

O uso de identificadores persistentes na recuperação da produção científica brasileira: análise de cobertura do Currículo Lattes na OpenAlex e ORCID

The use of persistent identifiers in the retrieval of Brazilian scientific production: analysis of the coverage of Currículo Lattes in OpenAlex and ORCID

El uso de identificadores persistentes en la recuperación de la producción científica brasileña: análisis de la cobertura del Currículo Lattes en OpenAlex y ORCID

Wellington Barbosa Rodrigues

Universidade Federal do ABC, Santo André, Brasil

Universidade de São Paulo, São Paulo, Brasil

Thamyres Vieira dos Santos

Universidade de São Paulo, São Paulo, Brasil

Hospital Israelita Albert Einstein, São Paulo, Brasil

Rogério Mugnaini

Universidade de São Paulo, São Paulo, Brasil

ORIGINAL

Resumo

Objetivo. Este estudo buscou analisar a recuperação da produção científica brasileira em ambientes informacionais internacionais e de acesso aberto, a partir da produção científica registrada nos Currículos Lattes de bolsistas de Produtividade em Pesquisa (PQ) do CNPq. Por meio dos identificadores persistentes DOI e ORCID, efetivou-se a busca nas fontes OpenAlex e ORCID. **Método.** Estudo exploratório da produção científica de 16.203 currículos de bolsistas PQ com ORCID registrado, e cerca de 1,1 milhão de registros de produção com DOI atribuído, do período de 2015-2024. Comparou o percentual de DOIs nas sobreposições entre as fontes Lattes, OpenAlex e ORCID, e também o grau de complementaridade entre as fontes. Considerou as seguintes variáveis secundárias: grande área, área específica da bolsa PQ e localização geográfica (Estado e Região) da afiliação institucional do pesquisador. **Resultados.** A totalidade da produção mostrou que cerca de 79% dos DOIs recuperados das três fontes estão representados na Plataforma Lattes, enquanto a OpenAlex complementa com 18,8% e a ORCID com 5,5%. Por outro lado, o maior percentual de DOIs encontra-se na intersecção entre o Currículo Lattes e a OpenAlex (37,9%), com destaque para as Ciências Agrárias, Ciências Biológicas e Ciências Exatas e da Terra, e na intersecção entre as três fontes (29,3%). As Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes destacam-se com os maiores percentuais advindos da OpenAlex, enquanto as Ciências de Saúde se destacam pelo percentual de DOIs exclusivos no Lattes (9,5%) e ORCID (2,8%). Entre as regiões, o Sudeste apresenta o maior percentual de DOIs provenientes da ORCID (2,1%) e na sobreposição Lattes e OpenAlex (40,1%). Por outro lado, o Centro-Oeste destaca-se com os maiores percentuais entre as fontes, sendo a região cujos DOIs presentes nos currículos foram mais complementados pela OpenAlex, o que pode significar que o identificador não tem sido registrado nos currículos. Ao comparar as áreas específicas e os estados, observou-se que, em ambos os casos, existe maior heterogeneidade no percentual de DOIs: na OpenAlex exclusivamente; na intersecção entre o Lattes e a OpenAlex; e na intersecção das três fontes. Dessa forma, observou-se que as áreas de Linguística, Letras e Artes, bem como de Ciências Agrárias, são as que

apresentam maior homogeneidade entre as suas subáreas, ao passo que nas Ciências Sociais Aplicadas se observa maior heterogeneidade. Entre os Estados, a maior heterogeneidade foi observada na região Norte, enquanto a região Sul é a mais homogênea, seguida da região Sudeste. **Considerações.** As diferenças em relação à cobertura das fontes, sugerem que as particularidades da comunicação científica, considerando o ecossistema editorial/informacional, divergem mais entre áreas do que entre regiões, o que concerne às características de publicação dos resultados de pesquisa. A sobreposição das produções entre as fontes evidencia o potencial de univocidade dos PIDs, possibilitando que pesquisas como esta sejam realizadas de maneira objetiva, sem que as variações de metadados textuais interfiram na qualidade dos resultados.

Palavras-chaves: identificadores persistentes, recuperação, fontes de informação abertas, cobertura, Plataforma Lattes

Abstract

Objective. This study sought to analyse the recovery of Brazilian scientific production in internationalised and open information environments, based on scientific production recorded in the Lattes Curricula of CNPq Research Productivity (PQ) scholarship holders. Using persistent identifiers DOI and ORCID, searches were carried out in the OpenAlex and ORCID sources. **Method.** Exploratory study of the scientific output of 16,203 CVs of PQ scholarship holders with registered ORCIDs, and approximately 1.1 million records of output with assigned DOIs, from the period 2015-2024. It compared the percentage of DOIs in the overlaps between the Lattes, OpenAlex and ORCID sources, as well as the degree of complementarity between the sources. It considered the following secondary variables: broad area, specific area of the PQ scholarship and geographical location (state and region) of the researcher's institutional affiliation. **Results.** The total production showed that around 79% of the DOIs retrieved from the three sources are represented on the Lattes Platform, while OpenAlex complements this with 18.8% and ORCID with 5.5%. On the other hand, the highest percentage of DOIs is found at the intersection between the Lattes Curriculum and OpenAlex (37.9%), with emphasis on Agricultural, Biological and Exact Sciences, and at the intersection between the three sources (29.3%). Human Sciences, Applied Social Sciences and Linguistics, Literature, and Arts stand out with the highest percentages coming from OpenAlex, while Health Sciences stand out for the percentage of exclusive DOIs in Lattes (9.5%) and ORCID (2.8%). Among the regions, the Southeast has the highest percentage of DOIs from ORCID (2.1%) and the Lattes and OpenAlex overlap (40.1%). On the other hand, the Midwest stands out with the highest percentages among the sources, being the region whose DOIs present in the curricula were most complemented by OpenAlex, which may mean that the identifier has not been registered in the curricula. When comparing specific areas and states, it was observed that, in both cases, there is greater heterogeneity in the percentage of DOIs: in OpenAlex exclusively; in the intersection between Lattes and OpenAlex; and in the intersection of the three sources. Thus, it was observed that the areas of Linguistics, Literature and Arts, as well as Agricultural Sciences, are the ones that present greater homogeneity among their subareas, while in Applied Social Sciences, greater heterogeneity is observed. Among the states, the greatest heterogeneity was observed in the North region, while the South region is the most homogeneous, followed by the Southeast region. **Conclusions.** The differences in source coverage suggest that the particularities of scientific communication, considering the editorial/informational ecosystem, diverge more between areas than between regions, which concerns the characteristics of publishing research results. The overlap in production between sources highlights the potential for univocity of PIDs, enabling research such as this to be conducted objectively, without variations in textual metadata interfering with the quality of the results.

