



DESVENDANDO O OPENALEX:

avanços e oportunidades em métricas bibliométricas

Diego José Macêdo

 <https://orcid.org/0000-0002-5696-0639>.

✉ diegomacedo@ibict.br.

🏢 Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

ROR  <https://ror.org/006c42y96> | Brasília, Brasil.

Ingrid Torres Schiessl

 <https://orcid.org/0000-0001-5815-2574>.

✉ ingridschiessl@ibict.br.

🏢 Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

ROR  <https://ror.org/006c42y96> | Brasília, Brasil.

Milton Shintaku

 <https://orcid.org/0000-0002-6476-4953>.

✉ shintaku@ibict.br.

🏢 Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) |

ROR  <https://ror.org/006c42y96> | Brasília, Brasil.

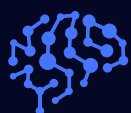
Eixo temático: Técnicas, Ferramentas e Infraestruturas

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.353

Resumo: Com o avanço da informática, surgem ferramentas métricas gratuitas, como o *OpenAlex*, ampliando o acesso. Este estudo investiga seu potencial inovador, descrevendo uma metodologia em duas etapas. A primeira etapa visa explorar a ferramenta, enquanto a segunda compara suas métricas com outras plataformas. Os resultados ressaltam a capacidade do *OpenAlex* de fornecer uma alternativa viável e complementar às ferramentas convencionais, especialmente na análise comparativa de métricas de publicação e citação.

Palavras-Chave: Software livre. *OpenAlex*. Ferramentas métricas. Acesso livre. Ferramenta bibliográfica.



1 INTRODUÇÃO

Desde os tempos imemoriais, a humanidade tem realizado medições, abrangendo tanto elementos tangíveis, como propriedades, quanto intangíveis, como o tempo. Na esfera das ciências, a medição tem sido fundamental ao longo de sua história, com a adoção de sistemas de medição reconhecidos internacionalmente. Para assegurar a precisão necessária nos estudos, diversas métricas foram desenvolvidas. Na biblioteconomia, várias dessas métricas são aplicadas, sendo a bibliometria uma das mais relevantes para este trabalho. A bibliometria, que se alinha com a estatística, é utilizada para quantificar diversos aspectos das publicações.

Historicamente a bibliometria nasce na década de 1920, ainda chamada como bibliografia estatística, mas em um pequeno artigo publicado por Groos e Pritchard (1969), algumas décadas depois, a métrica foi adaptada para a nomenclatura atual, assim como seu entendimento atual. Fonseca (1973), no entanto, relembra que o primeiro estudioso a utilizar o termo bibliometria foi o célebre Paul Otlet no *Traité de documentation*, no capítulo em que se refere ao livro e sua medida, defendendo a necessidade de consolidar a bibliometria como métrica dos livros e documentação (Otlet, 1934).

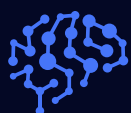
Independentemente da origem do termo, a bibliometria tornou-se uma das principais métricas das publicações para medir a informação, em que muitos dos seus princípios são utilizados por outras métricas. Assim, a bibliometria evoluiu com suas leis (Bradford, Lotka e Zipf) e indicadores, sendo um dos mais importantes o fator de impacto das publicações, mesmo com todas as suas críticas, proposto por Garfield (1955, p.111), em que “um artigo altamente significativo, o índice de citação tem um valor quantitativo, pois pode ajudar o historiador a medir a influência do artigo, ou seja, o seu ‘fator de impacto’”.

Com a evolução da informática, algumas ferramentas foram desenvolvidas para fazer as medições antes feitas manualmente, tendo algumas empresas atuando na área. Entretanto, uma das críticas a essas empresas é a necessidade de pagamento de licenças, entre outros, limitando o acesso à comunidade aos seus indicadores. Assim, algumas iniciativas estão disponíveis de forma gratuita, como o índice *H* ofertado pelo Google®, alinhado aos movimentos do acesso aberto à informação científica (*Open Access*) e a ciência aberta (*Open Science*).

