

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXV ENANCIB

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

DADOS DE PESQUISA NO BRASIL: INICIATIVAS NACIONAIS E EVIDÊNCIAS DE COMPARTILHAMENTO A PARTIR DA *OPENALEX*

RESEARCH DATA IN BRAZIL: NATIONAL INITIATIVES AND EVIDENCE OF SHARING FROM *OPENALEX*

Josimara Dias Brumatti – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT);
Universidade Federal Fluminense (UFF)

Fabiana de Melo Amaral Gonçalves Pinto – Universidade Federal Fluminense (UFF)

Maria Simone de Menezes Alencar – Universidade Federal do Estado do Rio de Janeiro
(UNIRIO)

Modalidade: Trabalho completo

Resumo: A promoção da ciência aberta tem incentivado pesquisadores e instituições a adotarem práticas mais transparentes e colaborativas, incluindo o compartilhamento dos dados de pesquisa. Este estudo teve como objetivo analisar os conjuntos de dados publicados por autores vinculados a instituições brasileiras e indexados na plataforma *OpenAlex*, entre 2019 e 2024. Adotou-se uma abordagem quali-quantitativa, de natureza descritivo-exploratória, onde foram analisados 9.652 conjuntos de dados. Os resultados indicam predominância do depósito em repositórios internacionais, acesso majoritariamente aberto e forte presença de coautoria internacional, especialmente com os Estados Unidos. Áreas como Ciências da Saúde, Ciências Ambientais e Ciência da Computação concentram a maior parte dos dados depositados. Algumas instituições brasileiras, como ICMBio, USP e Fiocruz, demonstram avanços institucionais importantes. No entanto, diversos repositórios nacionais ainda não estão indexados na *OpenAlex*, o que compromete sua visibilidade global. Conclui-se que, embora haja progressos, persistem lacunas na interoperabilidade entre infraestruturas nacionais e sistemas internacionais, demandando políticas coordenadas para fortalecer a presença dos dados brasileiros no ecossistema global da ciência aberta.

Palavras-chave: Ciência Aberta; Dados de pesquisa; *OpenAlex*; Brasil.

Abstract: The promotion of open science has encouraged researchers and institutions to adopt more transparent and collaborative practices, including the sharing of research data. This study aimed to analyze datasets published by authors affiliated with Brazilian institutions and indexed on the OpenAlex platform between 2019 and 2024. A qualitative-quantitative, descriptive-exploratory approach was adopted, through which 9,652 datasets were analyzed. The results indicate a predominance of deposits in international repositories, mostly under open access, and a strong presence of international coauthorship, especially with the United States. Fields such as Health Sciences, Environmental Sciences, and Computer Science account for most of the deposited datasets. Some Brazilian institutions, such as ICMBio, USP, and Fiocruz, demonstrate important institutional advances. However, several national repositories are still not indexed in OpenAlex, which compromises their global visibility. It is concluded that, despite the progress made, gaps remain in the

interoperability between national infrastructures and international systems, highlighting the need for coordinated policies to strengthen the presence of Brazilian research data within the global open science ecosystem.

Keywords: Open Science; Research datasets; OpenAlex; Brazil.

1 INTRODUÇÃO

Os avanços tecnológicos trazidos pela popularização das Tecnologias da Informação e Comunicação (TICs), impulsionaram a iniciativa como Arquivos Abertos e o Movimento do Acesso Aberto (MAA), que buscava democratizar o acesso à produção científica de modo gratuito e sem custo (Acesso Aberto – AA).

O Movimento do Acesso Aberto, materializado em múltiplas declarações a partir dos anos 2000, dentre as quais se destaca a *Budapest Open Access Initiative* (BOAI), atualmente em sua terceira versão, configurou-se, conforme Albagli (2017), como base para o surgimento da ciência aberta.

A dinâmica da ciência mundial vem se transformando a partir da proposição de novas práticas, tendências e tecnologias. O apelo crescente por abertura e transparência nos processos científicos, somado à valorização da cultura colaborativa e do acesso aos dados de pesquisa, tem se refletido nas políticas editoriais de diversos periódicos. Essas iniciativas resgatam princípios fundamentais da integridade científica, como a verificabilidade e a reprodutibilidade.

Segundo Sayão e Sales (2014) os dados e conjuntos de dados de pesquisa fornecem as evidências essenciais para assegurar a veracidade, autenticidade e reprodutibilidade do conhecimento científico publicado, sendo fundamentais para o avanço da ciência.

A prática de depósito ou publicação dos dados produzidos em pesquisas é relativamente recente, mas já está presente em diferentes fluxos de comunicação científica. Contudo, ainda carece de sistematização para que seja plenamente reconhecida como conduta científica legítima.

No Brasil, não há uma legislação nacional específica que regule o depósito de dados de pesquisa. O que se observa são iniciativas isoladas, predominantemente em instituições públicas, voltadas à formulação de diretrizes próprias para organização, preservação e disseminação dos dados científicos. Essa ausência de políticas unificadas

dificulta a consolidação de normas padronizadas e limita o avanço de uma cultura de dados abertos no cenário científico nacional.