Keywords: persistent identifiers, retrieval, open information sources, coverage, Lattes Platform

Resumen

Objetivo. Este estudio buscó analizar la recuperación de la producción científica brasileña en ambientes de información internacionalizados y abiertos, a partir de la producción científica registrada en los Currículos Lattes de los becarios de Productividad en Investigación de CNPq. La búsqueda se realizó utilizando los identificadores persistentes DOI y ORCID en las fuentes OpenAlex y ORCID. **Método.** Estudio exploratorio de la producción científica de 16,203 CV de becarios PQ con ORCID registrado, y alrededor de 1.1 millones de registros de producción con DOI asignado, entre 2015 y 2024. Comparó el porcentaje de DOIs en los solapamientos entre las fuentes Lattes, OpenAlex y ORCID, así como el grado de complementariedad entre las fuentes. Consideró las siguientes variables secundarias: área amplia, área específica de la beca PQ y localización geográfica (Estado y Región) de la afiliación institucional del investigador. **Resultados.** La producción total mostró que cerca del 79%

de los DOIs recuperados de las tres fuentes están representados en la Plataforma Lattes, mientras que OpenAlex complementa con el 18.8% y ORCID con el 5.5%. Por otro lado, el mayor porcentaje de DOIs se encuentra en la intersección entre el Plan de Estudios Lattes y OpenAlex (37.9%), especialmente en las Ciencias Agrícolas, Ciencias Biológicas y Ciencias Exactas y de la Tierra, y en la intersección entre las tres fuentes (29.3%). Las Humanidades, Ciencias Sociales Aplicadas y Lingüística, Letras y Artes destacan con los mayores porcentajes procedentes de OpenAlex, mientras que las Ciencias de la Salud sobresalen por el porcentaje de DOIs exclusivos en Lattes (9.5%) y ORCID (2.8%). Entre las regiones, el Sureste es la que presenta el mayor porcentaje de DOIs procedentes de ORCID (2.1%) y el solapamiento entre Lattes y OpenAlex (40.1%). Por otro lado, el Centro-Oeste destaca con los porcentajes más altos entre las fuentes, siendo la región cuyos DOIs en los CVs fueron más complementados por OpenAlex, lo que puede significar que el identificador no ha sido registrado en los CVs. Al comparar áreas y estados específicos, se observó que en ambos casos había mayor heterogeneidad en el porcentaje de DOIs: en OpenAlex exclusivamente; en la intersección entre Lattes y OpenAlex; y en la intersección de las tres fuentes. Como resultado, las áreas de Lingüística, Letras y Artes y Ciencias Agrarias son las más homogéneas entre sus subáreas, mientras que las Ciencias Sociales Aplicadas son las más heterogéneas. Entre los estados, la mayor heterogeneidad se observa en el Norte, mientras que el Sur es el más homogéneo, seguido del Sureste. **Conclusiones.** Las diferencias en la cobertura de las fuentes sugieren que las particularidades de la comunicación científica, considerando el ecosistema editorial/informativo, difieren más entre áreas que entre regiones, en cuanto a las características de publicación de los resultados de la investigación. El solapamiento de las producciones entre las fuentes pone de manifiesto la potencial univocidad de los PIDs, permitiendo que investigaciones como ésta se realicen de forma objetiva, sin que las variaciones en los metadatos textuales interfieran en la calidad de los resultados.

Palabras clave: identificadores persistentes, recuperación, fuentes de información abiertas, cobertura, Plataforma Lattes

1 Introdução

A informação científica é um ativo estratégico e o principal insumo utilizado para o progresso da ciência, uma vez que se torna um ponto de partida ou referência para outros pesquisadores na comunidade científica. Seu uso engloba o papel indispensável em disponibilizar insumo para a geração de novos conhecimentos até seu uso em processos avaliativos de instituições de ensino e pesquisa ou de pesquisadores. Em ambos os cenários de uso, as fontes de informação como periódicos, bases de dados, plataformas, redes acadêmico-científicas ou acervos físicos de serviços de informação de bibliotecas têm um papel crucial.

O produto da pesquisa científica é publicado em canais diversos, sendo disponibilizado nas fontes de informação, e portanto, passível de recuperação e uso. Meadows (1999, p. 161) afirma que “a realização de pesquisas e a comunicação de seus resultados são atividades inseparáveis”, uma vez que, “a visibilidade é uma característica desejável da comunicação científica, [...] ela representa a capacidade de exposição que uma fonte ou fluxo de informação possui de [...] ser acessada em resposta a uma demanda de informação” (Packer & Meneghini, 2006, p. 237).

Na perspectiva do usuário final, ou seja, aquele que deseja sanar suas necessidades informacionais, é necessário escolher uma fonte adequada, assim como definir uma estratégia de busca. Entretanto, tais escolhas implicam diretamente no conteúdo a ser recuperado. Entre as opções estão: a cobertura da fonte; a qualidade dos metadados de seus registros; e a chave de busca para recuperar o conteúdo desejado de forma inequívoca.

Dentre as fontes de informação científica, as bases de dados são caracterizadas por estabelecer uma política para indexação de produção. As bases de dados de periódicos científicos são as pioneiras, uma vez que as publicações periódicas são canais de comunicação estabelecidos a séculos. À medida que se proliferaram, com o crescimento da comunidade científica e a ênfase em avaliação de resultados, as políticas de indexação passaram a determinar critérios, que ora baseavam-se em aspectos temáticos, e crescentemente passaram a considerar o que entenderam corresponder a aspectos qualitativos.

