



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



ESTRATÉGIAS PARA ANÁLISE DA PRODUÇÃO DE PESQUISADORES: um estudo com Lattes, OpenAlex e Scopus

Marina Plentz

<https://orcid.org/0009-0004-4740-3920>
marina.plentz@ufrgs.br
Universidade Federal do Rio Grande do
Sul (UFRGS)

Ana Maria Mielniczuk de Moura

<https://orcid.org/0000-0002-7811-4660>
ana.moura@ufrgs.br
Universidade Federal do Rio Grande do
Sul (UFRGS)

RESUMO:

O trabalho comparou o uso das bases Scopus, OpenAlex e da Plataforma Lattes para recuperar a produção em acesso aberto de 30 pesquisadores do Instituto de Pesquisas Hidráulicas da UFRGS (2021-2025). Como resultados, enquanto o Lattes é superior para diagnósticos pontuais e exaustivos, bases como OpenAlex e Scopus são mais viáveis para o monitoramento periódico devido à facilidade de automação e menor tempo de processamento.

Palavras-chave: Fontes informacionais. Estudos métricos. Comparação de Fontes.

EIXO TEMÁTICO:

Bases e Fontes de Dados

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.976

COMO CITAR ESTE ARTIGO:

PLENTZ, Marina; MOURA, Ana Maria Mielniczuk de. Estratégias para análise da produção de pesquisadores: um estudo com Lattes, OpenAlex e Scopus. *In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. Anais [...].* Curitiba, 2026. DOI: 10.22477/x.ebbc.976. Disponível em: <https://doi.org/10.22477/x.ebbc.976>.



1 INTRODUÇÃO E FUNDAMENTAÇÃO

As bibliotecas acadêmicas, tem sido consideradas potenciais polos de apoio para os estudos métricos no contexto institucional, pois os bibliotecários já dominam ferramentas e recursos eletrônicos que os tornam atores estratégicos no apoio ao uso de métricas (Wolfram, 2020). Mas, o conhecimento de métricas das equipes de bibliotecas pode cumprir uma dupla função: promover o letramento em métricas entre pesquisadores e subsidiar diagnósticos sobre padrões de publicação que orientem estratégias institucionais de promoção da Ciência Aberta (Hoover; Coonin, 2021).

No contexto brasileiro o uso de bases comerciais é potencialmente problemático pela sub-representação do Sul Global e de periódicos Diamante (Borrego, 2023; Bosman *et al.*, 2021), ao buscar alternativas, o custo de limpeza e deduplicação pode se tornar um fator decisivo para a escolha do recurso. O procedimento precisa ser suficientemente leve para ser executado de forma recorrente pela biblioteca, sem depender de um esforço intensivo e pontual típico de projetos acadêmicos dedicados.

Alternativas (ou complementos) citadas na literatura incluem: Dimensions (Bosman *et al.*, 2021), Google Scholar (Digiampietri *et al.*, 2025; Pereira; Mugnaini, 2023); OpenAlex (Digiampietri *et al.*, 2025; Konch; Hazarika, 2025); e Plataforma Lattes (Digiampietri *et al.*, 2025; Pereira; Mugnaini, 2023).

As fontes bibliométricas distribuem-se em um gradiente entre o rigor do controle de qualidade (Scopus e Web of Science) e a amplitude documental (Google Scholar), com a OpenAlex buscando o equilíbrio via integração colaborativa (Digiampietri *et al.*, 2025). Já sistemas de currículos fornecem dados multidimensionais essenciais para avaliação, com uma cobertura documental ampla, mas de extração complexa. Plataformas nacionais como o Lattes, em contrapartida a situação europeia (Kousha;

Thelwall, 2014), facilitam a localização e padronização desses registros, embora a curadoria manual para ajustes e deduplicação permaneça indispensável.

O objetivo deste trabalho é analisar as implicações de esforço e cobertura na seleção de fontes para o monitoramento da produção em Ciência Aberta. Partiu-se do seguinte problema: quais os custos operacionais envolvidos no uso de diferentes bases de dados e como eles impactam a viabilidade administrativa dessas práticas em bibliotecas acadêmicas? Utilizamos uma base comercial, uma alternativa de acesso aberto e uma fonte curricular.

2 METODOLOGIA

O conjunto de pesquisadores a ser estudado são os professores, com vínculo nos cinco anos anteriores (2021-2025) com o Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental. Um total de 30 pesquisadores foi definido a partir dos dados do site do Programa¹. A seguir descreveremos como foi concebida a recuperação dos dados em cada fonte.