A questão que motivou esta pesquisa centra-se na lacuna que evidencia como a inexistência de políticas nacionais unificadas e a consequente fragmentação das iniciativas de depósito de dados de pesquisa no Brasil afetam a visibilidade, a interoperabilidade e a consolidação de uma cultura de dados abertos no país.

Neste contexto, o presente estudo propõe-se a analisar os conjuntos de dados de pesquisa publicados por autores brasileiros, vinculados a instituições nacionais, e indexados na plataforma *OpenAlex*. Busca-se identificar suas principais características, incluindo os repositórios utilizados, o tipo de acesso adotado, as áreas do conhecimento envolvidas e as instituições de afiliação dos autores.

A escolha da *OpenAlex* justifica-se por seu caráter totalmente aberto (open data, open API, open *source code*), com base nos princípios FAIR (*Findability, Accessibility, Interoperability e Reusability*). A plataforma disponibiliza metadados abrangentes sobre autores, instituições, publicações, dados e conceitos científicos, em conformidade com os padrões de ciência aberta (Neubert *et al.*, 2024, p.3).

Ao observar a produção científica brasileira sob a lente da *OpenAlex*, este estudo pretende evidenciar a importância do compartilhamento de dados de pesquisa para o avanço científico. Busca-se compreender o cenário atual, identificar desafios e oportunidades, e apontar caminhos para o desenvolvimento de políticas que incentivem a adesão de pesquisadores e instituições às práticas abertas de gestão e disseminação de dados. Espera-se, assim, contribuir para o fortalecimento de uma cultura de transparência, colaboração e inovação na ciência brasileira.

2 DADOS DE PESQUISA E POLÍTICAS DE ABERTURA CIENTÍFICA

Via de regra, a comunicação das descobertas e intentos científicos acontecem de maneira descritiva por meio de artigos convencionais. A matéria prima da pesquisa, ou seja, o(s) conjunto(s) de dados produzidos, por muito tempo ficou oculta e pouco acessível aos pares, o que por vezes esterilizou o potencial para colaboração científica e até mesmo a internacionalização da ciência (Pinto, 2022).

No atual cenário científico, os dados de pesquisa assumem um papel central na geração de conhecimento, impulsionados pelo avanço da tecnologia digital e pelo surgimento

de novos modelos de investigação científica. Como destacam Sayão e Sales (2014), a relevância dos dados em contextos de “*Big Science*”, como astronomia, física e biologia, somada à intensificação da colaboração em escala global, deu origem ao chamado quarto paradigma científico ou *e-Science*, caracterizado por processos de pesquisa orientados por dados e pela emergência de campos como a bioinformática e a astroinformática.

Nesse contexto, a noção de *Big Data* amplia-se para além do volume, velocidade e variedade incluindo também a veracidade e o valor dos dados gerados em diferentes áreas, exigindo novas práticas de curadoria digital e gestão da informação.

Ao mesmo tempo, a ideia de “cauda longa” da ciência, cunhado por Wyborn e Lehnert (2016), ressaltam o valor de conjuntos de dados menos utilizados ou provenientes de pesquisas de menor escala, mas que, quando preservados e disponibilizados em repositórios digitais, podem ser reutilizados, reinterpretados e gerar novos conhecimentos. Assim, a longevidade, a interoperabilidade e o acesso aberto aos dados tornam-se elementos essenciais para a transparência, a reprodutibilidade e a inovação no ecossistema científico contemporâneo.

Sayão e Sales (2019) complementam a noção de cauda longa ao afirmarem que “os dados da cauda longa representam a maior parcela de dados produzida pela ciência e constituem um território de constante criatividade e inovação que precisam ser revelados, integrados e compartilhados.” (2019, p. 6). Ao serem dispostos em “vitrines”, o conjunto de dados ganha potencial mobilidade, podendo ser reaproveitado em outras pesquisas que tragam nova utilidade científica.

2.1 Contexto internacional das políticas de dados de pesquisa

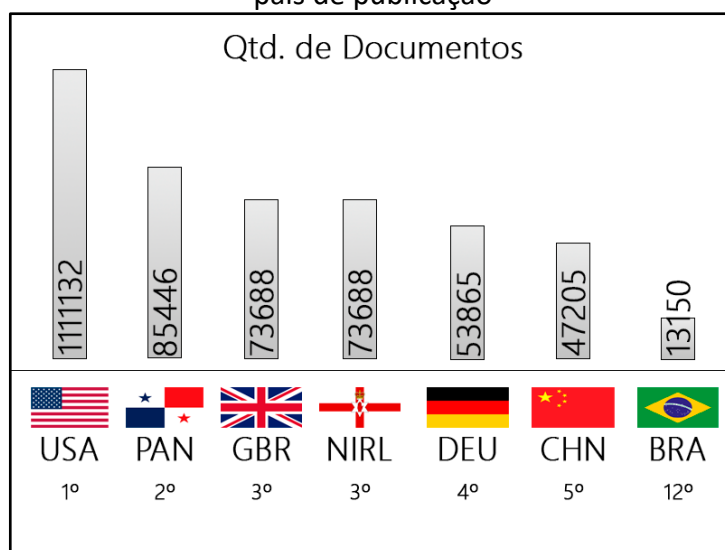
Nos países do Norte Global, a abertura dos dados de pesquisa constitui um dos pilares das políticas de ciência aberta, sendo cada vez mais incorporada por agências de fomento, universidades e instituições de pesquisa. A União Europeia tem liderado esse movimento por meio do programa *Horizon Europe*, que exige a elaboração de Planos de Gestão de Dados (*Data Management Plans – DMPs*) e o depósito dos dados em conformidade com os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*). Complementarmente, investe na criação de uma infraestrutura federada de dados por meio da *European Open Science Cloud (European Commission, 2021; EOSC Association, 2023)*.

