

O uso do OpenAlex nos estudos de métricas bibliométricas

The use of OpenAlex in bibliometric metrics studies

El uso de OpenAlex en los estudios de métricas bibliométricas

Diego José Macêdo

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, Brasil

Ingrid Torres Schiessl

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, Brasil

Milton Shintaku

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, Brasil

ORIGINAL

Resumo

Objetivo. Este estudo investiga o potencial inovador da ferramenta OpenAlex, que surge como uma alternativa gratuita no campo das métricas informacionais. O objetivo principal é avaliar a capacidade dessa ferramenta em comparação com outras plataformas métricas convencionais. **Método.** A pesquisa foi conduzida em duas etapas distintas. A primeira consistiu na exploração detalhada das funcionalidades oferecidas pelo OpenAlex, enquanto a segunda etapa envolveu a comparação direta das métricas de publicação e citação fornecidas pelo OpenAlex com aquelas obtidas nas plataformas Google Scholar Citations, Scopus, Web of Science e Dimensions. **Resultados.** Os resultados demonstraram que o OpenAlex é uma ferramenta viável e eficaz, oferecendo métricas que complementam as fornecidas pelas plataformas convencionais, com destaque para sua utilidade na análise comparativa de publicações e citações. **Conclusões.** Conclui-se que o OpenAlex possui um grande potencial para ser utilizado como uma ferramenta complementar às plataformas tradicionais, especialmente na análise de métricas de publicação e citação, contribuindo para uma maior acessibilidade e diversidade de fontes métricas.

Palavras-chave: acesso aberto, openalex, ferramentas métricas, acesso livre, ferramenta bibliográfica

Abstract

Objective. This study investigates the innovative potential of the OpenAlex tool, which emerges as a free alternative in the field of informational metrics. The primary objective is to evaluate the capability of this tool in comparison with other conventional metric platforms. **Method.** The research was conducted in two distinct stages: the first consisted of a detailed exploration of the functionalities offered by OpenAlex; the second stage involved a direct comparison of the publication and citation metrics provided by OpenAlex with those obtained from other traditional platforms. **Results.** The results demonstrated that OpenAlex is a viable and effective tool, offering metrics that complement those provided by conventional platforms, with a focus on its utility in the comparative analysis of publications and citations. **Conclusions.** It is concluded that OpenAlex has significant potential to be used as a complementary tool to traditional platforms, particularly in the analysis of publication and citation metrics, thereby contributing to greater accessibility and diversity of metric sources.

Keywords: open access, openalex, metric tools, free access, bibliographic tool

Resumen

Objetivo. Este estudio investiga el potencial innovador de la herramienta OpenAlex, que surge como una alternativa gratuita en el campo de las métricas informacionales. El objetivo principal es evaluar la capacidad de esta herramienta en comparación con otras plataformas métricas convencionales. **Método.** La investigación se llevó a cabo en dos etapas distintas. La primera consistió en la exploración detallada de las funcionalidades ofrecidas por OpenAlex, mientras que la segunda etapa implicó la comparación directa de las métricas de publicación y citación proporcionadas por OpenAlex con aquellas obtenidas en las plataformas Google Scholar Citations, Scopus, Web of Science y Dimensions. **Resultados.** Los resultados demostraron que OpenAlex es una herramienta viable y eficaz, que ofrece métricas que complementan las proporcionadas por las plataformas convencionales, destacándose por su utilidad en el análisis comparativo de publicaciones y citaciones. **Conclusiones.** Se concluye que OpenAlex posee un gran potencial para ser utilizado como una herramienta complementaria a las plataformas tradicionales, especialmente en el análisis de métricas de publicación y citación, contribuyendo a una mayor accesibilidad y diversidad de fuentes métricas.

Palabras clave: acceso abierto, openalex, herramientas métricas, acceso libre, herramienta 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 à estatística, é utilizada para quantificar diversos aspectos das publicações.

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

Independentemente da origem do termo, a bibliometria tornou-se uma das principais métricas utilizadas para medir informações em publicações, sendo que muitos dos seus princípios são utilizados por outras métricas. Assim, a bibliometria evoluiu com suas leis (Bradford, Lotka e Zipt) e indicadores, destacando-se o fator de impacto das publicações, proposto por Garfield (1955). Apesar das críticas, o fator de impacto é amplamente utilizado, pois, como afirma Garfield, “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’” (p. 111).

Com a evolução da informática, algumas ferramentas foram desenvolvidas para automatizar medições que antes eram feitas manualmente, e várias empresas passaram a atuar nessa área. No entanto, uma das principais críticas a essas empresas é a necessidade de pagamento de licenças, o que limita o acesso da comunidade a seus indicadores. Em resposta, algumas iniciativas foram disponibilizadas gratuitamente, como o índice H oferecido pelo Google®, alinhado aos movimentos do acesso aberto à informação científica (*Open Access*) e a ciência aberta (*Open Science*).

Nesse contexto, surge o *OpenAlex*, uma ferramenta on-line disponibilizada na web para métrica bibliométrica (*OpenAlex*, 2024). Priem et al. (2022) relatam que o *OpenAlex* foi desenvolvido devido ao descontinuamento do *Microsoft Academic Graph*, com o potencial de melhorar a transparência na avaliação, navegação, representação e descoberta de pesquisas. A ferramenta é totalmente livre, de código aberto e integrável por meio de *Application Programming Interface* (API), tornando-se, assim, uma alternativa viável para 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, 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 on-line pode ser

utilizada em pesquisas no contexto brasileiro, destacando-se não apenas por sua aplicabilidade em estudos métricos, 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 Formato de texto

Desde a Antiguidade, o conhecimento registrado tem sido o principal canal de disseminação das ciências, utilizando-se da tecnologia disponível em cada período histórico. Assim, os registros já assumiram diversos suportes, como papiros, tábuas de argila e madeira, ossos e tantos outros. Ao longo do tempo, a gestão desse conhecimento registrado, independentemente do suporte, ficou a cargo das bibliotecas e, em certos contextos, dos arquivos. No âmbito científico, inicialmente de caráter filosófico, essa responsabilidade foi exercida exclusivamente pelas bibliotecas.