2.1 SCOPUS

Scopus (Elsevier), acessada via Portal de Periódicos CAPES, foi utilizada por oferecer perfis de autor, o que facilita a desambiguação. A partir dos nomes da amostra, foram localizados os perfis correspondentes; homônimos e perfis com nomes semelhantes foram excluídos. Ao final, foram selecionados 39 perfis para os 30 pesquisadores. Essa diferença resulta da fragmentação de registros gerada pelos algoritmos da base, que em alguns casos criam múltiplas entradas para um mesmo autor. Tal cenário exigiu uma etapa de curadoria manual de aproximadamente 75 minutos para identificar e reunir as entradas pertinentes, garantindo a integridade da coleta antes de salvá-las como uma “lista de autores”.

¹ <https://www.ufrgs.br/ppgiph/corpo-docente-2>

Com a lista ativa, a base recuperou os documentos associados aos perfis e foram aplicados filtros de período e de tipo de documento compatíveis com a produção usualmente registrada como artigos no Lattes (article, review, editorial). Além do painel de filtros que permite identificar o tipo de acesso aberto (classificação em Gold, Green, Hybrid Gold e Bronze) a base oferece o recurso “Analyse results”, que permite visualização em gráficos dos autores mais produtivos, áreas de assunto, periódicos com mais publicações, financiadores.

2.2 OPENALEX

OpenAlex, mantida por organização sem fins lucrativos e agregando dados de múltiplas fontes, permite consulta via interface ou API.

Iniciou-se pela identificação do(s) identificador(es) de autoria (author.id), pois, apesar dos mecanismos de desambiguação, observou-se: autores com múltiplos IDs e casos em que um mesmo ID aparecia associado a pesquisadores distintos com nomes semelhantes. Assim, a etapa de identificação foi mais morosa do que na Scopus e levou cerca de 120 minutos; ao final, para os 30 autores foram reunidos 82 IDs considerados pertinentes. As consultas podem ser salvas (para usuários logados), para buscas recorrentes é recomendável revisar periodicamente a vinculação de novos IDs, de modo análogo à verificação de perfis duplicados na Scopus.

Para contornar a ambiguidade identificada principalmente em dois autores, a estratégia incluiu um filtro adicional por afiliação institucional (UFRGS). A consulta final adotou os mesmos critérios da Scopus quanto a tipos de documento e aplicou recorte temporal. Os resultados listam trabalhos, ano, instituições e assuntos, além da indicação de acesso aberto. A base não apresenta a mesma categorização tipológica do OA disponível na Scopus, mas os dados exportáveis (incluindo CSV) trazem o campo de acesso aberto explicitamente discriminado, permitindo tabu-

lação e análise detalhada.

2.3 LATTES

Para utilizar dados do Currículo Lattes, é necessário extrair a produção a partir dos IDs Lattes, com ferramentas externas como o ScriptLattes (Mena-Chalco; Junior, 2013) ou o extrator da BRAPCI Bibliometric Tools (Bufrem *et al.*, 2010). Em relação às bases bibliográficas, a principal vantagem é que os identificadores são únicos, reduzindo problemas de desambiguação de autoria. O procedimento consistiu em identificar os IDs na plataforma, reuni-los em um arquivo .txt e executar a extração nas ferramentas. Para este conjunto de pesquisadores, a identificação dos IDs levou cerca de 30 minutos; a extração foi rápida (8–20 minutos), embora possa aumentar significativamente em grupos maiores.

As ferramentas exportam a produção em CSV, mas não fornecem, a priori, informação sobre acesso aberto. Assim, a partir do CSV extraído via BRAPCI foi criada uma tabela no OpenRefine, e os registros foram deduplicados por combinações de título + DOI, apenas DOI (quando o mesmo item aparecia com título em dois idiomas) e apenas título (para itens sem DOI). A classificação do status de acesso aberto foi estruturada em um fluxo hierárquico de validação via OpenRefine. Primeiramente, realizou-se o cruzamento automatizado dos DOIs, via expressões GREL, com as APIs do Unpaywall e do DOAJ. Para os registros não cobertos por essas bases, a verificação foi complementada por uma etapa de curadoria manual, consultando o portal Miguilim (Instituto Brasileiro de Informação em Ciência e Tecnologia, 2026), os sites dos periódicos e as políticas de direitos autorais dos artigos. Esse procedimento garantiu a identificação de publicações em periódicos nacionais e da via diamante. Considerando identificação, extração, deduplicação e consultas externas, o procedimento demandou várias sessões, totalizando aproximadamente 16 horas.