Nos Estados Unidos, o memorando presidencial conhecido como “Nelson Memo” (*White House OSTP, 2022*) determina que, até 2026, todos os resultados de pesquisas financiadas por agências federais – incluindo os dados subjacentes – estejam disponíveis em acesso aberto. Iniciativas semelhantes são observadas no Reino Unido, por meio da *UK Research and Innovation (UKRI, 2016)*; no Canadá, com a *Tri-Agency Research Data Management Policy (Government of Canada, 2021)*; e na Austrália, com o apoio da *Australian Research Data Commons (ARDC, 2023)*.

Essas políticas têm promovido avanços importantes na governança dos dados científicos, incentivando o uso de repositórios confiáveis, a capacitação de pesquisadores e a valorização dos dados como parte da avaliação da produção acadêmica. Entretanto, mesmo em contextos com infraestrutura consolidada, persistem desafios como a padronização de metadados, a proteção de dados sensíveis e a equidade entre áreas do conhecimento (OECD, 2021; *Science Europe, 2018*). Ainda assim, observa-se uma convergência normativa no Norte Global, que legitima a abertura dos dados como prática científica estratégica.

O Gráfico 1 apresenta os depósitos de conjuntos de dados indexados na *OpenAlex* nos últimos cinco anos, com destaque para os cinco países mais bem posicionados no período. Os Estados Unidos lideram com mais de 1 milhão de depósitos, seguidos por Panamá (85.446), Reino Unido e Irlanda do Norte (73.688), Alemanha (53.865) e China (47.205). O Brasil figura na 12ª posição, com 13.150 conjuntos de dados registrados.

Gráfico 1 – Número de depósitos de conjuntos de Dados Indexados na *OpenAlex*, por país de publicação



Fonte: Elaborado pelas autoras com dados da *OpenAlex* (2025).

Apesar de sua colocação no ranking global, o Brasil apresenta uma trajetória emergente no campo da gestão de dados científicos. Essa posição não deve ser interpretada apenas como um reflexo de limitações, mas também como uma oportunidade para fortalecer e consolidar políticas institucionais e iniciativas estruturantes voltadas à publicação de dados de pesquisa.

2.2 Iniciativas brasileiras para a gestão e publicação de dados

Embora o Brasil ocupe uma posição ainda modesta na indexação de dados na *OpenAlex*, algumas iniciativas nacionais sinalizam avanços em direção à consolidação de uma cultura de dados abertos.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), ao exigir o plano de gestão de dados de pesquisa para projetos financiados pela fundação, corrobora para o fortalecimento de práticas e normativas de ciência aberta e gestão de dados de pesquisa. Segundo Pierro (2022) “Trata-se de uma iniciativa pioneira no país ao estabelecer políticas e diretrizes para o gerenciamento de dados científicos”.

Outra iniciativa relevante é o Consórcio Nacional para Ciência Aberta (CoNCienciA), criado em 2021 por meio da colaboração entre CNPq, RNP, IBICT, Fiocruz, Embrapa e CBPF. O consórcio tem como objetivo promover a criação e a interoperabilidade de repositórios de dados de pesquisa, além de incentivar a atribuição de identificadores persistentes (DOIs), alinhando o Brasil às práticas internacionais de ciência aberta (RNP, 2024).

Em abril de 2024, o Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT tornou-se signatário da Declaração de Barcelona sobre Informação Aberta de Pesquisa, um compromisso internacional que visa promover a abertura das informações relacionadas à condução e comunicação da pesquisa científica. A adesão do IBICT destaca o alinhamento do Brasil com os princípios da Ciência Aberta, buscando fortalecer infra estruturas e práticas que garantam maior transparência, acessibilidade e reprodutibilidade na produção científica. A Declaração propõe que informações como metadados bibliográficos, dados sobre financiamento, software, amostras e instrumentos de pesquisa sejam abertas por padrão, rompendo com modelos baseados em plataformas proprietárias e comerciais. Com mais de 50 instituições signatárias, incluindo universidades e organizações internacionais, o documento estabelece compromissos como apoiar infraestruturas abertas, adotar padrões de

interoperabilidade e fomentar ações coletivas para acelerar a transição para um ecossistema científico mais justo e inclusivo. (IBICT, 2024).

Entre as ferramentas institucionais já operacionais, destaca-se o repositório Deposita Dados¹, também coordenado pelo IBICT. Ele oferece uma opção para pesquisadores brasileiros vinculados a instituições que ainda não possuem repositórios próprios, ou para aqueles que colaboram com centros de pesquisa estrangeiros. O sistema é regionalizado e requer autorização prévia da equipe para realização do depósito.

