

## Pecha kucha

---

# Interoperabilidade entre Repositórios Científicos e OpenAlex: Desafios e Estratégias

*Interoperability between Scientific Repositories and OpenAlex:  
Challenges and Strategies*

*Interoperabilidad entre Repositorios Científicos y OpenAlex: Desafíos y Estrategias*

**Washington L. R. Carvalho Segundo \***

Doutorado em Informática

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

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

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

E-mail: [washingtonsegundo@ibict.br](mailto:washingtonsegundo@ibict.br)

**Marcel Garcia de Souza**

Mestrado em Educação em Ciências – Química da Vida e Saúde

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

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

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

E-mail: [marcelsouza@ibict.br](mailto:marcelsouza@ibict.br)

**Thiago Magela Rodrigues Dias**

Doutorado em Modelagem Matemática e Computacional

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

ORCID: <https://orcid.org/0000-0001-5057-9936>

Lattes: <http://lattes.cnpq.br/4687858846001290>

E-mail: [thiagomagela@cefetmg.br](mailto:thiagomagela@cefetmg.br)

**Fábio Lorensi do Canto**

Doutorado em Ciência da Informação

Universidade Federal de Santa Catarina (UFSC)

ORCID: <https://orcid.org/0000-0002-8338-1931>

Lattes: <http://lattes.cnpq.br/5914776544385758>

E-mail: [fabio.lc@ufsc.br](mailto:fabio.lc@ufsc.br)

## Resumo

Este trabalho avalia a representatividade de 199.446 publicações científicas brasileiras presentes, simultaneamente, nos repositórios indexados pelo Oasisbr e na OpenAlex, considerando o período de 2000 a 2024. Observou-se, ao longo dos anos, uma tendência de crescimento no número de publicações, com exceção dos períodos 2014-2015 e 2022-2024 – neste último caso, em razão do intervalo entre publicação, depósito e indexação. Cerca de 7% das publicações (13.976 artigos) não possuem o Digital Object Identifier (DOI), evidenciando a necessidade de estratégias para gerar identificadores persistentes, especialmente em áreas como medicina, agronomia e educação. A integração dessas bases de dados e a adoção de identificadores persistentes são fundamentais para enriquecer os repositórios locais, ampliar a visibilidade da ciência brasileira e fortalecer a interoperabilidade e a comunicação científica, em alinhamento com os princípios da ciência aberta.

**Palavras-chave:** Integração de Dados; Repositórios de Dados; Identificadores Persistentes; Oasisbr; OpenAlex.

## Abstract

This study evaluates the representativeness of 199,446 Brazilian scientific publications present simultaneously in repositories indexed by Oasisbr and OpenAlex, covering the period from 2000 to 2024. A growing trend in the number of publications was observed over the years, except for the 2014-2015 and 2022-2024 periods—in the latter case, due to the time lag between publication, deposit, and indexing. Approximately 7% of the publications (13,976 articles) do not have a Digital Object Identifier (DOI), which highlights the need for strategies to generate persistent identifiers, especially in fields such as medicine, agronomy, and education. The integration of these databases and the adoption of persistent identifiers are essential to enrich local repositories, increase the visibility of Brazilian science, and strengthen interoperability and scientific communication in alignment with the principles of open science.

**Keywords:** Data Integration, Data Repositories, Persistent Identifiers, Oasisbr, OpenAlex.

## Resumen

Este estudio evalúa la representatividad de 199.446 publicaciones científicas brasileñas presentes simultáneamente en repositórios indexados por Oasisbr y en OpenAlex, considerando el período de 2000 a 2024. Se observó una tendencia de aumento en el número de publicaciones a lo largo de los años, excepto entre 2014 y 2015, y 2022-2024 —en este último caso, debido al intervalo entre la publicación, el depósito y la indexación. Aproximadamente el 7% de las publicaciones (13.976 artículos) no cuentan con un Identificador de Objeto Digital (DOI), lo que pone de manifiesto la necesidad de establecer estrategias para generar identificadores persistentes, especialmente en áreas como la medicina, la agronomía y la educación. La integración de estas bases de datos y la adopción de identificadores persistentes son esenciales para enriquecer los repositórios locales, aumentar la visibilidad de la ciencia brasileña y fortalecer la interoperabilidad y la comunicación científica, en consonancia con los principios de la ciencia abierta.

**Palabras clave:** Integración de Datos, Repositórios de Datos, Identificadores Persistentes, Oasisbr, OpenAlex.

## Introdução

O avanço científico está, intrinsecamente, ligado à disseminação aberta, eficiente e acessível do conhecimento produzido por pesquisadores e instituições em escala mundial. No Brasil, iniciativas como o Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (Oasisbr) desempenham papel estratégico ao reunir e disponibilizar a produção científica nacional, promovendo sua reusabilidade, interoperabilidade e alcance global. O Oasisbr busca ampliar a visibilidade e o impacto da ciência brasileira, promovendo o intercâmbio e o alinhamento com as práticas da ciência aberta, por meio da agregação de dados de repositórios digitais, revistas científicas e bibliotecas digitais (Gibbon et al., 2023).