Nos últimos 20 anos, as bases de dados multidisciplinares e comerciais Scopus e Web of Science (WoS) disputam a posição de referência para a literatura científica mundial, aplicando critérios de indexação próprios, contudo deixando parte significativa da produção científica mundial fora deste contexto (Cebrián et al., 2025). Ao longo do

tempo, após as evidências relatadas na literatura, essas bases de dados começaram a rever suas estratégias, de modo a ampliar sua cobertura e minimizar essa lacuna. Entretanto, estudos recentes ainda evidenciam a preferência de indexação de periódicos da Europa, Oceania e América do Norte na Scopus e WoS, o que segundo os autores pode se dever ao fato de serem regiões de maior concentração de renda e predomínio da língua inglesa (Asubiaro et al., 2024). Além da questão regional e linguística, é prudente comentar que a cobertura dessas fontes privilegia os periódicos científicos em detrimento de outros tipos de documento, como livros ou anais de eventos, o que implica em uma problemática para as áreas de Artes e Humanidades que, por sua vez, não priorizam o periódico científico como principal meio de publicação de seus resultados de pesquisa (Glänzel & Schoepflin, 1999; Mongeon & Paul-Hus, 2015).

A produção científica brasileira também é impactada pela restrição imposta por bases de dados comerciais, como Scopus e WoS; portanto, estudos da produção científica nacional não devem se restringir ao universo de indexação oferecido pelas fontes tradicionais, ainda que se despenda tempo para lidar com problemas e limitações nos dados oferecidos por fontes de natureza diversa (Mugnaini et al., 2019). No país, a Plataforma Lattes, mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por meio do Currículo Lattes, hospeda a produção científica de pesquisadores brasileiros. A atualização do currículo demanda a ação direta dos próprios pesquisadores, que são os responsáveis por inserir, uma a uma, suas produções na plataforma. Por requerer uma ação manual, é esperado que a produção científica brasileira não esteja em sua totalidade cadastrada na plataforma e que se tenha inconsistências no cadastro, uma vez que se trata de um formulário com campos de texto livre, com poucas possibilidades de automatização de preenchimento de dados advindos da plataforma Crossref (instituição responsável pela gestão do principal identificador utilizado para registro individual das publicações).

Nos processos avaliativos de pesquisadores brasileiros, o Currículo Lattes é comumente utilizado como a principal fonte de consulta ou coleta das informações de produção científica e técnico-científica. Pode-se intuir que, por esse motivo, exista uma preocupação latente por parte dos pesquisadores em manter essa fonte atualizada. Contudo, a ausência de mecanismos que possibilitem a atualização automática pode ser um dos motivos que faz com que os pesquisadores despendam mais tempo nesse processo de atualização e, diante da rotina das atividades de ensino e pesquisa, essa tarefa possa ser adiada e os dados disponíveis no currículo possam não ser os mais recentes.

Mena-Chalco (2025) divulgou o resultado de um estudo em que foi possível constatar que, num olhar macro, cerca de 46% dos autores presentes no Currículo Lattes não atualizaram seu currículo nos últimos 5 anos, culminando inclusive em distorções nos processos avaliativos de pesquisadores e instituições, uma vez que a fonte de dados não representa a realidade, além de prejudicar a visibilidade e compreensão da produção científica brasileira de forma geral.

A Plataforma Lattes, embora se consolide como a fonte mais abrangente de literatura científica brasileira, não representa necessariamente uma ferramenta de busca, oferecendo poucas possibilidades de manejo de dados pelo usuário final. Assim, o levantamento de dados na Plataforma sempre exigiu grandes esforços, principalmente pelo fato do CNPq não permitir o download massivo de informações. Nesse cenário, para que o mapeamento da produção científica brasileira não se restrinja à cobertura oferecida pelas bases de dados comerciais, é pertinente discutir a consulta desse universo de produções em outras fontes também compreensivas, a fim de observar se e quanto a produção brasileira está nelas coberta.

Existem outras fontes informacionais consideradas mais abrangentes, que além de não terem uma limitação comercial, são utilizadas por pesquisadores tanto para a pesquisa de literatura quanto para estudos bibliométricos, como, por exemplo, o Google Scholar, SciELO ou Dimensions. Em 2022, foi lançada uma nova fonte informacional chamada OpenAlex (Priem et al., 2022), baseada na estrutura do extinto Microsoft Academic Graph (MAG), com o objetivo de conferir acesso às publicações científicas, seus autores e instituições, se mostrando uma ferramenta de alto potencial considerando a interconexão entre diversas fontes de dados (Murillo-Gonzalez & López, 2023).

Não obstante à escolha da fonte, a partir de sua cobertura, está a estratégia de busca utilizada para recuperação de literatura. A partir das necessidades do usuário, a busca poderá ser construída considerando uma série de variáveis como assunto, título, periódico, autores ou instituições. Todos esses elementos são suscetíveis a variações de nomenclatura seja por erros cometidos ao longo do processo de publicação, desde o manuscrito até o processo de indexação, ou por uma simples falta de consenso a respeito da forma correta de realizar a descrição bibliográfica, impactando na não recuperação da informação desejada. Os casos mais comumente discutidos são relativos às variações nos nomes de autores e organizações, que podem ser determinantes, por exemplo, no

processo avaliativo de pesquisadores e instituições (Mugnaini et al., 2019). Entretanto, a recuperação de publicações por seu título também pode ser impactada por variações de idioma ou caracteres especiais, ou a recuperação por assunto, que é comumente sanada com a adoção de vocabulários controlados por algumas bases de dados, como o Medical Subject Headings (MeSH) na PubMed, especializada na área de saúde. Este problema é ainda mais relevante quando da identificação do título de periódicos em referências bibliográficas, para cômputo de citações (Mugnaini et al., 2020).

Para contribuir na recuperação eficaz de produções, considerando a busca por diferentes elementos, são discutidos os identificadores persistentes ou persistent identifiers (PIDs), “[...] que na prática são códigos alfanuméricos com distintas e específicas configurações de acordo com seu contexto e finalidade, otimizando a recuperação dos recursos que representam” (Santos, 2021, p. 85). Alguns comumente conhecidos na comunidade científica são: o Digital Object Identifier (DOI), que é utilizado para representar as publicações de forma geral; o Open Researcher and Contributor ID (ORCID), que é utilizado para representar os autores das publicações; e o Research Organization Registry (ROR), que é utilizado para representar as instituições presentes nas afiliações dos autores das publicações.