Neste contexto, surge o *OpenAlex*¹, ferramenta online disponibilizada na web, para métrica bibliométrica. Priem, Piwowar e Orr (2022) relatam que o *OpenAlex* nasce com a preocupação do descontinuação do *Microsoft Academic Graph* (MAG), tendo o potencial para melhorar a transparência da avaliação, navegação, representação e descoberta de pesquisas, sendo totalmente livre, de código aberto e integrável por meio de *Application Programming Interface* (API). Dessa forma, torna-se uma opção aos estudos métricos em bibliometria.

Embora o *OpenAlex* represente uma ferramenta promissora para os Estudos Métricos da Informação, seu uso ainda não é amplamente disseminado na comunidade acadêmica. Isso se deve, em parte, ao conhecimento limitado sobre suas capacidades e funcionalidades, que são essenciais para a realização eficaz e eficiente de estudos métricos. Portanto, surge um questionamento central: de que maneira o *OpenAlex* pode ser empregado para ampliar os horizontes dos estudos métricos em bibliometria,

¹ Disponível em: <https://openalex.org>. Acesso em: 30 abr. 2024.



ultrapassando as barreiras impostas por ferramentas tradicionais e contribuindo para uma maior democratização do acesso à informação científica?

Com base nesse questionamento, o objetivo deste estudo é explorar o potencial do *OpenAlex* como uma ferramenta inovadora para estudos métricos. Pretende-se investigar como essa plataforma online pode ser utilizada em pesquisas no contexto brasileiro, destacando-se não apenas pela sua aplicabilidade em estudos métricos, que são fundamentais no campo da ciência da informação, mas também por sua capacidade de fomentar a geração de novos conhecimentos. Assim, este trabalho visa não apenas apresentar o *OpenAlex* como uma alternativa viável para pesquisa, mas também contribuir para o avanço da área, explorando as diversas possibilidades que essa ferramenta oferece para a comunidade científica.

2 METODOLOGIA

Neste estudo, adotou-se uma metodologia em duas etapas principais, visando explorar a plataforma *OpenAlex* e comparar suas métricas com as de outras plataformas acadêmicas. A primeira etapa concentrou-se em uma análise descritiva da *OpenAlex* para entender suas principais funcionalidades, interface e possíveis aplicações acadêmicas. Esta análise qualitativa focou na ferramenta de busca e filtros disponíveis, bem como na apresentação de dados sobre publicações e citações, com o objetivo de avaliar a usabilidade e eficiência da plataforma em prover informações relevantes para pesquisadores em diversos campos.

Na segunda etapa, realizou-se uma análise comparativa envolvendo a *OpenAlex* e outras plataformas como *Google Scholar Citations*², *Scopus*³, *Web of Science*⁴ e *Dimensions*⁵. Esta comparação teve como foco as métricas de publicações e citações oferecidas por cada plataforma. Para tanto, selecionou-se inicialmente no *Google Scholar Citations* os perfis dos dez autores mais citados na área de interesse indicada pelo pesquisador em “Ciência da Informação”, garantindo uma amostra de autores reconhecidos nas demais plataformas. Esses mesmos autores foram posteriormente localizados em cada uma das plataformas para comparar o número total de publicações e citações, mantendo o anonimato dos autores para assegurar a imparcialidade da análise. Esta metodologia visa proporcionar uma compreensão detalhada e sistemática das funcionalidades e métricas das plataformas analisadas, destacando a importância de selecionar recursos de pesquisa adequados para otimizar o acesso e uso de informações científicas relevantes.