3 RESULTADOS E COMPARAÇÃO

A Scopus recuperou inicialmente 350 artigos em 137 revistas, a OpenAlex 400 artigos em 149 revistas e no Lattes foram identificados 432 artigos em 172 periódicos, caracterizando o Lattes como a fonte de maior cobertura.

Para avaliar diferenças de cobertura entre Scopus e OpenAlex, bases que podem excluir conteúdo por critérios de indexação ou limitações técnicas, os registros foram importados no Zotero e deduplicados, resultando em 441 itens únicos. Em seguida, testaram-se os 42 registros aparentemente exclusivos da Scopus (identificados pela procedência preservada no Zotero). Esses itens foram exportados em CSV e tratados no OpenRefine, onde expressões de consulta a API da OpenAlex verificaram sua existência na base. Apenas dois artigos não estavam na OpenAlex por ausência de cobertura; dez foram excluídos por divergência de data (publicação online em 2020 na OpenAlex vs. data final 2021 na Scopus); e, nos 30 restantes, a ausência de afiliação UFRGS impediu a recuperação quando se exigiu o filtro institucional.

O impacto dos filtros varia entre as bases. Na Scopus, o filtro institucional pouco alterou a recuperação (350 para 338 itens), validando a precisão dos perfis de autor. Na OpenAlex, a ausência de filtro gerou inflação de 79,41% em identificadores compartilhados. Embora o cruzamento de autor e instituição refine a precisão em homônimas, ele pode excluir registros com metadados de afiliação ausentes.

Mesmo com filtro UFRGS, a OpenAlex retorna 102 itens exclusivos; a inspeção (CSV dos Works no OpenRefine) mostrou que parte são anais de eventos (11) e preprints (8), reforçando a necessidade de limpeza.

4.1 RETRATO DO ACESSO ABERTO

A definição de acesso aberto adotada é deliberadamente restrita e considera apenas a disponibilização da versão final do artigo. Por isso, não foram incluídas as vias verde e bronze: a via verde nem sempre permite o depósito do texto final revisado, e a bronze é instável, podendo mudar conforme políticas editoriais.

A Scopus classifica o acesso aberto em Total, Gold, Green, Hybrid e Bronze. Dos 350 artigos recuperados, 174 estão em modalidades que garantem acesso à versão final (Gold e Hybrid Gold), correspondendo a 49,71% do conjunto.

Na interface da OpenAlex, o indicador geral aponta que 67,5% da produção (270 de 400 artigos) está em acesso aberto. Entretanto, a partir da exportação dos registros e de uma análise mais detalhada dos tipos de acesso – a base diferencia Gold, Gold Hybrid, Diamante, Green e Bronze – o percentual de itens em acesso aberto cai para 58,5%, conforme apresentado na Tabela 1. A Scopus agrupa o modelo diamante ao Gold, mas para as bibliotecas é útil identificar publicações sem custos já utilizadas. Valorizar experiências positivas da própria comunidade nesses canais ajuda a disseminar boas práticas e fortalecer a confiança entre pares.

Tabela 1 - Tipo de acesso de acordo com a base

Base	Fechado	Gold	Hybrid Gold	Diamante	Total	Percentual em acesso aberto
Scopus	176	157	17	N/D	350	49,71%
OpenAlex	166	80	14	140	400	58,5%
Lattes	166	134	21	111	432	61,57%

Fonte: elaboração com base nos resultados da Scopus, OpenAlex e dados do Lattes

A alta incidência de registros diamante na OpenAlex decorre, em parte, de falsos positivos na categorização, como já apontado por Pilatti et al. (2025). Isso é obser-

vado em periódicos com cobrança de APC, como a Revista Brasileira de Recursos Hídricos e títulos da International Water Association, cujas políticas de isenção podem gerar inconsistências na identificação automatizada do modelo de acesso pela base.

4 CONCLUSÃO

O currículo Lattes garante maior precisão e abrangência, mas a complexidade de limpeza dificulta o monitoramento contínuo. Já Scopus e OpenAlex são mais eficientes para consultas recorrentes, pois permitem salvar e replicar estratégias de busca com baixo esforço.

A Scopus apresenta maior precisão na recuperação, mas como apontado na literatura é mais limitada em termos de cobertura de fontes e, quando comparada ao Lattes, tomado aqui como padrão de referência, oferece um retrato do acesso aberto aquém daquilo que parece ser a realidade.