Por isso, pode-se afirmar que os catálogos de bibliotecas foram as primeiras bases indexadoras, pois facilitaram a recuperação das obras existentes na biblioteca. No entanto, esses catálogos eram locais, restringindo o acesso ao conhecimento. Com o avanço da tecnologia, tornou-se possível exportar e agregar esses catálogos, criando bases integradas, especialmente com o crescente uso das Tecnologias da Informação e Comunicação (TIC), o que permitiu identificar em quais bibliotecas estão determinados livro ou revista estava disponível.

A integração de catálogos de várias bibliotecas foi o caminho da primeira base indexadora aberta desenvolvida no Brasil, o Catálogo Coletivo Nacional (CCN), proposto pelo então Instituto Brasileiro de Bibliografia e Documentação (IBBD), posteriormente Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Segundo Oddone et al. (2023), a ideia do CCN foi inspirada no Serviço de Intercâmbio de Catalogação (SIC), mantido pela Imprensa Nacional. Com isso, a primeira base indexadora aberta nacional ainda está ativa e on-line, tendo passado por evoluções, como relatado Costa (2024) e agora faz parte de um serviço informacional aberto maior, denominado de *Pinakes*.

Com o surgimento da Web, na década de 1990, inúmeras opções foram surgindo, sendo a criação do projeto arXiv um marco para a ciência, pois ofereceu acesso ao texto completo dos chamados preprints, ou versão do autor, já que são artigos que ainda não passaram pelo processo editorial (Ginsparg, 2011). Com a possibilidade de os autores depositarem obras científicas em sistemas disponibilizados na web, foi criado o primeiro movimento de abertura das ciências, o *Open Archives*, cujo maior expoente no Brasil é a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), mantida pelo Ibict. A BD TD é uma base indexadora de teses e dissertações defendidas no Brasil e participa da maior base mundial de conhecimento do gênero, a *Networked Digital Library of Theses and Dissertations*.

Com a informatização, grandes bases de conhecimento proprietárias e pagas também ganharam espaço, à medida que agregam várias fontes, incluindo as dos grandes editores. Nesse sentido, cabe destacar duas das principais: a *Web of Science* (WoS) e a Scopus, que agregam grande parte da produção científica mundial, principalmente de bases privadas. Pode-se dizer que a WoS, como é conhecida atualmente, foi criada em 1997 pelo *Institute of Science Information* (ISI) e, posteriormente, vendida para a empresa *Clarivate*. Schötten et al (2017), ao descreverem a história da base de conhecimento *Scopus*, relatam que essa base, que indexa grande parte das fontes científicas do norte global, foi disponibilizada em 2004. Apesar de serem pagas, essas bases ainda desempenham um papel importante para o ecossistema científico, na medida em que agregam outras bases pagas.

No Brasil, a primeira grande base indexadora a disponibilizar o texto completo, seguindo os preceitos do movimento de acesso aberto (*Open Access*), foi a *Scientific Electronic Library Online* (SciELO), criada em 1997. A metodologia SciELO obteve um grande sucesso, sendo exportada para vários países da América Latina e África do Sul. De certa forma, a SciELO inovou e foi a precursora do movimento de acesso aberto. Inicialmente a SciELO tinha maior ênfase nas ciências da saúde, mas gradativamente incorporou outras áreas do conhecimento, tornando-se uma base de conhecimento gratuita, amplamente reconhecida e confiável.

Nesse seguimento, o Ibict lançou o *OasisBr*, uma base de conhecimento nacional que agrega quase todos os repositórios institucionais das universidades e unidades de pesquisa, além de alguns portais de revistas, incluindo a própria SciELO. Com isso, o *OasisBr* se tornou a maior base de conhecimento aberta do Brasil, fazendo parte

de outras bases, como o Repositórios Científicos de Acesso Aberto de Portugal (RCAAP) e o *LA Referencia* da América Latina.

Essas grandes bases abertas representam o que advoga Harnad et al (2008) sobre o impacto positivo da abertura do acesso à produção científica. Tanto que, o Brasil foi um dos precursores da criação de bases de acesso aberto na América Latina, com a SciELO e posteriormente fomento à criação e integração de repositórios e revistas diamante pelo Ibict, esses modelos foram disseminados para vários países pelo projeto do *LA Referencia*.

Com o sucesso do buscador Google, foi criado em 2004 o Google Acadêmico, que indexa apenas as fontes consideradas científicas, com base na tecnologia utilizada. Para alguns especialistas, o Google Acadêmico é visto mais como um buscador e do que uma base de conhecimento propriamente dita. Entretanto, tornou-se uma ferramenta importante na pesquisa, pois indexa grande parte das fontes científicas existentes. No entanto, isso pode ser um complicador, já que retorna uma grande quantidade de registros, o que dificulta a verificação do que é realmente relevante. Outro aspecto a ser observado é que o Google Acadêmico não disponibiliza um formato de download e nem API que permita estudos bibliométricos. Nesse sentido, Costa (2006) defende que o acesso aberto precisa utilizar tecnologia livres, de forma a transformar se em uma filosofia aberta.

Nesse contexto, revela-se que há várias bases de conhecimento, abertas ou restritas, cada uma com suas vantagens e desvantagens, e os pesquisadores precisam avaliar qual fonte é a mais adequada para seus estudos. Nesse cenário, surge uma nova proposta: o *OpenAlex*, que não apenas agrupa uma base de conhecimento, mas também oferece serviços informacionais que facilitam a recuperação de dados e a descoberta de literatura relevante.

