica. O Deposita Dados é de caráter público. Pode ser usado por qualquer pesquisador brasileiro vinculado a instituições científicas que ainda não possuem repositórios próprios de dados de pesquisa e/ou pesquisadores brasileiros que executaram seus conjuntos de dados por meio de colaboração científica em instituições estrangeiras, com interesse em depositar seus dados. O Maenduar, criado pelo LARHUD¹⁶, é um repositório institucional dedicado a dados de pesquisa em Humanidades Digitais e Humanidades. Já o Aleia, repositório digital oficial do Ibict, faz o arquivamento, a publicação, a divulgação, a preservação e o compartilhamento de dados digitais de pesquisa produzidos pela comunidade científica do Ibict (RE3DATA.ORG, 2024).

Cada um desses repositórios cobre áreas temáticas específicas, como Medicina, Ciências da Vida, Ciências Naturais, Ciências Humanas, Ciências Sociais, Ciências da Engenharia, Medicina Veterinária, Zoologia, Biologia, Agricultura, Engenharia de Sistemas, Ciência da Computação, Engenharia Elétrica, entre outras, que são disponibilizadas em uma variada tipologia documental – como bancos de dados, formatos de dados científicos e estatísticos, imagens, código fonte, gráficos estruturados, aplicativos de software e dados arquivados.

Considerações finais

Este estudo mostra a presença de RDPs brasileiros no re3data.org, proporcionando visibilidade aos dados de pesquisa. Também identifica as instituições que possuem os RDPs institucionalizados, os assuntos e os tipos de conteúdo armazenado em cada repositório. Demonstrou-se ainda o benefício de permitir que os usuários consultem os dados de pesquisas brasileiras em uma única fonte, o re3data.org.

¹⁶ Laboratório em Rede de Humanidades Digitais.

Entre os RDPs levantados, constatou-se a presença de 12 genuinamente brasileiros, indexados nesse diretório de registro de dados de pesquisa de âmbito internacional. Isso demonstra que as instituições brasileiras, assim como seus pesquisadores, estão dando um passo importante na cultura da ciência aberta. No entanto, ainda são poucos os RDPs brasileiros nessa condição – existe uma variedade de áreas de conhecimento que ainda não adotou essa prática. Com a análise, percebeu-se que os RDPs vinculados a instituições de ensino superior e a campos das humanidades são os que mais têm dados disponibilizados.

O diálogo acerca da disponibilização dos dados de pesquisa, pelos pesquisadores, em repositórios de dados começa a ganhar força perante as instituições que fomentam a pesquisa, e será a consolidação desse avanço que tornará acessíveis os dados de todas as etapas, antes restritos aos pesquisadores. Isso vai proporcionar celeridade e reduzir os custos das novas pesquisas, contribuindo, portanto, para o avanço da ciência brasileira.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

CRediT – Contribuições dos Autores

Gisela Danin | Conceitualização, Escrita – redação original, Supervisão, Recolha de dados

Hamilton Vieira | Escrita – revisão e edição, Supervisão

Alegria Benchimol | Escrita – revisão e edição, Supervisão

Referências

Danin, G. F. M. (2021). Acesso aberto à informação científica: uma análise dos repositórios institucionais dos Institutos Federais de Educação, Ciência e Tecnologia do Brasil. [Dissertação, Instituto de Ciências Sociais Aplicadas]. Belém. <https://repositorio.ufpa.br/handle/2011/14112>

Panpel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M. (2013). Making research data repositories visible: the re3data.org registry. PLOS ONE, 8(11), 1-18. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>

Re3data (2024). Registry of research data repositories Consultado a 1 de março de 2024. <https://www.re3data.org/>

Sayão, L. F., Sales, L. F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. Informação & informação, 21(2), 90-115. <https://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>

Silva, F. C. C. da, Silveira, L. da. (2019). O Ecossistema da Ciência Aberta. Transinformação, 31, 1-13. <https://www.scielo.br/pdf/tinf/v31/2318-0889-tinf-31-e190001.pdf>