Outro passo significativo em direção ao planejamento e ações de natureza tecnológica para depósito, preservação e acesso a dados de pesquisa nacional é a ferramenta Metabuscaador de Dados de Pesquisa², que consiste em um agregador de repositórios institucionais no Estado de São Paulo, iniciativa da FAPESP.

A plataforma *SciELO Data*³ também merece destaque como repositório multidisciplinar voltado à preservação e disseminação de dados associados a artigos da Rede SciELO e ao SciELO *Preprints*. Essa iniciativa reforça o compromisso da comunidade científica latino-americana com a ciência aberta e promove maior consciência sobre o papel dos dados na integridade e reprodutibilidade das pesquisas.

O cenário brasileiro conta com uma variedade de repositórios institucionais dedicados ao depósito e disseminação de dados de pesquisa. A Tabela 1 apresenta um panorama quantitativo de algumas dessas iniciativas, com destaque para o ARCA Dados (Fiocruz) e o OasisBR, que reúnem, respectivamente, 194.825 e 173.029 itens.

Tabela 1 – Repositórios nacionais de dados de pesquisa e número de itens depositados

Repositórios	Itens depositados
LattesData	299
Deposita Dados - Ibict	300
Metabuscaador de Dados de Pesquisa - FAPESP	2.095
SciELO Data	6.015
OasisBR	173.029

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

² Disponível em: <https://metabuscaador.uspdigital.usp.br/>

³ Disponível em: <https://data.scielo.org/>

ARCA Dados - Fiocruz	194.825
----------------------	---------

Fonte: Elaborado pelas autoras (2025).

Outra iniciativa que merece destaque no contexto das ações em prol da publicação e depósito de dados de pesquisa é o repositório LattesData⁴ que tem como objetivo reunir, armazenar e divulgar os conjuntos de dados científicos de pesquisadores beneficiários do CNPq. Futuramente, será também disponível para depósito de dados de pesquisa oriundos de pesquisadores não financiados pelo CNPq, mediante acordos celebrados com suas instituições no âmbito do CoNCiência.

2.3 Metodologia

A pesquisa adota uma abordagem quali-quantitativa, de natureza descritivo-exploratória. A análise qualitativa envolveu a interpretação dos resultados obtidos, relacionando-os ao contexto nacional e internacional das políticas de ciência aberta e de gestão de dados de pesquisa. Essa análise interpretativa permitiu compreender padrões de depósito, identificar lacunas de interoperabilidade e apontar implicações para a visibilidade e o reuso dos dados no cenário científico global.

A pesquisa quantitativa baseou-se na base de dados *OpenAlex* que foi utilizada como fonte principal para a coleta e análise dos dados. A busca exploratória baseou-se na aplicação de filtros sobre os metadados disponíveis na plataforma, especificamente: *Type*, para identificar o tipo de material (neste caso, *datasets*); *Year*, para delimitar o período de interesse (2019 a 2024); e *Country*, para selecionar o país de afiliação institucional dos autores (Brasil). Foram incluídos na análise os conjuntos de dados (*datasets*) publicados nos últimos cinco anos, cuja autoria está vinculada a instituições brasileiras.

O recorte temporal escolhido se justifica por esse período abranger os anos mais recentes e representativos da consolidação de políticas e iniciativas voltadas à ciência aberta e à gestão de dados de pesquisa, tanto no cenário internacional quanto no Brasil. Além da fase de expansão e reconhecimento da *OpenAlex* e por fim, o período inclui avanços institucionais relevantes, que refletem mudanças significativas no ecossistema brasileiro de dados científicos.

⁴ Disponível em: <https://lattesdata.cnpq.br/>

Foram identificados 13.150 conjuntos de dados indexados na *OpenAlex*, no período analisado. Para extração dos dados utilizou-se a ferramenta de exportação de dados disponibilizada pela própria interface da *OpenAlex*.

Após coleta e seleção dos conjuntos de dados elegíveis para o estudo, foi necessário a exclusão de alguns dados por inconsistências nas informações, o que poderia impactar a análise, como por exemplo, metadados em branco, chegando ao resultado de 9.652 conjuntos de dados aptos para análise, aproximadamente (73% do total).

As ferramentas utilizadas incluíram o exportador da própria interface da *OpenAlex*, o software Python (para organização e filtragem), o Microsoft Excel (para tabulação e análise quantitativa) e o *VOSviewer*, utilizado para a construção de visualizações gráficas de coautoria.

2.4 Resultados e análise dos dados

A análise revelou que 100% dos conjuntos de dados brasileiros indexados na *OpenAlex* estão depositados em repositórios digitais, sendo a maioria hospedada em plataformas internacionais, como *Figshare* e *Zenodo*, conforme ilustrado no Gráfico 2. Este apresenta os cinco repositórios mais utilizados para depósito dos conjuntos de dados por autores brasileiros. As plataformas predominantes são de abrangência global, o que reforça a inserção internacional dos dados brasileiros, embora também aponte para a ausência de maior representação dos repositórios nacionais nos sistemas de indexação da *OpenAlex*.

Ainda segundo Nicodemo (2025?) projetos de autores brasileiros frequentemente publicam seus dados em repositórios internacionais, e ressalta a importância de fortalecer a infraestrutura nacional para evitar a dependência de plataformas estrangeiras e garantir a soberania do conhecimento científico.