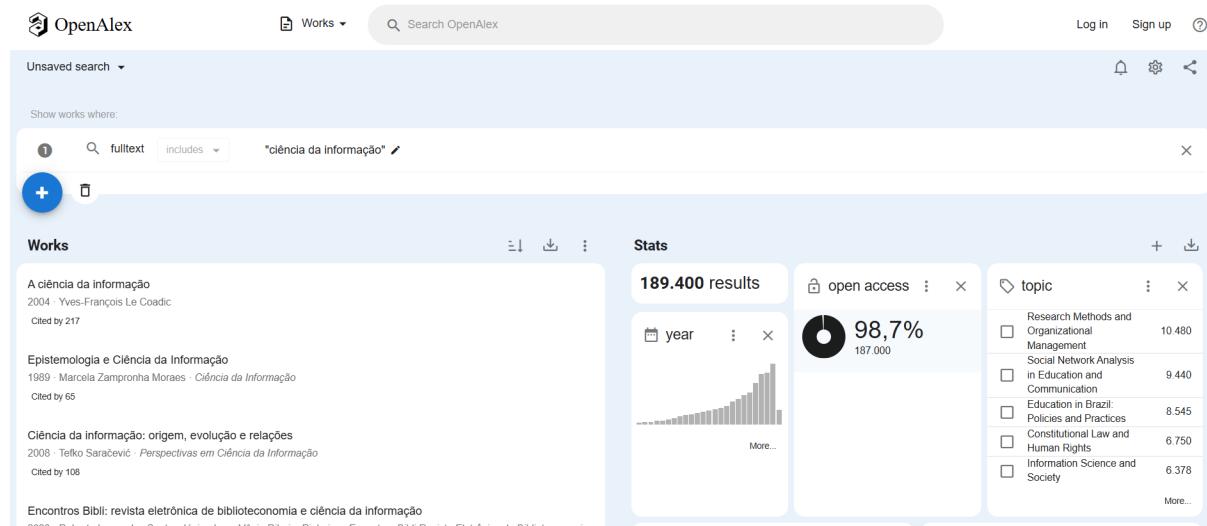
3 O OpenAlex

O nome *OpenAlex*, conforme descrito em seu site (*OpenAlex*, 2024), faz referência à grande Biblioteca de Alexandria, localizada no Egito durante a Antiguidade. Além disso, o nome reflete a filosofia de seus criadores, em trabalhos sem fins lucrativos e de forma gratuita, que nos movimentos científicos traduziu-se por aberto (Open), como nos casos dos arquivos abertos, acesso aberto, dados abertos e outros. Assim, o *OpenAlex* poderia ser entendido como uma grande biblioteca aberta, que disponibiliza seu catálogo digital livremente, fazendo alusão à grande Biblioteca de Alexandria.

A Figura 1 apresenta a interface da plataforma *OpenAlex* com uma busca realizada pelo termo "ciência da informação". Na área central, estão listados os resultados correspondentes, organizados por relevância, destacando-se: "A ciência da informação" (2004, Yves-François Le Coadic), citado 217 vezes; "Epistemologia e Ciência da Informação" (1989, Marcela Zampronha Moraes), citado 65 vezes; e "Ciência da informação: origem, evolução e relações" (2008, Tefko Saračević), citado 108 vezes. Na lateral direita, a seção de estatísticas informa que foram encontrados 189.400 resultados, sendo 98,7% em acesso aberto (187.000 artigos disponíveis). Um gráfico apresenta a distribuição dos resultados por ano, indicando o volume de publicações ao longo do tempo, enquanto os tópicos mais relacionados incluem "Research Methods and Organizational Management", "Social Network Analysis in Education and Communication" e "Information Science and Society", entre outros. A interface oferece ainda ferramentas para refinar os resultados com filtros, realizar downloads ou salvar buscas.

Figura 1

Página de resultados da busca pelo termo "Ciência da Informação"



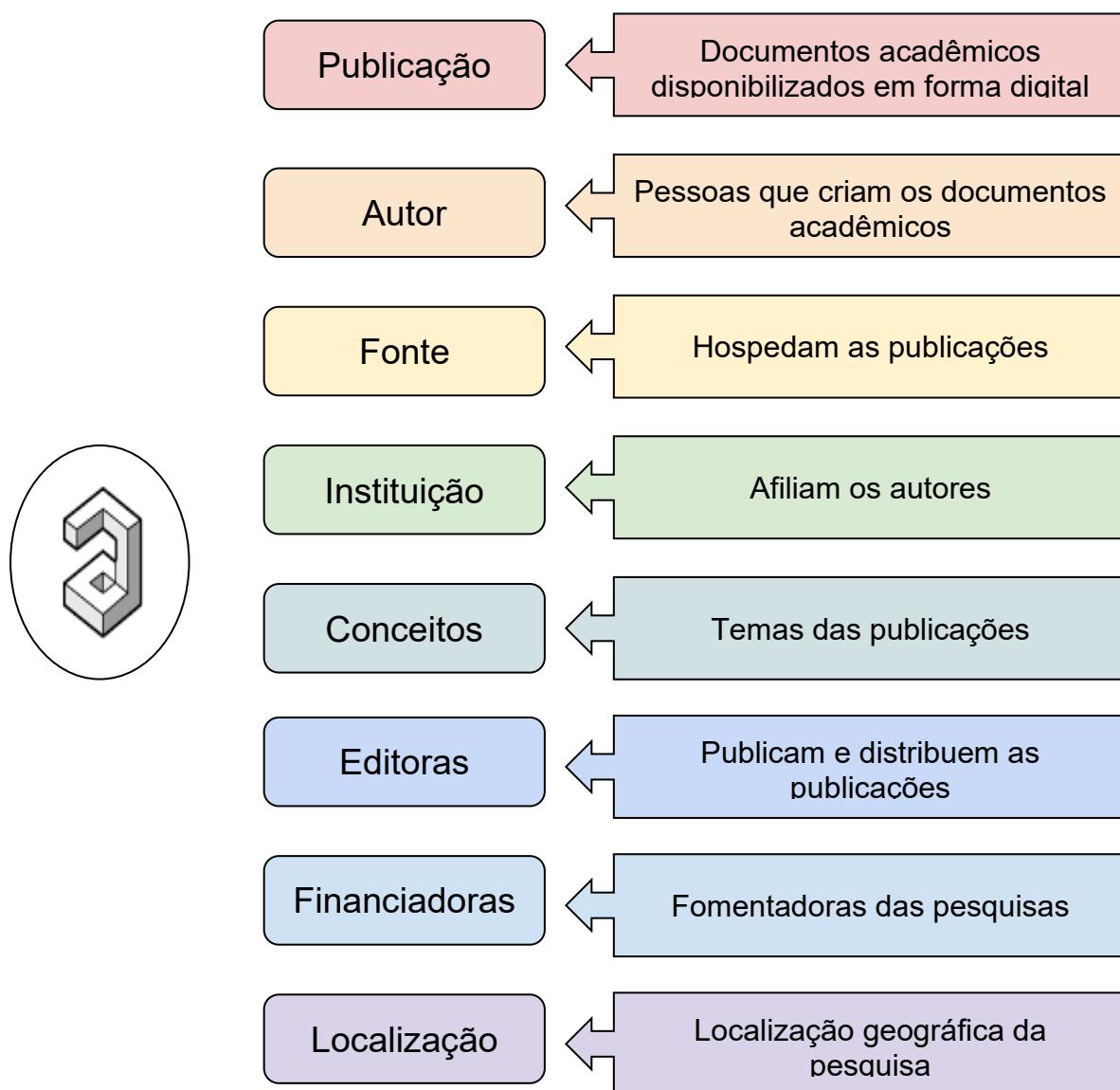
Nota. Fonte: Captura de tela da página openalex.org (2024). [Início da Descrição] A imagem apresenta a interface do OpenAlex buscando o termo "ciência da informação". No topo, está a barra de busca com o filtro fulltext inclui "ciência da informação". Na seção central à esquerda, encontra-se a lista de resultados de trabalhos acadêmicos. A ciência da informação (2004) foi citado 217 vezes. Epistemologia e Ciência da Informação (1989), com 65 citações. Ciência da informação: origem e evolução (2008), referenciado 108 vezes. Encontros Bibli: revista online sobre biblioteconomia e ciência da informação. À direita, está a seção Estatísticas: Total de 189.400 resultados. Gráfico de barras com a evolução anual. Indicador mostrando que 98,7% (187 mil trabalhos) estão em acesso aberto. Tópicos relacionados: Métodos de Pesquisa e Gestão Organizacional (10.480), Análise de Redes Sociais na Educação e Comunicação (9.440), Educação no Brasil: Políticas e Práticas (8.545), Direito Constitucional e Direitos Humanos (6.750) e Ciência da Informação e Sociedade (6.378). No canto superior direito, estão os botões de login, cadastro e ícones de configuração. [Fim da Descrição].