Embora os identificadores persistentes tenham um objetivo bem delineado, destaca-se o quanto é indispensável o engajamento e participação da comunidade científica em sua implementação. Há diversos fatores que podem corroborar em uma não clara compreensão dos identificadores persistentes, como a falta de treinamento ou divulgação por parte das instituições e seus respectivos serviços de informação, até a adoção indeterminada de diferentes identificadores, sem a devida orientação de uso e atualização, que pode culminar em um desinteresse por sua adoção (Cousijn et al., 2021).

É importante salientar ainda que a disponibilidade e uso dos PIDs pode ser prejudicada em regiões nas quais a infraestrutura de pesquisa esteja menos desenvolvida e que conte com recursos escassos. No Brasil, as regiões Centro-Oeste, Nordeste e Norte concentravam pouco mais que 15% da produção científica nacional, no período de 1992-1994, aumentando pouco sua participação entre 2007-2009 (quase 25%), e revelando forte dependência das regiões Sudeste e Sul, segundo dados da Plataforma Lattes (Sidone et al., 2016).

Diferentes fontes informacionais, assim como Scopus e WoS, possuem como chave de busca os identificadores persistentes DOI e ORCID. A OpenAlex realiza uma interconexão entre diversas fontes de dados bibliográficos, incluindo aquelas que utilizam os identificadores DOI, ORCID e ROR (Murillo-Gonzalez & López, 2023; Murillo & Zapata, 2023), proporcionando um cenário abrangente.

Tendo em vista o uso irrestrito da Plataforma Lattes como fonte abrangente para levantamento da produção científica e técnico-científica, mesmo com as limitações já mencionadas, torna-se indispensável questionar em que medida outras fontes informacionais poderiam ser utilizadas para o enriquecimento dos dados de pesquisa, como afiliação institucional de autores, classificação temática e contagem de citações, considerando as peculiaridades da produção nacional e as possibilidades de sua recuperação de forma inequívoca. Assim, considerando a cobertura abrangente proposta pela OpenAlex, o presente estudo tem como objetivo analisar a recuperação da produção científica brasileira de bolsistas de Produtividade em Pesquisa (PQ) do CNPq na OpenAlex e no ORCID, por meio dos identificadores persistentes DOI e ORCID, tendo como base o levantamento de publicações registradas no Currículo Lattes desses pesquisadores. Embora Mena-Chalco (2025) tenha apontado para a falta de atualização de quase metade dos currículos, seus resultados também mostram que os doutores são os que mais os atualizam, sendo que apenas cerca de 19% desse perfil se encontra desatualizado. Além disso, Digiampietri et al. (2014) já haviam mostrado que os bolsistas PQ figuram entre os grupos que atualizam seus currículos com maior frequência, o que justifica o recorte adotado.

Para situar o estudo em relação à literatura existente, foi realizado um levantamento exploratório em bases multidisciplinares (Scopus e Web of Science), complementado por buscas no Google Scholar. Foram utilizados termos em inglês e português relacionados à temática, como "OpenAlex", "ORCID", "DOI", "identificadores persistentes" e "cobertura", combinados com operadores booleanos. As buscas indicaram um número bastante reduzido de estudos diretamente relacionados à articulação entre OpenAlex, identificadores persistentes e recuperação da produção científica, mesmo com a ampliação e flexibilização das combinações de termos. Diante disso, o levantamento assumiu caráter não sistemático, com a finalidade de identificar estudos correlatos e delimitar a lacuna abordada neste artigo, privilegiando trabalhos que discutem cobertura de bases de dados e uso de identificadores persistentes em contextos bibliométricos.

Nesse contexto, o presente artigo contribui ao investigar a cobertura das produções registradas no Currículo Lattes na OpenAlex e no ORCID a partir da perspectiva dos identificadores persistentes, com atenção especial aos

processos de vinculação, desambiguação e interoperabilidade dos dados científicos. Estudos prévios têm se concentrado, sobretudo, na comparação da cobertura e das características estruturais de diferentes bases. Singh et al. (2021), por exemplo, analisam comparativamente Web of Science, Scopus e Dimensions, mostrando diferenças significativas na indexação de periódicos e na representação da produção científica entre essas fontes. Em direção semelhante, Madi et al. (2026) mostram que a OpenAlex apresenta cobertura superior da produção científica brasileira em relação a bases comerciais, com maior inclusão de periódicos nacionais, publicações em língua portuguesa e conteúdos em acesso aberto. Estudos no campo da Saúde também apontam maior cobertura da OpenAlex em comparação com outras fontes (Cebrian et al., 2025; Mezquita et al., 2025). Além disso, Schnieders et al. (2022) mostram que a discussão sobre cobertura e uso de identificadores persistentes já se coloca em contextos institucionais, embora sem focalizar a recuperação da produção registrada em bases curriculares e sua articulação específica com OpenAlex e ORCID.

O Quadro 1 sintetiza os estudos mais próximos identificados na literatura, permitindo explicitar seus focos analíticos e evidenciar a contribuição específica deste trabalho.

Quadro 1

Estudos correlatos sobre cobertura e identificadores persistentes

Estudo	Foco Principal	Bases analisadas	Uso de PIDs	Contribuição principal
Singh et al. (2021)	Comparação de cobertura entre bases	WoS, Scopus, Dimensions	Não central	Diferenças na indexação e impacto bibliométrico
Maricato et al. (2023)	Cobertura da SciELO em bases internacionais	Scopus, WoS, Dimensions, OpenAlex	DOI	Maior cobertura em Dimensions e OpenAlex
Schnieders et al. (2022)	Cobertura de ORCID para reporting institucional	ORCID, OpenAlex, FREYA PID Graph	Central (ORCID, DOI, ROR)	Avalia prontidão de fontes abertas e diferenças de cobertura
Madi et al. (2026)	Cobertura da produção brasileira	OpenAlex vs bases comerciais	Não central	Maior inclusão da produção nacional
Cebrian et al. (2025); Mezquita et al. (2025)	Cobertura em Ciências da Saúde	OpenAlex e outras bases	Não central	Maior abrangência da OpenAlex
Este estudo	Recuperação da produção do Lattes via PIDs	Lattes, OpenAlex, ORCID	Central (DOI, ORCID)	Vinculação, interoperabilidade e recuperação baseada em PIDs
Singh et al. (2021)	Comparação de cobertura entre bases	WoS, Scopus, Dimensions	Não central	Contribuição principal
Maricato et al. (2023)	Cobertura da SciELO em bases internacionais	Scopus, WoS, Dimensions, OpenAlex	DOI	Diferenças na indexação e impacto bibliométrico
Schnieders et al. (2022)	Cobertura de ORCID para reporting institucional	ORCID, OpenAlex, FREYA PID Graph	Central (ORCID, DOI, ROR)	Maior cobertura em Dimensions e OpenAlex
Madi et al. (2026)	Cobertura da produção brasileira	OpenAlex vs bases comerciais	Não central	Avalia prontidão de fontes abertas e diferenças de cobertura
Cebrian et al. (2025); Mezquita et al. (2025)	Cobertura em Ciências da Saúde	OpenAlex e outras bases	Não central	Maior inclusão da produção nacional