Gráfico 2 – Repositórios utilizados para o depósito de dados

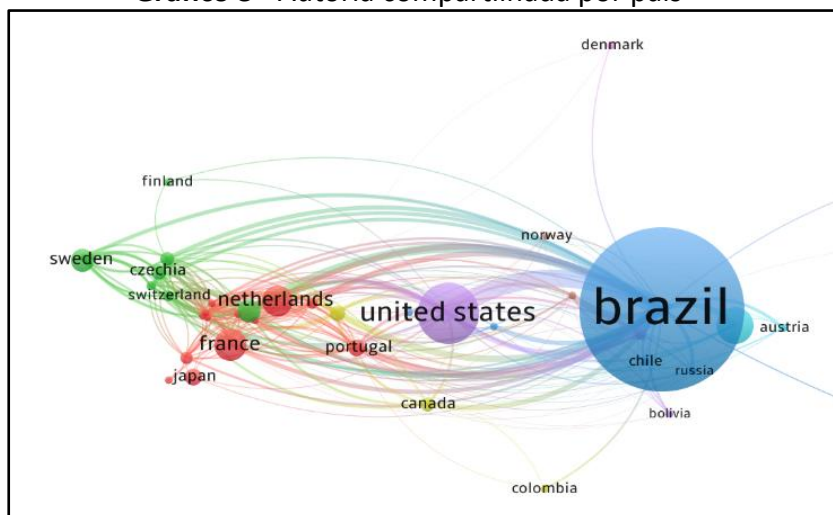


Fonte: Elaborado pelas autoras com dados da *OpenAlex* (2025).

Além da predominância de depósitos em repositórios internacionais, observa-se no Gráfico 3, um volume expressivo de autoria compartilhada nos conjuntos de dados analisados. A partir da análise dos metadados, identificou-se que aproximadamente 37% dos registros envolvem coautoria com pesquisadores estrangeiros, refletindo a presença de redes de colaboração internacional. Dentre os países parceiros, os Estados Unidos lideram as associações, seguidos por outras nações do Norte Global.

A presença recorrente de múltiplas nacionalidades nos registros evidencia não apenas a internacionalização da ciência brasileira, mas também a potencialização do reuso e da visibilidade dos dados compartilhados por pesquisadores brasileiros.

Gráfico 3 - Autoria compartilhada por país



Fonte: Elaborado pelas autoras com dados da *OpenAlex* (2025).

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Aproximadamente 100% dos conjuntos de dados identificados no estudo estão disponíveis em acesso aberto⁵, conforme a classificação da *OpenAlex*. Dentre os 9.652 conjuntos válidos, 9.632 estão com status de acesso livre.

Dentre os *status*⁶ de acesso aberto classificados pela *OpenAlex*, o tipo de maior prevalência é o *gold*, que totalizam 7.698 registros (80%). Essa predominância indica uma tendência de alinhamento às políticas editoriais que exigem ou incentivam a abertura dos dados vinculados às publicações.

A série histórica entre 2019 e 2024 aponta para um crescimento no número de depósitos ao longo dos anos, com pico em 2023 (3.653 conjuntos de dados), seguido por 2020 (2.954). Chama a atenção o decréscimo registrado em 2024 (506 registros) até o momento da coleta, o que pode estar relacionado a fatores como o atraso na indexação ou o aumento de depósitos em repositórios institucionais brasileiros ainda não integrados à *OpenAlex*.

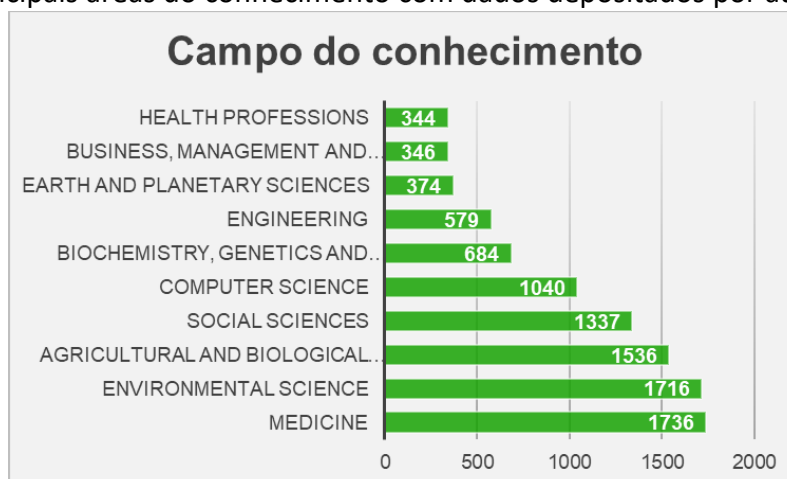
Segundo os metadados obtidos, os conjuntos de dados analisados acumulam cerca de 21 mil citações, demonstrando não apenas sua visibilidade, mas também seu potencial de reuso. Este número reforça a importância de políticas institucionais voltadas à gestão qualificada e disseminação estratégica de dados, como forma de ampliar o impacto da produção científica.