Demers (2024) denomina o OpenAlex como um Índice Aberto do Ecossistema de Pesquisa Global, que indexa uma grande quantidade de documentos científicos digitais, disponíveis de forma totalmente gratuita e aberta. Nesse sentido, Priem et al. (2022) relatam que o OpenAlex é semelhante aos grandes indexadores de conteúdo, como o Scopus ou Web of Science, mas que apresenta vantagens sobre essas bases pagas, especialmente em termos de inclusão, acessibilidade e disponibilidade. Para os autores, as três grandes vantagens do OpenAlex são: 1 - Cobertura, indexando grande parte das publicações, inclusive em várias línguas e do sul global; 2 - Facilidade de consulta; e 3 - Ser gratuito e aberto, com a licença mais aberta possível.

Conforme os autores citados anteriormente, na documentação técnica, o OpenAlex atua com oito entidades informacionais, como apresentado na Figura 2. Com isso, o OpenAlex cobre todos os elementos relacionados com os resultados da pesquisa, sendo as publicações a parte mais conhecida do ecossistema. Entretanto, para que uma publicação seja realizada são necessários diversos elementos. Um estudo científico é realizado por um ou mais autores, que são filiados a uma ou mais instituições, que está publicado em uma fonte e assim por diante.

Figura 2

Entidades informacionais do OpenAlex



Nota. Fonte: Elaborado pelos autores (2024). [Início da Descrição] A imagem exibe um diagrama em blocos colorido que ilustra os principais componentes do OpenAlex. À esquerda, está o logotipo do OpenAlex. À direita, há oito retângulos coloridos, cada um com uma entidade e sua definição. Publicação (na cor rosa): documentos acadêmicos disponibilizados em formato digital. Autor (na cor laranja): pessoas que criam os documentos acadêmicos. Fonte (na cor amarelo): locais que hospedam as publicações. Instituição (na cor verde): organizações que afiliam os autores. Conceitos (na cor azul claro): temas das publicações. Editoras (na cor azul): entidades que publicam e distribuem as publicações. Financiadoras (roxo azulado): fomentadoras das pesquisas. Localização (na cor roxo): localização geográfica da pesquisa. Setas apontam da entidade para a respectiva definição [Fim da Descrição].

Na maioria das bases indexadoras, as publicações estão no centro, uma vez que são esses elementos que se tornam a unidade informacional básica da base. Tradicionalmente, as publicações acadêmicas incluem monográficas (livros ou capítulos) e artigos (revistas científicas e eventos), mas podem assumir outras tipologias, como teses e dissertações, *preprints*, relatórios e pareceres, entre outros, que podem ser hospedados em fontes como repositórios. Outra fonte importante para o acesso aberto são as revistas, que não apresentam restrição ao acesso ao texto completo, mesmo aquelas que cobram para publicar, por meio dos chamados *Article Processing Charge* (APC). Assim, revistas de acesso aberto e repositórios se tornam as fontes prioritárias do OpenAlex.