Estudo	Foco Principal	Bases analisadas	Uso de PIDs	Contribuição principal
Este estudo	Recuperação da produção do Lattes via PIDs	Lattes, OpenAlex, ORCID	Central (DOI, ORCID)	Maior abrangência da OpenAlex

Nota. Fonte: elaborado pelos autores (2025). *[Descrição do quadro]* O quadro apresenta seis estudos comparados quanto ao foco principal, bases analisadas, uso de PIDs e contribuição. Os estudos abordam temas como comparação de cobertura entre bases bibliográficas, cobertura da SciELO em bases internacionais, uso de ORCID para relatórios institucionais, cobertura da produção científica brasileira e cobertura da literatura em Ciências da Saúde. As análises envolvem bases como Web of Science, Scopus, Dimensions, OpenAlex e ORCID. A maioria dos estudos concentra-se na comparação de cobertura entre bases de dados, enquanto apenas parte deles utiliza os identificadores persistentes como elemento central da investigação. O presente estudo diferencia-se por investigar a recuperação da produção científica do Currículo Lattes por meio de DOI e ORCID, utilizando Lattes, OpenAlex e ORCID, com foco na interoperabilidade entre essas fontes e na recuperação baseada em PIDs. *[Fim da descrição]*.

Apesar dessas contribuições, tais trabalhos concentram-se principalmente na cobertura e nas características das bases, sem aprofundar o papel dos identificadores persistentes na recuperação e integração da produção científica. Desse modo, este estudo avança na compreensão de uma dimensão ainda pouco explorada na literatura consultada, especialmente no contexto brasileiro. Sua relevância reside em evidenciar, a partir da produção científica registrada no Currículo Lattes, o potencial e os limites dos identificadores persistentes como via de recuperação em fontes abertas, contribuindo para o debate sobre visibilidade, integração e qualidade dos dados da ciência brasileira.

2 Indexação e recuperação da produção científica em fontes informacionais por identificadores persistentes

As fontes informacionais possuem um universo limitado de materiais indexados, ou seja, no presente contexto, não há uma fonte informacional em que a produção científica global esteja indexada e tão logo seja passível de recuperação, sobretudo com o crescimento exponencial de literatura científica disponibilizada em ambiente digital. A escolha de uma fonte informacional considerada adequada às necessidades é um desafio para parte dos acadêmicos e pesquisadores que, por desconhecimento de suas limitações, podem ter seu universo de busca e recuperação limitados, além da qualidade de seus dados de pesquisa comprometidos. Tão importante quanto escolher a fonte, é delimitar a estratégia de busca utilizada visando uma recuperação eficaz.

Considerando o uso da recuperação da informação científica com objetivos avaliativos, Krüger (2020) afirma que existe uma competição por disponibilização de dados da literatura científica para gerar indicadores bibliométricos, sendo que o problema de maior destaque nesse cenário está na falta de clareza a respeito de sua metodologia de coleta e edição, uma vez que, de alguma forma, esse método se torna seu segredo de negócio.

O rigor dos critérios declarados como adotados para seleção de materiais elegíveis a serem indexados variam entre as fontes, mas comumente se trata de uma avaliação a respeito de aspectos variados de sua qualidade editorial e gráfica, tipo de conteúdo, cobertura geográfica e idioma (Castro, 2011). Ao estabelecer esses critérios, as conhecidas Scopus e WoS, que são fontes que mais possuem critérios formais de seleção estabelecidos em comparação com outras fontes como Google Scholar (Krüger, 2020), possuem deficiências semelhantes em sua cobertura, como em relação às áreas de Ciências sociais e Humanas, e a falta de literatura indexada em idiomas diferentes do inglês (Akness & Sivertsen, 2018).

A WoS foi criada em 1963 por Eugene Garfield sendo, na ocasião, chamada de Science Citation Index (SCI) e se consolidando ao longo do tempo como uma das fontes mais utilizadas para estudos bibliométricos (Danesh & Mardani-Nejad, 2022). Com a criação da Scopus e do Google Scholar em 2004, cresceu exponencialmente a quantidade de estudos que comparam a abrangência de cobertura entre essas duas fontes e a WoS, se tornando um dos maiores tópicos de discussão nas áreas de bibliometria e cientometria (Prancutê, 2021), incluindo também nos últimos anos alguns comparativos com as bases Dimensions e OpenAlex. Além disso, o uso do Google Scholar para estudos avaliativos, pela comunidade especialista em bibliometria e cientometria, evidencia a participação de autores de países não-anglófonos, e revela ser esta uma fonte frequentemente utilizada graças ao seu tamanho e diversidade documental, assim como pela disponibilidade da contagem de citações (Pereira &

Mugnaini, 2023). Contudo, a obtenção de dados em maior volume não é uma opção viável desta fonte, explicando a popularidade alcançada pela OpenAlex, em tão curto período de tempo.