De acordo com a classificação da *OpenAlex*, as áreas do conhecimento com maior incidência de conjuntos de dados depositados por autores brasileiros incluem as Ciências da Saúde (com destaque para a Medicina), as Ciências Ambientais, Ciências Agrárias e Biológicas, e a Ciência da Computação, conforme representação no Gráfico 4.

⁵A *OpenAlex* não exige acesso aberto ao conteúdo completo, apenas metadados estruturados e acessíveis. Assim, conjuntos de dados com acesso restrito, sob embargo ou mediante solicitação também podem ser indexados, desde que vinculados a produções científicas.

⁶AA Status: diamond: Publicado em um periódico totalmente AA: indexado pelo DOAJ ou sem custos de processamento de artigos (ou seja, gratuito para leitores e autores). gold: Publicado em um periódico totalmente AA. green: Acesso pago na página inicial do editor, mas há uma cópia gratuita em um repositório. hybrid: Gratuito sob uma licença aberta em um periódico de acesso pago. bronze: Leitura gratuita na página inicial da editora, mas sem nenhuma licença identificável. closed: Todos os outros artigos. Fonte: *OpenAlex*, 2025.

Gráfico 4 – Principais áreas do conhecimento com dados depositados por autores brasileiros

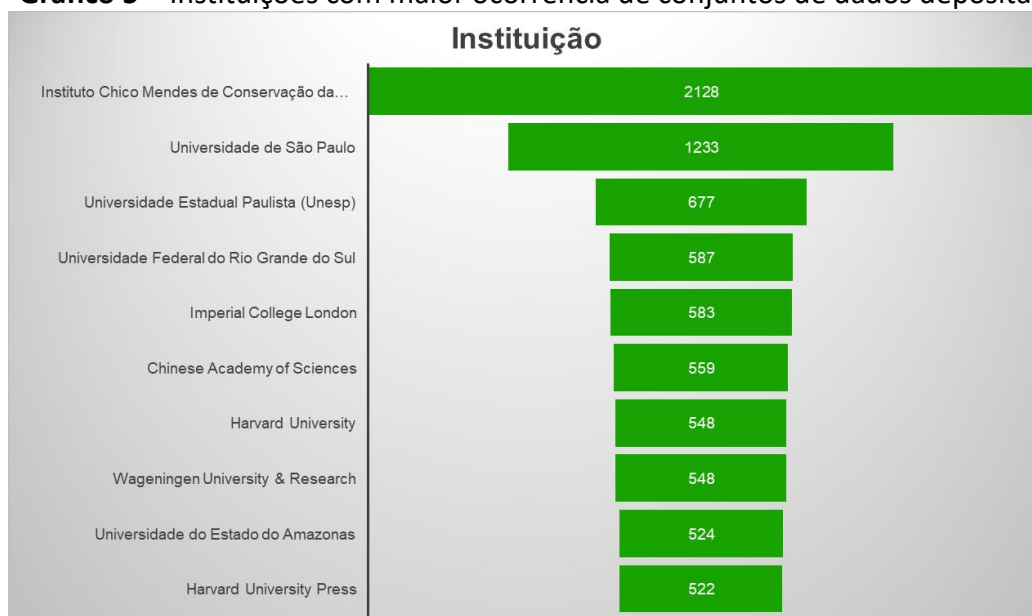


Fonte: Elaborado pelas autoras com dados da OpenAlex (2025).

Esse padrão pode ser atribuído à natureza das pesquisas conduzidas nessas áreas, que frequentemente lidam com temas de alto impacto social, como saúde pública, sustentabilidade ambiental e inovação tecnológica. Nessas áreas, há uma maior exigência por transparência, reprodutibilidade e compartilhamento de dados, o que pode justificar o engajamento mais expressivo com práticas de ciência aberta.

As instituições nacionais com maior volume de conjuntos de dados indexados na *OpenAlex* foram o Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), a Universidade de São Paulo (USP), a Universidade Estadual Paulista (UNESP), Universidade Federal do Rio Grande do Sul (UFRGS), conforme ilustrado no Gráfico 5.

Gráfico 5 – Instituições com maior ocorrência de conjuntos de dados depositados



Fonte: Elaborado pelas autoras com dados da *OpenAlex* (2025).

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

O protagonismo do ICMBio na indexação de dados na *OpenAlex* pode ser atribuído a iniciativas como o SALVE – Sistema de Avaliação do Estado de Conservação da Biodiversidade⁷, que organiza e disponibiliza informações sobre espécies ameaçadas com base em critérios da *International Union for Conservation of Nature* (UICN). A política institucional de abertura e a padronização dos dados ambientais reforçam a visibilidade da produção da instituição e seu alinhamento com os princípios da ciência aberta.

A forte presença de Instituições do Estado de São Paulo, com destaque entre as Instituições com maior número de depósitos, fato que pode ser atribuído às Universidades pertencerem ao primeiro Estado nacional a exigir Plano de Gestão de Dados e seus posteriormente seus depósitos de dados.

Identificou-se também conjuntos de dados com autoria vinculada a instituições estrangeiras, como a *Imperial College London*, a *Chinese Academy of Sciences* e a *Harvard University*. Essa internacionalização dos dados está associada à presença de pesquisadores brasileiros em colaborações internacionais, evidenciando a importância de redes globais para a disseminação dos dados de pesquisa.