4 Metodologia

A realização deste estudo ocorreu em duas fases principais. A primeira etapa envolveu uma análise descritiva do *OpenAlex*, visando compreender suas funcionalidades, interface e potenciais aplicações no âmbito das métricas informacionais. O acesso foi efetuado diretamente por meio da interface web da plataforma, utilizando-se do campo de pesquisa principal e dos menus de navegação. Neste estágio, foram analisadas as funcionalidades de busca por autor, publicação e instituição, além da utilização de filtros por período, tipo de documento (com foco em artigos de periódicos) e área de conhecimento (Ciência da Informação). Além da análise das alternativas de aprimoramento, foi observado o modo como os resultados são apresentados, englobando os metadados exibidos, como título, ano, periódico, afiliação institucional e número de citações, assim como os recursos disponibilizados para exportação de dados. A coleta foi efetuada de maneira manual, registrando-se os resultados de publicações e citações obtidas por meio das consultas, bem como verificou-se a consistência das informações através da repetição de buscas em períodos diferentes. Esse processo permitiu examinar a disponibilização das informações, a eficácia das ferramentas de busca, a facilidade de utilização dos filtros e a transparência na definição das métricas de citação, com o objetivo de avaliar a utilidade da plataforma como um recurso de apoio para pesquisadores de diversas áreas.

Na etapa subsequente, foi realizada uma análise comparativa entre o *OpenAlex* e as plataformas *Google Scholar Citations*, *Scopus*, *Web of Science* e *Dimensions*, visando avaliar as métricas relacionadas às publicações e citações disponibilizadas por cada uma dessas ferramentas. A seleção dessas plataformas fundamentou-se em sua extensa aplicação na área da avaliação científica (Bakkalbasi et al., 2006; Gerasimov et al., 2024; Martín-Martín et al., 2018), o que permite não apenas avaliar a viabilidade do *OpenAlex* como uma opção, mas também analisar sua complementaridade em relação a sistemas já estabelecidos no ambiente acadêmico.

Para a definição da amostra, adotou-se como critério a seleção dos dez autores mais citados no campo da Ciência da Informação, a partir do *Google Scholar Citations*. A pesquisa foi realizada utilizando o filtro por área, pela expressão: *label:ciência_da_informação*. O critério de dez autores foi implementado em virtude de assegurar a representatividade do grupo de maior relevância na área, no qual um número restrito de pesquisadores detém a maior parte das citações. Simultaneamente, essa abordagem proporciona viabilidade analítica, possibilitando comparações coerentes entre distintas plataformas, sem expandir de forma excessiva o escopo da coleta de dados. Após realizar a busca, os resultados são apresentados em forma uma lista organizada em ordem decrescente pelo número total de citações.

A partir dessa lista, foram selecionados os dez primeiros autores, independentemente de sua nacionalidade. No entanto, em um caso específico, o autor inicialmente identificado não possuía vínculo direto com a área de Ciência da Informação e, por esse motivo, foi substituído pelo autor subsequente da lista. Após a definição da amostra, os mesmos autores foram localizados manualmente nas plataformas *OpenAlex*, *Scopus*, *Web of Science* e *Dimensions*. Para cada um deles, registrou-se o número total de publicações e o número total de citações. Na análise, os nomes dos autores foram suprimidos, assegurando que os dados fossem tratados de forma imparcial. A coleta foi realizada em fevereiro de 2024 e, a partir dos dados obtidos, procedeu-se à comparação entre as cinco plataformas, com o objetivo de identificar convergências e divergências nos resultados referentes às métricas de publicações e citações.

Figura 3

Página de resultados da busca pelo termo "label:ciência_da_informação" no Google Scholar, com omissão dos nomes, informações e foto dos autores.

The screenshot shows a Google Scholar search results page with the query "label:ciência_da_informação". The results are filtered to show only profiles (Perfis). Each result includes a user icon (graduation cap), the name (redacted), a list of topics (redacted), and citation counts (e.g., "Citado por 1"). Below the results, there is a navigation bar with page numbers and arrows.

Nome (Redactado)	Topics (Redactado)	Citado por
ciência da informação		Citado por 1
	ciência da informação biblioteconomia arquivologia museologia information science	Citado por
		Citado por
Ciência da Informação		Citado por
	Ciência da Informação Gestão da Informação Gestão do Conhecimento Biblioteconomia Arquivologia.	Citado por
	Ciência da Informação	Citado por
		Citado por
	biblioteca digital ciência da informação informação científica estudo de usuários	Citado por
	Ciência da Informação Epidemiologia Qualidade da Informação	Citado por
	Ciência da Informação Organização e Representação ...	Citado por
	Ciência da Informação Biblioteconomia Arquivologia Museologia Mediação da Informação	Citado por
	ciência da informação comunicação científica epistemologia da ciência da info... interdisciplinaridade bibliomet	Citado por

Nota. Fonte: Captura de tela e edição da imagem pelos autores (2024). [Início da Descrição] A imagem exibe uma tela do Google Acadêmico em português, com busca pelo termo label: ciência_da_informação. No topo, está a barra de busca com o termo pesquisado. O título "Perfis" indica que os resultados são de autores. A lista contém dez perfis, cada um com um ícone de usuário usando um chapéu de formatura. Após cada ícone, surgem: Nome

do pesquisador (oculto na figura). As áreas de interesse do perfil incluem "Ciência da Informação", "Biblioteconomia", "Arquivologia", "Museologia", "Gestão da Informação", "Gestão do Conhecimento", "Estudo de Usuários", "Qualidade da Informação" e "Organização e Representação". Informe citações com "Citado por [número]". Na parte inferior da tela, aparece a paginação: "1 – 10", indicando que são os dez primeiros perfis encontrados [Fim da Descrição].

Foram selecionados os dez pesquisadores mais citados dessa lista, sem verificação da nacionalidade. Após essa seleção, esses mesmos autores foram localizados nas demais plataformas para a comparação do número total de publicações e citações. No entanto, verificou-se, a partir das publicações, que o primeiro autor identificado no Google Scholar não possuía relação com a área de Ciência da Informação e, por isso, foi descartado, sendo substituído pelo autor subsequente. Os nomes dos autores foram omitidos na análise, garantindo que os dados fossem tratados de forma imparcial.

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 cada plataforma. Essa comparação teve como objetivo analisar as métricas apresentadas por cada uma coleta de dados foi realizada em fevereiro de 2024. Para isso, foram inicialmente selecionados no *Google Scholar Citations* os perfis dos dez autores mais citados na área de interesse indicada pelo pesquisador em "Ciência da Informação", visando obter uma amostra também relevante nas demais bases. De maneira equivalente, os mesmos dez autores foram identificados no *OpenAlex*, *Scopus*, *Web of Science* e *Dimensions*, permitindo uma comparação efetiva da quantidade de publicações e citações. Após a seleção dos autores, procedeu-se à escolha das quantidades de trabalhos e citações associadas, com o intuito de comparar as métricas apresentadas nas plataformas. Os nomes dos autores foram suprimidos, pois não eram necessários para a análise.