A OpenAlex tem sido alvo de intenso interesse na comunidade científica e diversos estudos comparativos já foram publicados (Jiao et al., 2023; Velez-Estevez et al., 2023; Céspedes et al., 2025; Ortega & Delgado-Quirós, 2024; Turgel & Chernova, 2024; Culbert et al., 2025; entre outros estudos). Um estudo realizado por Alperin et al. (2024), em que é feito um comparativo entre a OpenAlex e a Scopus, os autores concluem que a primeira possui uma cobertura superior em relação à segunda e que pode ser utilizada como alternativa para a realização de análises bibliométricas, sobretudo análises à nível de país. Contudo, nesse mesmo estudo os autores mencionam que a acurácia e completude dos metadados na OpenAlex é um tópico que ainda precisa ser investigado na comunidade científica, com o objetivo de compreender melhor suas limitações e consequentes implicações nas análises realizadas e dados extraídos a partir dessa fonte.

Dentre os trabalhos já publicados, é possível encontrar artigos que, para traçar seu comparativo, utilizam os identificadores persistentes para realizar a recuperação da produção científica de forma inequívoca. Um exemplo seria o estudo de Maricato et al. (2023), que comparou a cobertura da produção científica da SciELO com a Scopus, WoS, Dimensions e OpenAlex, utilizando o DOI como chave de busca, concluindo que Dimensions e OpenAlex possuíram maior cobertura, com cerca de 95% e 93%, respectivamente, em comparação à Scopus e WoS, com aproximadamente 57% e 33%, respectivamente.

Os PIDs, além de identificarem inequivocamente os recursos que representam, corroboram para que esses recursos sejam localizados facilmente em ambiente digital ao longo do tempo, além de propiciar acesso aos metadados a eles vinculados, colaborando para uma interoperabilidade entre plataformas que vise beneficiar a infraestrutura da pesquisa científica de forma global. Quando se fala a respeito de identificadores persistentes relacionados ao universo da pesquisa, não se fala apenas em identificar o artigo, seus autores e instituições, mas também, por exemplo, as referências citadas ou dados de pesquisa. Nesse sentido, Mugnaini et al. (2020) afirmam que a presença dos DOIs nas referências citadas contribui com a promoção da transparência de dados relacionados, colaborando também com um movimento de abertura da pesquisa científica - e mais uma vez, fica latente que as áreas de Ciências Sociais e Humanas apresentam menor cobertura, mesmo considerando os hábitos de citação.

A adoção de identificadores persistentes no ecossistema científico, em especial o ORCID, tem sido estimulada por instituições, periódicos e agências de fomento, sobretudo em função de seu potencial para desambiguação de autoria e integração de dados. No entanto, sua efetividade depende diretamente do engajamento da comunidade científica, o que ainda se mostra desigual. Estudos indicam que parte dos pesquisadores apresenta resistência ou baixa adesão ao ORCID, frequentemente associada à falta de compreensão sobre sua finalidade, à percepção de aumento de carga administrativa ou à ausência de incentivos institucionais claros (Houghton & Foster, 2024). Além disso, a adoção do identificador não garante, por si só, a qualidade ou completude das informações, uma vez que o preenchimento e atualização dos perfis dependem de ação ativa dos próprios pesquisadores.

Esse cenário se agrava quando se observa que muitos perfis permanecem incompletos ou desatualizados, o que compromete o potencial dos identificadores persistentes para recuperação e integração da produção científica. Estudos mostram que uma parcela significativa de perfis ORCID não contém publicações vinculadas ou apresenta múltiplos registros para um mesmo autor, evidenciando limitações tanto na adoção quanto na manutenção dessas informações (Boudry & Durant-Barthez, 2020). Situação semelhante é observada no Currículo Lattes, em que a atualização depende de inserção manual e pode ser negligenciada ao longo do tempo (Mena-Chalco, 2025). Em conjunto, esses fatores indicam que, embora os identificadores persistentes representem uma solução técnica relevante para a organização da informação científica, sua efetividade está condicionada à qualidade dos metadados e ao grau de familiarização e engajamento dos pesquisadores com essas ferramentas.

Conforme explicitado anteriormente, a literatura oferece um panorama da cobertura de fontes informacionais além das tradicionais e também o uso dos PIDs na recuperação da informação científica. Contudo, são poucos os estudos como o de Maricato et al. (2023), que buscaram estudar o uso de PIDs para recuperação da produção científica brasileira da SciELO em fontes mais abrangentes, como a OpenAlex. Nesse sentido, evidencia-se a importância de estudos sobre a cobertura da produção científica registrada na Plataforma Lattes e seu mapeamento na OpenAlex e ORCID.

3 Metodologia

A revisão de literatura utilizada para justificar e embasar a pesquisa, foi inicialmente realizada com base em um levantamento bibliográfico nas fontes informacionais OpenAlex, Scopus, Web of Science, combinando termos como “persistent identifiers”, “OpenAlex”, “Currículo Lattes”, “coverage” e “database”, combinando entre eles os operadores booleanos “e” e “ou”, conforme a fonte e os termos buscados, uma vez que já era esperado que não seria localizado volume significativo de publicações que tratassem sobre a temática proposta para a pesquisa de forma tão específica. O objetivo do levantamento bibliográfico inicial era compreender o volume e as discussões existentes sobre as temáticas que envolvem a pesquisa na literatura e, a partir dela, foi realizada uma análise das referências, para que então pudessem ser agregados outros estudos à discussão. Também foram incluídas referências tradicionais que embasam conceitos de comunicação científica, além de um estudo publicado em uma rede de cunho acadêmico e profissional.