A análise revela que as instituições brasileiras que se destacam nesse cenário são, em sua maioria, aquelas que já implementaram normativas ou diretrizes institucionais voltadas à gestão e publicação de dados de pesquisa. Essa associação é ilustrada no Quadro 1, que reúne os documentos institucionais identificados na pesquisa.

Quadro 1 – Normativas e diretrizes para dados de pesquisa em instituições brasileiras

Instituição	Documento institucional
ICMBio - Instituto Chico Mendes de Conservação da Biodiversidade	Normativa nº 2, de 25 de novembro de 2015 - Política de Dados de Pesquisa do ICMBio
USP - Universidade de São Paulo	Resolução nº 7.900/2019 e Portaria PRP/PRPG/STI/AGUIA nº 001/2019
UNESP - Universidade Estadual Paulista "Júlio de Mesquita Filho"	Resolução Unesp nº 08, de 04/02/2025
UFRGS - Universidade do Rio Grande do Sul	Guia para Integridade em Pesquisa Científica

Fonte: Elaborado pelas autoras (2025).

Se compararmos os dados de pesquisa indexados na *OpenAlex* com os dados de pesquisas depositados em outras iniciativas institucionais nacionais, como as mencionadas anteriormente na Tabela 1, podemos identificar que só o ARCA Dados, que é o repositório

⁷ Disponível em: <https://salve.icmbio.gov.br/>

institucional de dados de pesquisa da Fundação Oswaldo Cruz - FioCruz, possui mais itens depositados do que os indexados na *OpenAlex*. Considera-se que tal constatação seja consequência de uma infraestrutura institucional que prioriza a abertura dos dados de pesquisa, institucionalizando uma política de acesso aberto e a conscientização dos pesquisadores para a importância desta prática.

3 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi investigar o panorama do depósito de conjuntos de dados de pesquisa realizados por autores vinculados a instituições brasileiras, indexados na plataforma *OpenAlex*, no contexto das práticas de Ciência Aberta.

Os resultados revelam uma predominância do uso de repositórios internacionais para o depósito de dados por parte da comunidade científica brasileira. Esse padrão parece refletir a ausência de diretrizes nacionais consolidadas, bem como o estágio ainda incipiente de implementação de políticas institucionais voltadas à gestão e disseminação de dados de pesquisa. Tais fatores indicam que o país se encontra em um processo de transição e construção de uma cultura institucional em dados abertos.

No que se refere às áreas do conhecimento que já publicam dados de pesquisa em repositórios, revela-se o protagonismo de algumas áreas em detrimento de outras. Tal observação sugere iniciativas para sensibilização e transformação da cultura científica em prol da adoção de práticas em ciência aberta de forma mais ampla.

Outro achado relevante refere-se à interoperabilidade dos sistemas de indexação. Embora diversas instituições brasileiras estejam avançando na criação e consolidação de repositórios institucionais, muitos desses ambientes — como o ARCA Dados (FioCruz), o SciELO Data e o Metabuscador FAPESP — ainda não estão integrados à *OpenAlex*, o que limita sua visibilidade internacional. A ausência de mecanismos técnicos que favoreçam essa integração compromete o alcance global da produção científica nacional e restringe o potencial de reuso e impacto dos dados produzidos no país.

No que tange às instituições com maior volume de dados indexados, observa-se uma correlação entre a existência de normativas específicas e o desempenho no depósito de dados, bem como a participação em colaborações internacionais. Isso evidencia que a estrutura institucional e o engajamento com redes globais são fatores decisivos para o fortalecimento das práticas de ciência aberta.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Em síntese, embora o Brasil apresente avanços importantes na constituição de repositórios e políticas voltadas à gestão de dados de pesquisa, ainda persiste um descompasso entre os esforços nacionais e a integração com os sistemas internacionais de indexação. Superar essa lacuna demanda a adoção de estratégias coordenadas de interoperabilidade, o fortalecimento das políticas institucionais e o reconhecimento da ciência aberta como um vetor estratégico de visibilidade, colaboração e inovação.

REFERÊNCIAS

ALBAGLI, Sarita. Ciência aberta como instrumento de democratização do saber. EDITORIAL. **Trab. educ. saúde** v. 15, n. 3, 2017. Disponível em: <https://doi.org/10.1590/1981-7746-sol00093>. Acesso em: set, 2024.

AUSTRALIAN RESEARCH DATA COMMONS. **Strategic Plan: 2023–2028**. 2023. Disponível em: <https://ardc.edu.au>. Acesso em: 9 maio 2025.

EOSC ASSOCIATION. **Strategic Research and Innovation Agenda (SRIA)**. 2023. Disponível em: <https://www.eosc.eu>. Acesso em: 9 maio 2025.

EUROPEAN COMMISSION. **Horizon Europe – The EU Research and Innovation Programme (2021–2027): Programme Guide**. 2021. Disponível em: <https://ec.europa.eu/programmes/horizon2020/>. Acesso em: 9 maio 2025.

GOVERNMENT OF CANADA. **Tri-Agency Research Data Management Policy**. 2021. Disponível em: <https://science.gc.ca>. Acesso em: 9 maio 2025.