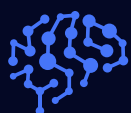
Adicionalmente, realizou-se uma análise comparativa entre o *Google Scholar Citations*, *OpenAlex*, *Scopus*, *Web of Science* e *Dimensions* com o propósito de identificar as publicações e citações disponibilizadas por ambas as plataformas. Essa comparação visa analisar as métricas apresentadas por cada plataforma. Para tanto, selecionou-se inicialmente no *Google Scholar Citations* os perfis dos dez autores mais citados na área de interesse indicada pelo pesquisador em “Ciência da Informação”, esse tipo de

² Disponível em: <https://scholar.google.com/citations>. Acesso em: 30 abr. 2024.

³ Disponível em: <https://www.scopus.com>. Acesso em: 30 abr. 2024.

⁴ Disponível em: <https://www.webofscience.com>. Acesso em: 30 abr. 2024.

⁵ Disponível em: <https://www.dimensions.ai>. Acesso em: 30 abr. 2024.



seleção foi para selecionar uma amostra que também fosse relevante nas demais bases. De maneira equivalente, identificaram-se os mesmos dez autores no *OpenAlex*, *Scopus*, *Web of Science* e *Dimensions* para possibilitar uma comparação efetiva entre a quantidade de publicações e citações. Após a seleção dos autores, procedeu-se com a escolha das quantidades de trabalhos, citações associadas para fins de comparação das métricas apresentadas nas plataformas. Os nomes dos autores foram suprimidos, pois para análise realizada não seria necessária.

3 RESULTADOS

Para compreender as características e o potencial da plataforma *OpenAlex*, foi inicialmente realizada uma análise detalhada da documentação disponível no site oficial, complementada por uma investigação prática de suas funcionalidades. Esta análise revelou que o *OpenAlex*, administrado por uma entidade sem fins lucrativos, não apenas se inspira no histórico legado da Biblioteca de Alexandria, mas também se esforça para transcender a simples compilação do conhecimento. Ao invés disso, a plataforma se dedica a aprimorar a indexação e a recuperação do conhecimento, fazendo uma reverência moderna ao catálogo Pinakes - um símbolo do início dos esforços para organizar e disseminar o conhecimento. Esta conexão entre a consulta inicial e as descobertas subsequentes destaca a missão da *OpenAlex* de facilitar o acesso ao conhecimento de maneira eficaz e inovadora.

A premissa fundamental da base é sua natureza livre, com os dados sendo disponibilizados para reuso mediante a utilização da licença *Creative Commons* CC0. Essa licença implica que os dados fornecidos estão em domínio público, proporcionando assim total liberdade para sua utilização.

O *OpenAlex* disponibiliza duas alternativas principais para o acesso aos seus dados: por meio da *Representational State Transfer* (API REST) ou por meio de uma cópia integral do banco de dados, a qual pode ser implementada em um *Data Warehouse* ou em um banco de dados relacional. Contudo, a plataforma sugere, de modo preferencial, a utilização da API REST devido à sua eficácia e simplicidade de integração com diferentes aplicações.

Uma das características principais observadas é a representação dos dados em entidades para facilitar a recuperação da informação. Essas entidades são representadas por um conjunto de nove categorias: trabalhos, autores, fontes, instituições, assuntos, publicadores, agências de fomento, geolocalizações e conceitos, sendo que esta última está em processo de descontinuação. Para cada uma dessas entidades, é possível realizar buscas, aplicar filtros e utilizar recursos de facetas por meio da interface web, além de acessar as informações dessas entidades por meio da API, aumentando as possibilidades de descobertas de informações.

Para as entidades mencionadas, a plataforma obtém informações das fontes do MAG, Crossref, PubMed, repositórios institucionais, ORCID, Wikidata e ROR, fornecendo dados sobre publicações e utilizando identificadores internos persistentes para uma melhor identificação e deduplicação das entidades. A ferramenta também emprega identificadores externos canônicos para reconhecer de maneira unívoca os objetos das entidades, como os seguintes: para trabalhos, o DOI; para autores, o ORCID; para fontes, o ISSN-L; para instituições, o ROR ID; para conceitos, o Wikidata ID; e para publicadores,



também o Wikidata ID.