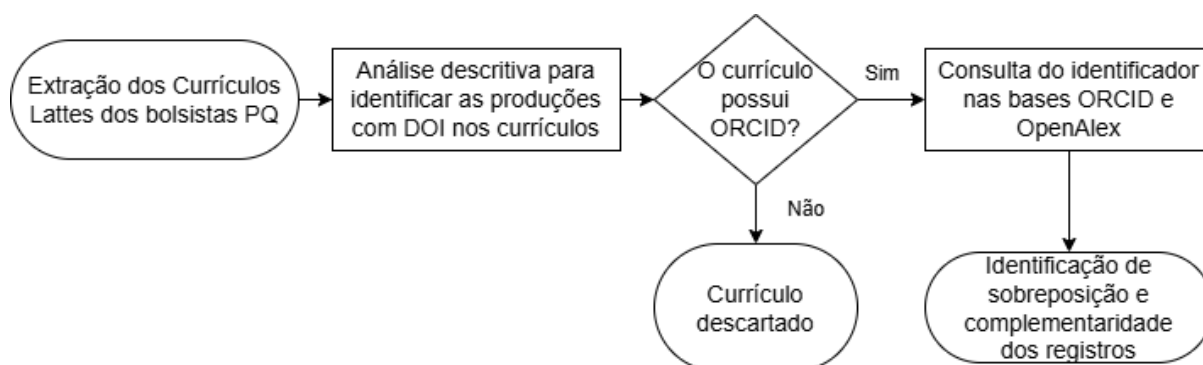
Em relação aos dados da pesquisa, a metodologia adotada teve início com a identificação dos Bolsistas Produtividade em Pesquisa do CNPq (PQ) por meio do painel do mapa de fomento em ciência, tecnologia e inovação, em fevereiro de 2025. A partir desses dados foram levantadas suas informações adicionais, como área, grande área e estado de afiliação. Em posse dessas informações, um total de 16.203 currículos Lattes foram baixados, por meio do programa scriptLattes (Mena-Chalco & Cesar Junior, 2009), no período compreendido entre fevereiro e março de 2025. Na sequência, procedeu-se à verificação da presença do identificador ORCID em cada um dos currículos, para tal, utilizou-se código próprio, desenvolvido em Python, para leitura do código HTML dos currículos baixados e identificação da tag indicativa de ORCID, para leitura do código HTML dos currículos baixados e identificação da tag indicativa de ORCID, bem como para a integração dos dados provenientes das diferentes fontes. Deste universo, 11.324 currículos (69,9%) continham o ORCID declarado e foram selecionados para as fases seguintes da análise. Os 4.879 currículos restantes, que não possuíam o identificador, foram descartados do escopo deste estudo, uma vez que a ausência do ORCID inviabiliza a recuperação automatizada e a ligação pretendida entre as fontes Lattes, ORCID e OpenAlex. Os dados processados estão disponíveis em repositório aberto, para futuras reproduções (Rodrigues et al., 2025). Com uso dos ORCIDs identificados nos Lattes, foi realizada a busca nas plataformas ORCID e OpenAlex, por meio das APIs oficiais, para recuperar as produções científicas associadas a cada pesquisador. Os DOIs encontrados nas três fontes foram reunidos para cada bolsista, sendo previamente normalizados por meio da remoção de prefixos (como <https://doi.org/> e [doi:](#)), padronização em minúsculas e eliminação de valores inválidos. Para formar o conjunto de dados final, realizou-se um processo de consolidação e deduplicação centrado no pesquisador: para cada bolsista PQ individualmente, os DOIs provenientes das três fontes foram agrupados e as duplicatas foram removidas. Assim, um mesmo DOI não foi contado múltiplas vezes para o mesmo pesquisador, mesmo que esteja presente em mais de uma fonte. A deduplicação não foi realizada entre pesquisadores. Logo, um mesmo DOI, correspondente a um artigo em coautoria, foi mantido no conjunto de dados final para cada um dos bolsistas PQ coautores presentes na amostra. Esta abordagem foi adotada, pois o objetivo foi verificar a recuperação da produção sob a perspectiva de cada pesquisador, reconhecendo que a presença e o registro de um DOI podem variar entre os currículos e perfis dos coautores, mesmo referindo-se à mesma publicação.

Publicações que possuíam DOI registrado no ORCID ou OpenAlex, mas que não constavam no Currículo Lattes de um determinado pesquisador, foram incorporadas à sua lista de produção, ampliando o conjunto de dados. Essa consolidação, focada no par pesquisador-DOI e respeitando o recorte temporal de 2015 a 2024 (pelas razões apresentadas adiante), resultou em um universo total de 1.111.147 registros de produção. Este conjunto representa a base de dados utilizada para as análises.

O estudo, portanto, não se limitou à identificação de sobreposições, mas buscou evidenciar também o grau de singularidade (ou complementaridade) entre as fontes, de modo a avaliar a efetividade dos identificadores persistentes na recuperação ampliada da produção científica brasileira em sistemas informacionais abertos. A Figura 1 apresenta o fluxograma que sintetiza as etapas de coleta, filtragem e análise, incluindo extração dos currículos Lattes, identificação de produções com DOI, verificação de ORCID, consulta às bases OpenAlex e ORCID e análise de sobreposição entre as fontes.

Figura 1

Fluxograma do pipeline de coleta, filtragem e análise dos dados



Nota. Fonte: elaborado pelos autores (2025). [Descrição da figura] Fluxograma do processo metodológico de coleta, filtragem e análise de dados. O processo inicia com a extração dos currículos Lattes dos bolsistas produtividade. Em seguida, realiza-se uma análise descritiva para identificar as produções com DOI presentes nesses currículos. Após essa etapa, há um ponto de decisão: verifica-se se o currículo possui identificador ORCID. Caso não possua, o currículo é descartado. Caso possua ORCID, o fluxo segue para a consulta desse identificador nas bases ORCID e OpenAlex. Por fim, os registros recuperados passam por uma etapa de identificação de sobreposição e complementaridade dos dados. [Fim da descrição].

Trata-se de um estudo exploratório, com vistas a mensurar: a) a representatividade dos identificadores únicos (DOI e ORCID) nos currículos dos bolsistas PQ registrados na Plataforma Lattes e demais fontes; b) o potencial de levantamento de produção científica complementar à registrada nos currículos, a partir dos identificadores persistentes, nas fontes ORCID e OpenAlex. As variáveis secundárias consideradas são: grande área (9 áreas CNPq) e área da bolsa PQ (84 áreas) e a localização geográfica (Estado e Região) da afiliação institucional do pesquisador. Essas informações foram obtidas a partir do painel do mapa de fomento em ciência, tecnologia e inovação e consideradas para todo o período conforme registradas no momento da coleta.