A OpenAlex produz resultados mais próximos aos do Lattes. Enfrenta desafios na desambiguação de autores, mas a expansão do uso de identificadores únicos pode trazer melhorias futuras nesse aspecto. Outro elemento de atenção é a classificação equivocada de periódicos que cobram APC como diamante, fator problemático quando se pretende mapear canais alternativos de publicação institucional sem custos para os autores. Assim, se por um lado bases comerciais podem subdimensionar o percentual de acesso aberto, por outro, a OpenAlex pode inflá-lo. Ainda assim, pela capacidade de recuperar um conjunto mais amplo de resultados e pelo uso da API para gerar buscas reproduzíveis e compartilháveis, a OpenAlex se apresenta como o recurso mais promissor para o monitoramento sistemático da produção com vistas ao uso por equipes de bibliotecas.

REFERÊNCIAS

- BORREGO, Á. Article processing charges for open access journal publishing: a review. **Learned Publishing**, Watford, v. 36, n. 3, p. 359–378, 2023. Disponível em: <https://doi.org/10.1002/leap.1558>. Acesso em: 16 out. 2025.
- BOSMAN, J.; FRANTSVÅG, J. E.; KRAMER, B.; LANGLAIS, P.-C.; PROUDMAN, V. **OA diamond journals study. Part 1: findings**. [S. l.]: [S. n.], 2021. DOI: 10.5281/ZENODO.4558704. Disponível em: <https://zenodo.org/record/4558704>. Acesso em: 2 out. 2025.
- BUFREM, L. S.; COSTA, F.D.de O.; GABRIEL JUNIOR, R.F.; PINTO, J.S.de P. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, Campinas, v. 15, p. 22–41, 2010. Disponível em: <https://doi.org/10.1590/S1413-99362010000200003>. Acesso em: 19 fev. 2026.
- DIGIAMPIETRI, L. *et al.* Approaching bibliometrics and prosopography. **Encontros Bibli**, Florianópolis, v. 30, p. 1–34, 2025. Disponível em: <https://doi.org/10.5007/1518-2924.2025.e103528>. Acesso em: 20 jan. 2026.
- HOOVER, J.; COONIN, B. Tips from the experts talking with your STEM faculty about open access: a checklist. **Issues in Science and Technology Librarianship**, Chicago, v. 2021, n. 97, 2021. Disponível em: <https://doi.org/10.29173/istl2589>. Acesso em: 10 out. 2025.
- INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Miguilim**. Brasília, DF, 2026. Disponível em: <https://miguilim.ibict.br/static/pages/miguilim.jsp>. Acesso em: 19 fev. 2026.
- KONCH, P.; HAZARIKA, H. Scientometric analysis of adaptive learning literature: insights from the OpenAlex database. **Education for Information**, [s. l.], 2025. Disponível em: <https://doi.org/10.1177/01678329251364338>. Acesso em: 10 out. 2025.
- KOUSHA, K.; THELWALL, M. Disseminating research with web CV hyperlinks. **Journal of the Association for Information Science and Technology**, Hoboken, v. 65, n. 8, p. 1615–1626, 2014. Disponível em: <https://doi.org/10.1002/asi.23070>. Acesso em: 14 out. 2025.
- MENA-CHALCO, J. P.; JUNIOR, R. M. C. **Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes**. [S. l.]: Unpublished, 2013. Disponível em: <https://www.researchgate.net/doi/10.13140/RG.2.1.5183.8561>. Acesso em: 19 fev. 2026.
- PEREIRA, F. A.; MUGNAINI, R. Mapping the use of Google Scholar in evaluative bib-



liometric or scientometric studies: A bibliometric review. **Quantitative Science Studies**, Cambridge, Massachusetts, v. 4, n. 1, p. 233–245, 2023. Disponível em: https://doi.org/10.1162/qss_a_00231. Acesso em: 20 jan. 2026.

PILATTI, L. E. *et al.* From fees to free: comparing APC-based and diamond open access journals in engineering. **Publications**, Basel, v. 13, n. 2, p. 16, 2025. Disponível em: <https://doi.org/10.3390/publications13020016>. Acesso em: 14 out. 2025.

WOLFRAM, D. O papel da biblioteca acadêmica na promoção efetiva da comunicação científica e das aplicações bibliométricas para a avaliação das pesquisas. *In*: GRÁCIO, M. C. C. *et al.* (org.). **Tópicos da bibliometria para bibliotecas universitárias**. São Paulo: Cultura Acadêmica, 2020. Disponível em: <https://books.scielo.org/id/zvdpp>. Acesso em: 20 jan. 2026.