No contexto internacional, bases de dados de acesso aberto, como a OpenAlex, oferecem uma visão abrangente do ecossistema de pesquisa global. Lançada em 2022, a OpenAlex surge como uma alternativa às fontes comerciais, promovendo a transparência e a reutilização de dados e de informações sobre a pesquisa. A capacidade de integrar dados de múltiplas fontes, incluindo repositórios, instituições e Identificadores Persistentes (PIDs), faz da OpenAlex uma ferramenta importante para compreender padrões de publicações, colaborações e impacto científico (Priem; Piwowar; Orr, 2022).

Apesar do crescente volume de informações, a representatividade da produção brasileira, em fontes internacionais, ainda é pouco explorada. A comparação entre artigos em repositórios nacionais indexados pelo Oasisbr e os presentes em bases como a OpenAlex pode revelar lacunas e oportunidades para aprimorar a integração entre as plataformas. Além disso, o uso de PIDs, como o DOI (Digital Object Identifier), desempenha um papel estratégico na rastreabilidade e na interoperabilidade dos registros bibliográficos.

Este estudo tem por objetivo analisar a representatividade de artigos brasileiros publicados em periódicos nacionais e internacionais e depositados em repositórios indexados pelo Oasisbr, em comparação com aqueles disponíveis na OpenAlex. Por meio de uma abordagem baseada na integração e na análise de dados, busca-se compreender melhor aspectos relativos à visibilidade da produção nacional, em especial o uso de PIDs. Essa análise pode contribuir tanto para a compreensão do cenário atual como para a formulação de políticas públicas e para o desenvolvimento de tecnologias que facilitem a integração e a disseminação de dados e informações sobre a pesquisa nacional.

## Metodologia

O estudo utilizou uma abordagem quantitativa e comparativa, com dados extraídos das APIs (Application Programming Interfaces) públicas do Oasisbr e da OpenAlex. Foram considerados como artigos nacionais aqueles em que pelo menos um dos autores está vinculado a alguma instituição brasileira. Foram coletados dados de 500.134 artigos

brasileiros publicados em periódicos indexados pelo Oasisbr e 2.940.546 indexados na OpenAlex. Os metadados extraídos foram: DOI, nome dos autores e das instituições, fontes de publicação, entre outros campos.

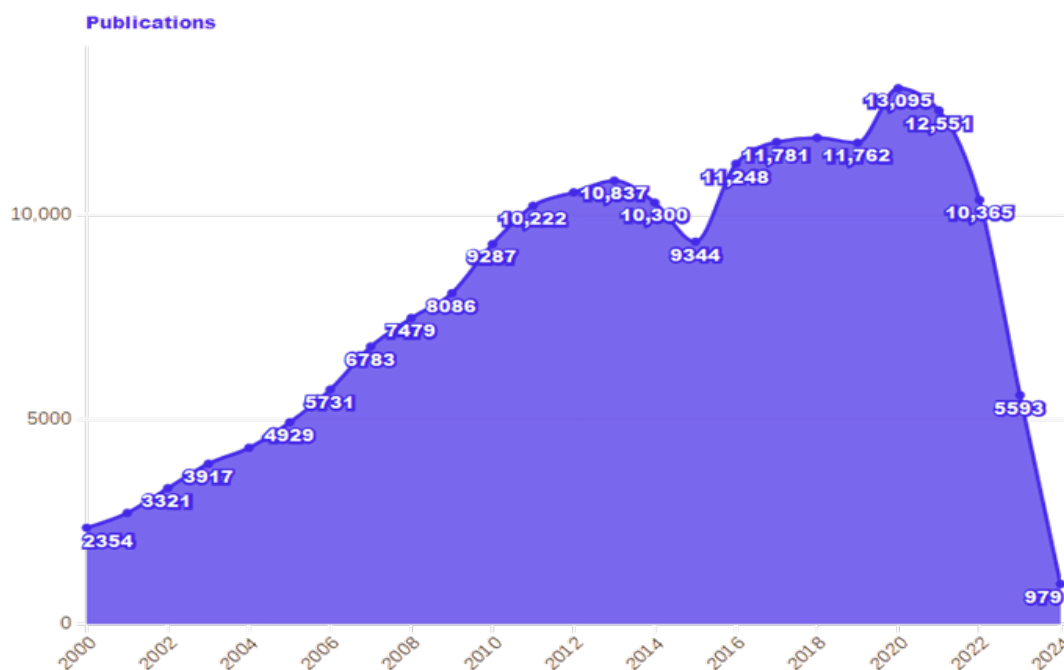
Posteriormente, os registros de DOIs extraídos do Oasisbr foram cruzados com os existentes na OpenAlex. Para os artigos que não possuíam DOI ou que não estavam registrados em uma das fontes, adotou-se a estratégia de gerar um identificador persistente, garantindo a comparabilidade dos artigos (Dias; Moita, 2015).