Para avaliar a extensão da cobertura da base de dados *OpenAlex* em comparação com outras, realizou-se uma análise concentrada no volume de publicações e no número de citações de autores especializados em ciência da informação, os quais foram selecionados através do *Google Scholar Citations*. Esta investigação foi aprofundada pelo exame do Quadro 1, que oferece uma comparação entre o número de publicações e citações desses autores nas bases de dados apresentadas. Destaca-se, por exemplo, o autor A3, que apresenta o maior volume de trabalhos publicados no *Google Scholar Citations*. Em contraste, o autor A7, embora não possua o maior número de publicações, destaca-se pelo elevado número de citações recebidas no *OpenAlex*, indicando a significativa influência de seus trabalhos na comunidade acadêmica. Esta análise sublinha a importância de considerar tanto a quantidade de publicações quanto o impacto das mesmas, por meio do número de citações, para uma avaliação precisa do desempenho acadêmico.

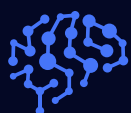
Para avaliar a extensão da cobertura da base de dados *OpenAlex* em comparação com as outras, realizou-se uma análise concentrada no volume de publicações e no número de citações de autores em ciência da informação, os quais foram selecionados por meio do *Google Scholar Citations*. Prosseguindo com a análise do Quadro 1 que apresenta um comparativo entre as publicações e citações de diferentes bases de dados para dez autores, é possível observar diversas nuances. Como ilustração, o autor A3 destaca-se com o maior número de trabalhos publicados no *Google Scholar Citations*, indicando uma alta produtividade. Por outro lado, o autor A7, apesar de não ter o maior número de publicações, possui um alto número de citações no *OpenAlex*. Isso ressalta a importância de considerar tanto a quantidade de trabalhos quanto o total de citações ao avaliar o desempenho acadêmico, além de considerar o uso de outras bases para comparação da produção.

Quadro 1 - Comparativo de volume de publicações e citações das bases

Código Autor	Google Scholar Citations		OpenAlex		Scopus		Web of Science		Dimensions	
	Qtd. trab.	Total Cit.	Qtd. trab.	Total Cit.	Qtd. trab.	Total Cit.	Qtd. trab.	Total Cit.	Qtd. trab.	Total Cit.
A1	293	6823	213	411	35	103	27	36	87	185
A2	188	5721	150	395	17	64	14	48	70	178
A3	527	5063	343	298	47	98	30	37	158	171
A4	185	4699	117	630	12	45	10	34	39	301
A5	189	4151	163	276	23	40	32	26	66	112
A6	205	3390	109	183	10	6	5	18	22	52
A7	143	3096	110	882	28	559	16	314	61	768
A8	245	3023	118	111	2	0	*	*	36	47
A9	134	2890	88	313	5	23	6	30	32	143
A10	417	2655	176	248	16	63	8	42	35	123

Fonte: Elaborado pelos autores (2024).

Além disso, ao comparar as bases de dados, nota-se que o *Google Scholar Citations* frequentemente exibe um número maior de trabalhos e um volume significativamente maior de citações. Isso ocorre



devido à sua política de indexação aberta e à inclusão de uma ampla variedade de documentos, tais como memorandos técnicos, literatura cinzenta, guias e apresentações, e também pela ausência de um processo de deduplicação.

O *OpenAlex*, promove uma abordagem de indexação e recuperação do conhecimento que é aberta, conforme discutido na metodologia do estudo que originou o quadro. A presença significativa de alguns autores nesta base, como o autor A7 com 882 citações, destaca seu potencial para ampla disseminação e reconhecimento acadêmico.