5 Resultados e Discussões

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. Essa análise revelou que o *OpenAlex*, administrado pela entidade sem fins lucrativos *Our Research*, 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 (Assis et al., 2023). Essa 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 reúso 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 API REST (*Representational State Transfer*) 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 *Microsoft Academic Graph* (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 focada no volume de publicações e no número de citações de autores especializados em ciência da informação, selecionados a partir do *Google Scholar Citations*. O Quadro 1 detalha esses aspectos, permitindo observar diferenças importantes entre as métricas analisadas. Por exemplo, o autor A3 se destaca pelo maior volume de publicações no *Google Scholar Citations*, enquanto o autor A7 apresenta um número expressivo de citações no *OpenAlex*, mesmo sem ser o mais prolífico em termos de publicações. Essa disparidade evidencia que as métricas de desempenho acadêmico podem variar significativamente dependendo da base utilizada e do critério considerado, ressaltando a necessidade de avaliar tanto a quantidade de publicações quanto o impacto medido pelo número de citações para uma análise mais abrangente.

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

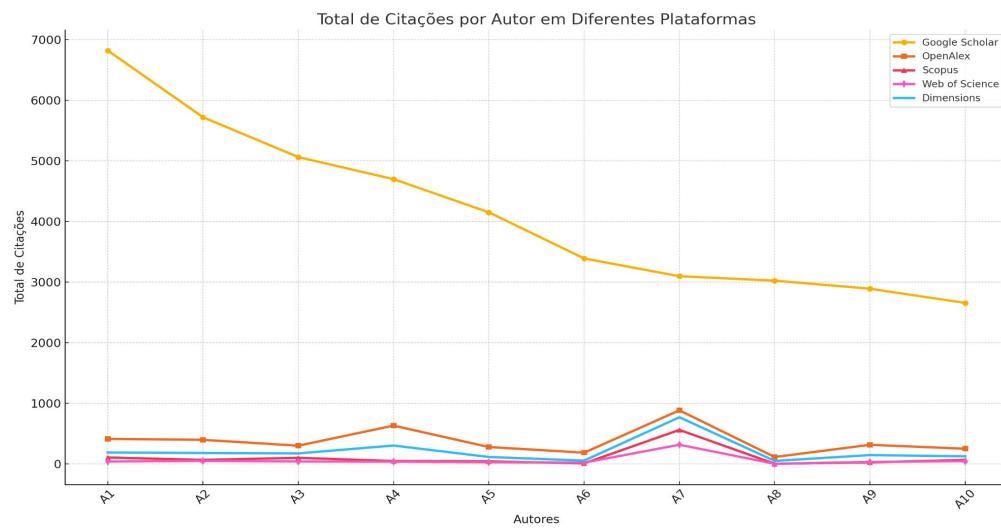
Nota. Fonte: Elaborado pelos autores (2024). [Início da Descrição] A imagem apresenta uma tabela comparando dez autores (A1 a A10) em cinco bases de dados — *Google Scholar*, *OpenAlex*, *Scopus*, *Web of Science* e *Dimensions* — indicando, para cada uma, a quantidade de trabalhos e o total de citações. O cabeçalho é laranja com texto branco, e os dados estão organizados em colunas e linhas para facilitar a comparação dos indicadores bibliométricos [Fim da Descrição].

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 da informação que é aberta, conforme discutido na metodologia do estudo que originou o Quadro 1. A presença significativa de alguns autores nessa base, como o autor A7 com 882 citações, destaca seu potencial para ampla disseminação e reconhecimento acadêmico.

Gráfico 1

Total de citações por autor em diferentes plataformas.

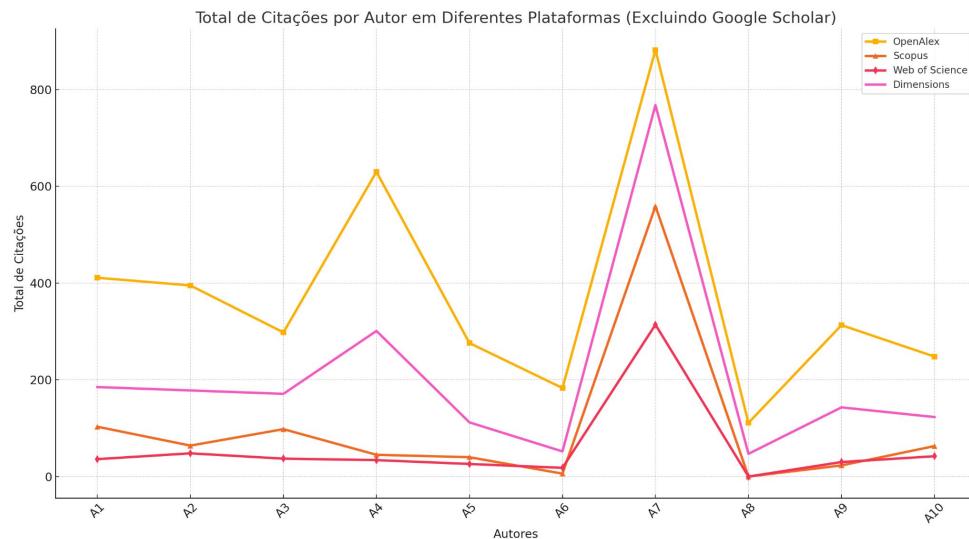


Nota. Fonte: Elaborado pelos autores (2024). [Início da Descrição] O gráfico de linhas tem o título “Total de Citações por Autor em Diferentes Plataformas”. Eixo X (horizontal): representa dez autores, identificados como A1 até A10. Eixo Y (vertical): mostra o total de citações, variando de 0 até 7000. As linhas coloridas representam as diferentes plataformas: Google Scholar (linha amarela): apresenta valores muito mais altos que as demais plataformas, variando de aproximadamente 6800 citações para o autor A1 até pouco acima de 3000 citações para o autor A10. OpenAlex (linha vermelha), Scopus (linha rosa), Web of Science (linha roxa) e Dimensions (linha azul-clara): mostram números bem menores de citações, geralmente abaixo de 1000, com pequenas variações entre os autores. Observa-se que as plataformas, exceto o Google Scholar, mantêm linhas próximas entre si, mas com diferenças pontuais, como no caso do autor A7, em que há um pico em todas as bases) [Fim da Descrição].