## Resultados

Uma vez estabelecida a correspondência entre os artigos das duas fontes, foi possível identificar aqueles que estavam presentes em ambas as bases de dados ou em apenas uma delas. Uma sumarização das publicações identificadas em ambas as bases pode ser observada na Figura 1.

Foi identificado que 199.446 artigos estavam presentes nas duas fontes. Considerando a janela temporal de 2000 a 2024, observou-se uma tendência de crescimento no número de publicações ao longo dos anos, com exceção dos anos de 2014 e 2015.

**Figura 1 – Evolução Temporal dos Artigos identificados.**



Fonte: dados da pesquisa (2025).

A partir de 2022, verificou-se, novamente, uma queda no número de publicações, o que pode ser explicado pelo intervalo entre a publicação, o depósito em repositórios e a

indexação na OpenAlex (Figura 1). Essa dinâmica reflete não apenas o comportamento das publicações, mas também os prazos para o processamento e integração de dados entre sistemas.

Constatou-se ainda que um total de 13.976 publicações (aproximadamente 7% do total analisado) não possuía registro de DOI. A identificação dessas publicações foi possível graças à utilização de elementos básicos para a geração de identificadores, conforme descrito anteriormente. Essa abordagem evidenciou a relevância de implementar estratégias que permitam gerar PIDs em publicações que ainda não os possuem, facilitando sua integração em diferentes bases de dados.

Ao analisar as áreas de conhecimento das publicações sem DOI, observa-se que algumas se destacam, como medicina, agronomia e educação. Esse resultado evidencia a necessidade de maior atenção à adoção de PIDs em áreas específicas, contribuindo para aprimorar a interoperabilidade e aumentar a visibilidade da produção científica dessas áreas.

## Conclusões

A integração entre bases de dados locais, como o Oasisbr, e agregadores globais, como o OpenAlex, mostra-se uma estratégia fundamental para enriquecer os repositórios científicos locais e ampliar sua visibilidade. Essa integração não apenas fortalece a disseminação da produção científica brasileira, mas também possibilita a inclusão de dados mais robustos e completos, favorecendo análises mais precisas e a identificação de lacunas ou oportunidades de pesquisa.

Um aspecto central para o sucesso desse processo é a adoção de PIDs, como o DOI. Esses identificadores atuam como pontos de referência para as publicações, facilitando o cruzamento de dados entre diferentes sistemas, mesmo em cenários com volumes massivos de informações. A ausência de PIDs pode comprometer, significativamente, o potencial de integração e a análise, aumentando o esforço necessário para a deduplicação e a desambiguação de dados.

Portanto, a promoção da adoção de PIDs e o desenvolvimento de estratégias automatizadas para gerar e validar esses identificadores, especialmente em bases de dados que ainda apresentam lacunas nesse aspecto, são ações importantes. Elas não apenas fortalecem a interoperabilidade entre sistemas, mas também consolidam a confiabilidade e a rastreabilidade dos dados científicos, contribuindo para o avanço da ciência aberta e da comunicação científica global.

## Agradecimentos

Os autores agradecem ao Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) pelo suporte à pesquisa.

## Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

## Disponibilização dos Dados de Investigação

<https://doi.org/10.48472/aleia/NIET4Q>

## CRedit – Contribuições dos Autores

Washington L. R. Carvalho Segundo | Escrita – redação original, Recolha de dados.

Marcel Garcia de Souza | Escrita – revisão e edição.

Thiago Magela Rodrigues Dias | Escrita – revisão e edição.

Fábio Lorensi do Canto | Escrita – revisão e edição.

## Referências

DIAS, T. M. R. & MOITA, G. F. (2015). A method for the identification of collaboration in large scientific databases. Em *Questão*, Porto Alegre, v. 21, n. 2, p. 140-161. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/53259/34340>. Acesso em: 10 mar. 2025.

GIBBON, C. A.; MIRANDA, A. C. D.; CARVALHO SEGUNDO, W. L. R.; MORAES, M. H. M. & SILVA, L. S. (2023). Ciência Aberta brasileira: uma análise a partir do Oasisbr. In: *WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA - WIDaT*, 6., 2023, Brasília, DF. Anais [...]. Brasília, DF: [s. n.]. DOI: 10.22477/vi.widat.07. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/7>. Acesso em: 15 jan. 2025.

PRIEM, J.; PIWOWAR, H. & ORR, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv. Disponível em: <https://doi.org/10.48550/arXiv.2205.01833>. Acesso em: 10 mar. 2025.