Bases como *Scopus* e *Web of Science*, com critérios de inclusão mais rigorosos, tendem a ter números menores de trabalhos e citações. As variações observadas entre diferentes autores e bases de dados podem refletir diferenças disciplinares e a aderência de determinadas áreas do conhecimento a certas bases de dados.

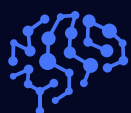
Identificar tendências, como o aumento da adoção do acesso aberto, é possível ao observar o crescimento de citações em plataformas como o *OpenAlex*, alinhado com o movimento de ciência aberta e acesso livre à informação científica, potencialmente aumentando a visibilidade e o impacto dos trabalhos publicados nessas plataformas. Para pesquisadores e instituições, compreender essas dinâmicas é crucial para maximizar a visibilidade e o impacto de suas pesquisas, considerando a diversidade de plataformas e suas audiências específicas. O Quadro 1 não apenas revela as diferenças em números de publicações e citações entre várias bases de dados para um conjunto de autores, mas também sublinha a importância de uma estratégia de publicação bem pensada que leve em conta a diversidade de plataformas e suas audiências específicas, ressaltando o valor das ferramentas de acesso aberto como o *OpenAlex* no avanço da ciência aberta.

4 CONSIDERAÇÕES FINAIS

O estudo apresentado oferece uma visão abrangente da relevância e do potencial do *OpenAlex* como uma ferramenta bibliográfica inovadora e de livre acesso para a comunidade acadêmica. A investigação detalhada de suas funcionalidades, juntamente com a comparação metódica com outras plataformas estabelecidas, revela o papel significativo do *OpenAlex* na promoção de uma maior transparência e democratização do acesso à informação científica.

Os resultados obtidos, especialmente a partir da análise comparativa de métricas de publicação e citação, sublinham a capacidade do *OpenAlex* de oferecer uma alternativa viável e complementar às ferramentas tradicionais. A plataforma com sua filosofia de código aberto e abordagem inclusiva, alinha-se com os princípios da ciência aberta e do acesso livre.

Por fim, o trabalho realizado sugere que iniciativas como o *OpenAlex* são fundamentais para o futuro da pesquisa acadêmica, ao proporcionar ferramentas que não só ampliam o acesso à informação científica, mas também promovem uma cultura de transparência e colaboração.



REFERÊNCIAS

BFONSECA, Edson Nery da. Bibliografia estatística e bibliometria: uma reivindicação de prioridades.

Ciência da Informação, Rio de Janeiro, v. 2, n. 1, p. 5-7, 1973. DOI: <https://doi.org/10.18225/ci.inf.v2i1.19>. Disponível em: <https://revista.ibict.br/ciinf/article/view/19>. Acesso em: 15 maio 2024.

GARFIELD, Eugene. Citation indexes for science; a new dimension in documentation through association of ideas. **Science**, New York, v. 122, n. 3159, p. 108-111, 15 jul. 1955. DOI: <https://doi.org/10.1126/science.122.3159.108>. Disponível em: <https://www.science.org/doi/10.1126/science.122.3159.108>. Acesso em: 19 maio 2024.

OTLET, Paul. **Traité de Documentation**: le livre sur le livre. Bruxelles: Mundaneum, 1934. Disponível em: https://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf. Acesso em: 9 fev. 2024.

PRIEM, Jason; PIWOWAR, Heather; ORR, Richard. OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts. *In*: INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS, 26., 2022, Granada. **Anais [...]**. 17 jun. 2022. Itachi, NY: Cornell University, 17 June 2022. DOI: <https://doi.org/10.48550/arXiv.2205.01833>. Disponível em: <http://arxiv.org/abs/2205.01833>. Acesso em: 9 mar. 2024.

GROOS, Ole V.; PRITCHARD, Alan. Documentation Notes. **Journal of Documentation**, v. 25, n. 4, p. 344-349, 1 abr. 1969. DOI: <https://doi.org/10.1108/eb026482>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/eb026482/full/html>. Acesso em: 10 maio 2024.