Para comparar o total de citações em diferentes plataformas, o Gráfico 1 destaca a discrepância nas citações da plataforma *Google Scholar Citations* em relação às demais. Apesar de terem sido selecionados os autores mais citados na área de Ciência da Informação na base do Google, o Gráfico 1 sugere que essa plataforma pode apresentar consideráveis discrepâncias, especialmente se for utilizada para análises bibliométricas.

Gráfico 2

Total de citações por autor nas plataformas: OpenAlex, Scopus, Web of Science e Dimensions.



Nota. Fonte: Elaborado pelos autores (2024). [Início da Descrição] O gráfico de linhas tem o título “Total de Citações por Autor em Diferentes Plataformas (Excluindo Google Scholar)”. Eixo X (horizontal): representa dez

autores identificados como A1 até A10. Eixo Y (vertical): mostra o total de citações, variando de 0 até 900. As linhas representam quatro plataformas: *OpenAlex* (linha amarela): apresenta os valores mais altos em quase todos os autores, destacando-se especialmente em A4 (cerca de 630 citações) e A7 (acima de 900). *Dimensions* (linha rosa): ocupa geralmente a segunda posição, com valores entre 150 e 300, e também mostra pico em A7 (cerca de 750). *Scopus* (linha laranja): aparece em terceiro lugar, com citações entre aproximadamente 20 e 120, alcançando pico em A7 (em torno de 550). *Web of Science* (linha vermelha): apresenta os valores mais baixos, variando entre 0 e 300, com pico em A7 (cerca de 300). Observa-se que todos os sistemas têm um aumento expressivo no autor A7, mas o *OpenAlex* e o *Dimensions* superam significativamente as demais plataformas [Fim da Descrição].

O Gráfico 2 revela uma certa homogeneidade na distribuição de citações por autor entre as diferentes plataformas, cabe destacar que o Gráfico 2 utiliza os mesmos dados do Gráfico 1, porém exclui o *Google Scholar*, de forma a permitir uma comparação de dados da mesma ordem de grandeza. No entanto, o *OpenAlex* apresenta valores de citação maiores em comparação com as demais. Isso se deve à sua capacidade de indexar um maior número de trabalhos e citações, mantendo, ao mesmo tempo, a qualidade e a confiabilidade de suas análises.

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.

É possível identificar tendências, como o aumento da adoção do acesso aberto, ao observar o crescimento das citações em plataformas como o *OpenAlex*. Isso está em sintonia com o movimento da ciência aberta e o acesso livre à informação científica, o que pode aumentar 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.

Este estudo apresenta limitações devido à ausência de avaliação da qualidade dos periódicos nas plataformas *OpenAlex* e *Google Scholar*. A falta de metadados que identificam a revisão por pares ou o *status* de *preprint* dos artigos também representa uma limitação. Além disso, existe a possibilidade de inclusão de periódicos predatórios e a dificuldade no controle de duplicações em algumas bases de dados. Essas limitações podem afetar a generalização dos resultados e a identificação precisa de periódicos de alta qualidade.

6 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 com outras plataformas estabelecidas, revela o papel significativo que o *OpenAlex* pode desempenhar 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, sugerem a capacidade do *OpenAlex* de oferecer uma alternativa viável e complementar às ferramentas tradicionais. Considera-se que a plataforma, com sua filosofia de código aberto e abordagem inclusiva, alinha-se com os princípios da ciência aberta e do acesso livre.

No entanto, o estudo também revela que, embora o *OpenAlex* seja promissor, ele ainda está começando a ser adotado pela comunidade acadêmica, e mais pesquisas são necessárias para entender suas limitações em relação a outras plataformas mais consolidadas. Desafios relacionados à deduplicação de dados e à variação na cobertura de determinadas áreas do conhecimento são questões que precisam ser monitoradas e pesquisadas.

Nesse contexto, surgem diversas possibilidades para trabalhos futuros, especialmente no que diz respeito a explorar como ferramentas abertas como o *OpenAlex* podem influenciar as políticas de avaliação científica, particularmente em áreas que tradicionalmente dependem de métricas baseadas em plataformas pagas. Além disso, o *OpenAlex* oferece uma oportunidade para o desenvolvimento de novas métricas bibliométricas, alinhadas

aos princípios da ciência aberta. Explorar novas abordagens para medir o impacto acadêmico, levando em consideração o compartilhamento aberto de dados, pode contribuir significativamente para o avanço da área.

Cabe destacar que o *OpenAlex*, assim como o Google Acadêmico, não tem o objetivo de avaliação de qualidade das publicações, enquanto as bases comerciais têm critérios para esse "filtro", falhos ou não. Essa ausência de um metadado claro sobre a natureza dos artigos, como se são preprints ou se passaram por revisão por pares, permite a inclusão de publicações de periódicos predatórios e aquelas que não foram submetidas a esse processo de revisão. Além disso, o aumento do volume de publicações e a complexidade inerente ao processo de publicação científica dificultam a implementação eficaz desses filtros. A evolução tecnológica, por sua vez, promete facilitar uma seleção mais precisa no futuro.