INSTITUTO BRASILEIRO DE INFORMAÇÃO E TECNOLOGIA. **Ibict assina Declaração de Barcelona sobre Informação Aberta de Pesquisa**. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2024/abril/ibict-assina-declaracao-de-barcelona-sobre-informacao-aberta-de-pesquisa>. Acesso em: 25 maio. 2025.

INSTITUTO BRASILEIRO DE INFORMAÇÃO E TECNOLOGIA. **Sobre o Oasisbr**. Disponível em: <https://www.gov.br/ibict/pt-br/aceso-a-informacao/perguntas-frequentes/sobre-o-oasisbr>. Acesso em: 9 maio 2025.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **METABUSCADOR de dados de pesquisa FAPESP: Home**. Disponível em: <https://metabuscador.uspdigital.usp.br/>. Acesso em: 25 maio 2025.

FUNDAÇÃO OSWALDO CRUZ. **Arca Dados é o repositório oficial de dados para pesquisa da Fiocruz**. 2023. Disponível em: <https://fiocruz.br/noticia/2023/09/arca-dados-e-o-repositorio-oficial-de-dados-para-pesquisa-da-fiocruz>. Acesso em: 9 maio 2025.

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE. **SALVE: Sistema de Avaliação do Estado de Conservação da Biodiversidade**. Brasília: ICMBio, [2024?]. Disponível

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

em: <https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/fauna-brasileira/salve>.
Acesso em: 20 maio 2025.

NICODEMO, Thiago Lima. Dados abertos de pesquisa no Brasil: diagnóstico e perspectivas futuras. **Ciência & Cultura**, [2025?]. Disponível em: <https://revistacienciaecultura.org.br/?artigos=dados-abertos-de-pesquisa-no-brasil-diagnostico-e-perspectivas-futuras>. Acesso em: 6 de Agosto de 2025.

NEUBERT, Patrícia; *et al.* OpenAlex como fonte de dados para sistemas nacionais de informação científica: a experiência do projeto Laguna. *In: VII Workshop de Informação, Dados e Tecnologia - WIDaT 2024*, [S. l.], v. 7, p. e184, 2024. Disponível em: <https://doi.org/10.22477/vii.widat.184>. Acesso em: 22 maio 2025.

OPENALEX. **The open catalog to the global research system**. 2025. Disponível em: <https://openalex.org/>. Acesso em: 9 maio 2025.

OPENALEX. **OpenAlex technical documentation**. 2025. Disponível em: <https://docs.openalex.org/>. Acesso em: 10 Agosto 2025.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Recommendation of the Council on Access to Research Data from Public Funding**. Paris: OECD Publishing, 2021. Disponível em: <https://www.oecd.org/sti/inno/recommendation-access-to-research-data.htm>. Acesso em: 9 maio 2025.

PIERRO, Bruno de. Uma estratégia para dados: pesquisadores são estimulados a gerenciar e compartilhar as informações científicas que produzem. **Revista Pesquisa Fapesp**. ed. 267, 25 jul 2022. Disponível em: <https://revistapesquisa.fapesp.br/uma-estrategia-para-dados/>. Acesso em: 24 maio 2025.

PINTO, Fabiana de Melo Amaral Gonçalves. **Publicação de dados de pesquisa em Enfermagem: novas práticas em Ciência Aberta**. 2022. 112 f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2022. Disponível em: <http://app.uff.br/riuff/handle/1/26650>. Acesso em: 5 ago. 2025.

REDE BRASILEIRA PARA EDUCAÇÃO E PESQUISA. **RNP e Ibict lançam o Núcleo de Dados de Pesquisa da Rede Brasileira de Repositórios Digitais**. Disponível em: <https://www.rnp.br/pesquisa-e-desenvolvimento/rnp-e-ibict-lancam-o-nucleo-de-dados-de-pesquisa-da-rede-brasileira-de-repositorios-digitais/>. Acesso em: 25 maio 2025.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS**, [S. l.], v. 8, n. 2, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 5 ago. 2025.

SALES, Luana Farias; SAYÃO, Luís Fernando. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA*

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

DA INFORMAÇÃO, 19., Londrina, 2018. **Anais [...]**. Londrina: UEL/ANCIB, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103678>. Acesso em: 5 ago. 2025.

SCIENCE EUROPE. **Practical Guide to the International Alignment of Research Data Management**. Bruxelas: Science Europe, 2018. Disponível em: <https://www.scienceeurope.org>. Acesso em: 9 maio 2025.

UK RESEARCH AND INNOVATION. **UKRI Research Data Policy**. 2016. Disponível em: <https://www.ukri.org>. Acesso em: 9 maio 2025.

WHITE HOUSE. **Office of Science and Technology Policy**. Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. 2022. Disponível em: <https://www.whitehouse.gov>. Acesso em: 9 maio 2025.

WYBORN, Lesley; LEHNERT, Kerstin. Exploiting the long tail of scientific data: Making small data BIG. *In*: ERESEARCH AUSTRALASIA CONFERENCE, Melbourne, Australia, 10-14 Oct. 2016. **Anais [...]**. Melbourne, Australia, 2016. Disponível em: https://eresearchau.files.wordpress.com/2016/03/eresau2016_paper_88.pdf. Acesso em: 5 ago. 2025.