Por fim, o trabalho realizado sugere que iniciativas como o *OpenAlex* podem ser 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

- Assis, T. B., Costa, B. C. C., & Amaro, B. (2023). Projeto Pinakes: Uma reflexão para a construção de um ecossistema digital integrado de recursos bibliográficos técnicos e científicos. *Anais. XXII Seminário Nacional de Bibliotecas Universitárias*, Florianópolis.
<https://portal.febab.org.br/snbu2023/article/view/2957>
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(1), 7. <https://doi.org/10.1186/1742-5581-3-7>
- Costa, B. C. C., Santos, G. H., Viana, J. G. G., Nascimento, R. L. S., Santos, R. B., & Arakaki, A. C. S. (2024). Análise de dados bibliográficos: Estudo de caso no Catálogo Coletivo Nacional de Publicações Seriadas (CCN). *Anais. VII Workshop de Informação, Dados e Tecnologia - WIDaT*, Porto Velho.
<https://doi.org/10.22477/vii.widat.206>
- Costa, S. M. S. (2006). Filosofia aberta, modelos de negócios e agências de fomento: Elementos essenciais a uma discussão sobre o acesso aberto à informação científica. *Ciência da Informação*, 35(2), Artigo 2.
<https://doi.org/10.18225/ci.inf.v35i2.1139>
- Demes, K. (2024). *Open Alex*. University of Kentucky Libraries. <https://doi.org/10.13023/AXTF-GT03>
- Fonseca, E. N. (1973). Bibliografia estatística e bibliometria: Uma reivindicação de prioridades. *Ciência da Informação*, 2(1), 5–7. <https://doi.org/10.18225/ci.inf.v2i1.19>
- Garfield, E. (1955). Citation indexes for science; a new dimension in documentation through association of ideas. *Science (New York, N.Y.)*, 122(3159), 108–111. <https://doi.org/10.1126/science.122.3159.108>
- Gerasimov, I., Kc, B., Mehrabian, A., Acker, J., & McGuire, M. P. (2024). Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite. *Scientometrics*, 129(7), 3681–3704. <https://doi.org/10.1007/s11192-024-05073-5>
- Ginsparg, P. (2011). ArXiv at 20. *Nature*, 476(7359), 145–147. <https://doi.org/10.1038/476145a>
- Groos, O. V., & Pritchard, A. (1969). DOCUMENTATION NOTES. *Journal of Documentation*, 25(4), 344–349.
<https://doi.org/10.1108/eb026482>
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., & Hilf, E. R. (2008). The access/impact problem and the green and gold roads to Open Access: An update. *Serials Review*, 34(1), 36–40. <https://doi.org/10.1080/00987913.2008.10765150>
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177. <https://doi.org/10.1016/j.joi.2018.09.002>

Oddone, N. E., Shintaku, M., & Saldanha, G. S. (2023). Lydia de Queiroz Sambaquy: 1954 a 1965. Em D. A. P. Cunha (Org.), *Ibict 70 anos: Um resgate histórico daqueles que fizeram o instituto* (p. 18–32). Editora Ibict. <https://doi.org/10.22477/9786589167457.cap1>

OpenAlex. (2024). <https://openalex.org/>

Otlet, P. (1934). *Traité de Documentation: Le livre sur le livre*. Mundaneum.
https://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (N. arXiv:2205.01833). arXiv. <http://arxiv.org/abs/2205.01833>

Schötten, M., El Aisati, M., Meester, W. J. N., Steginga, S., & Ross, C. A. (2017). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. Em *Research Analytics* (p. 31–58). CRC Press. <https://doi.org/10.1201/9781315155890>

Dados de publicação

Diego José Macêdo

Mestre

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil

diegomacedo@ibict.br

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

Mestre em Ciência da Informação pela Universidade de Brasília. Bacharel em Sistema de Informação pela Universidade Católica de Brasília. Atualmente é tecnologista do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

Ingrid Torres Schiessl

Mestre

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil

ingridschiessl@ibict.br

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

Mestre em Ciência da Informação e bacharela em Biblioteconomia pela Universidade de Brasília (UnB). Bibliotecária e pesquisadora no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

Milton Shintaku

Doutor

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil

shintaku@ibict.br

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

Doutor em Ciência da Informação pela Universidade de Brasília. Coordenador de Tecnologia para Informação (Cotec) do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

Originalidade

Declaro que o texto é original e não está sendo revisado por nenhuma outra publicação

Preprint

O manuscrito não foi submetido a alguma plataforma de preprint.

Contribuição dos autores

Concepção e preparação do manuscrito: DJ Macêdo, IT Schiessl, M Shintaku

Coleta de dados: IT Schiessl

Análise dos dados: DJ Macêdo, IT Schiessl

Discussão dos resultados: M Shintaku

Revisão e aprovação: DJ Macêdo, M Shintaku

Uso de inteligência artificial

Não se aplica.

Financiamento

Não se aplica.

Permissão para usar imagens

Não se aplica.

Aprovação do Comitê de Ética em Pesquisa

Não se aplica.

Conflito de interesses

Não se aplica.

Declaração de disponibilidade de dados

Schiessl, Ingrid & Macêdo, Diego José (2025). Dados coletados para o estudo "O uso do OpenAlex nos estudos de métricas bibliométricas" [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.17282233>

Licença de uso

Os autores concedem à Biblios direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob uma Licença Creative Commons Atribuição (CC BY) 4.0 Internacional. Esta licença permite que terceiros remixem, adaptem e desenvolvam o trabalho publicado, dando os devidos créditos pela autoria e publicação inicial neste periódico. Os autores estão autorizados a firmar acordos adicionais separados para distribuição não exclusiva da versão publicada do trabalho no periódico (por exemplo, publicação em um repositório institucional, em um site pessoal, publicação de uma tradução ou como um capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

Editor

Publicado pelo Sistema de Bibliotecas Universitárias da Universidade de Pittsburgh. Responsabilidade compartilhada com universidades parceiras. As ideias expressas neste artigo são dos autores e não representam necessariamente as opiniões dos editores ou da universidade.

Editores

João de Melo Maricato, Janicy Aparecida Pereira Rocha e Lúcia da Silveira

Histórico

Recebido: 13-09-2024 – Aprovado: 29-09-2025 – Publicado: 10-12-2025



The articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by [Pitt Open Library Publishing](#).