Um estudo prévio analisou a ocorrência dos ORCIDs e DOIs nos currículos dos bolsistas PQ, considerando as variáveis secundárias supracitadas. Constatou-se que 21,4% da produção intelectual até 2024 possui DOI. Tal percentual é de 35,4% entre 2015 e 2024 e apenas 11,9% entre 1900 e 2014. O tipo de produção com maior ocorrência de DOI são os artigos completos em periódicos (69,4%), seguido de artigos aceitos para publicação (20,9%) e capítulos de livro (11,3%). Adicionalmente, os dados mostraram que os currículos de bolsistas das áreas de Ciências Exatas e da Terra (área com maior percentual, com 35,3%), quando considerados o período mais recente apresenta aumento significativo (53,0%). E no extremo oposto, a área de Linguística, Letras e Artes apresenta o menor percentual (3,9%), aumentando para 8,1% quando delimitado o período. A região do país com maior percentual de DOI é a região Nordeste, com 23,6% quando se considera o período completo, e 38,5% filtrando pelo período mais recente. Os percentuais mais baixos são apresentados pelas regiões Centro-Oeste e Sul: 20,7% e 20,9%, respectivamente, com o período completo; e 33,8% e 33,7% no período mais recente. Como se pôde notar, a diferença do percentual entre as regiões não é tão pronunciada.

O ORCID, por sua vez, apresentou percentuais de incidência variados entre as grandes áreas. Ciências Agrárias, Ciências da Saúde e Outras (congrega áreas interdisciplinares como Biomedicina, Ciências Ambientais, Divulgação Científica, entre outras) são as que apresentaram maior percentual (acima de 75%), enquanto os menores percentuais são observados em Ciências Exatas e da Terra, Ciências Humanas e Linguística, Letras e Artes (abaixo de 65%). As regiões Centro-Oeste e Sul são as que apresentam os maiores percentuais (aproximadamente 80%) e a região Sudeste o menor (64,5%). Deve-se considerar que a região Sudeste, onde estão 56,3% dos bolsistas, devem ter os bolsistas mais antigos, o que pode explicar o menor percentual de ORCIDs declarados.

Finalmente, buscando garantir um maior percentual de DOIs disponíveis, optou-se por delimitar o período das produções, entre 2015 e 2024. Considerando o recorte temporal adotado, foi realizada uma verificação exploratória da distribuição das combinações entre as fontes, não sendo observadas alterações substantivas nos padrões ao longo do período analisado.

Para uma adequada contextualização e interpretação dos resultados que seguem, são destacadas algumas limitações metodológicas. A primeira delas se refere ao tratamento das publicações em coautoria. O estudo

contabilizou cada trabalho para todos os bolsistas PQ coautores presentes na amostra, sem realizar uma deduplicação global entre eles. Assim, ao se proceder agregações por áreas ou localização geográfica, um mesmo documento figura múltiplas vezes no conjunto de dados, associado a cada respectivo coautor, refletindo a perspectiva de recuperação individual da produção e a forma como cada pesquisador registra seus trabalhos.

Adicionalmente, a acurácia dos DOIs registrados na Plataforma Lattes é uma variável a ser considerada. A inserção destes identificadores é manual por parte dos pesquisadores, um processo que está sujeito a erros de digitação ou à inserção de identificadores inválidos. Tais incorreções podem comprometer o correto pareamento das publicações com os dados provenientes de ORCID e OpenAlex, com potencial impacto nas análises de sobreposição e singularidade entre as fontes.

Por fim, a própria estrutura da Plataforma Lattes impõe restrição quanto ao registro de DOIs para certos tipos de produção. A plataforma não disponibiliza campos específicos para a inserção deste identificador em produções que não sejam estritamente bibliográficas, como é o caso de produções técnicas, por exemplo. Essa característica impede a recuperação sistemática de tais itens via DOI a partir do Lattes, mesmo que possuam o identificador atribuído em outra fonte, podendo influenciar a avaliação da cobertura comparativa com as demais fontes que podem ter registrado esses DOIs.

Adicionalmente, é importante considerar que a adoção do identificador DOI não é homogênea entre os diferentes tipos de produção científica. Artigos em periódicos tendem a apresentar maior incidência desse identificador, enquanto capítulos de livro, livros e trabalhos em anais de eventos apresentam menor padronização. Nesse sentido, parte das diferenças observadas entre áreas do conhecimento pode refletir não apenas a cobertura das fontes analisadas, mas também a composição do portfólio de produção científica característico de cada área. Assim, áreas com maior predominância de artigos tendem a apresentar maior recuperação baseada em DOI, ao passo que áreas com maior diversidade de tipos documentais podem apresentar menor cobertura relativa, independentemente da fonte considerada.

4 Resultados

A análise da presença dos DOIs nas diversas fontes informacionais permite avaliar o grau de interoperabilidade proporcionado pelos PIDs, mas também a singularidade entre as mesmas. Mensurar a sobreposição e singularidade das fontes permite avaliar a utilidade de cada uma, ou em conjunto, para a cobertura adequada da ciência brasileira.

O fato de o Currículo Lattes ser autopreenchível e não haver ferramentas de importação automatizadas, pode ocasionar sua incompletude, por conta do simples esquecimento do lançamento de uma nova produção, ou até mesmo pela não declaração do DOI. É recorrente, por exemplo, a necessidade de registro da produção imediatamente à sua publicação, por conta do prazo limite de submissão de proposta a um edital, contudo um artigo pode estar disponível no site do periódico, mas sem que o respectivo DOI tenha sido atribuído até aquele momento. Contudo, isto não impede o lançamento da produção.

A OpenAlex funciona como um grande índice agregado, não sendo uma plataforma de preenchimento ou edição direta por parte dos pesquisadores. Ela constrói seu vasto agrupamento de conhecimento coletando e integrando dados de diversas fontes. Além do extinto MAG, suas fontes primárias incluem Crossref, ROR, Directory of Open Access Journals (DOAJ), dentre outras. A OpenAlex processa esses dados para criar entidades e estabelecer conexões entre elas. Seu modelo é puramente baseado na indexação e interligação, de maneira a oferecer uma visão ampla e conectada da produção científica global.

O ORCID, por sua vez, opera em um modelo híbrido. Nele, o pesquisador pode preencher manualmente suas informações e lista de publicações, mas seu grande diferencial é a capacidade de atualização automática por meio de fontes confiáveis. Ao conceder permissão, o pesquisador permite que editoras, bases de dados (como Scopus, Web of Science e Crossref) e outras instituições enviem automaticamente os dados de suas novas publicações para seu perfil ORCID quando uma nova publicação com o identificador é registrada, desde que o pesquisador autorize essa conexão. Isso cria um sistema que combina a curadoria e validação do pesquisador com a automação advinda do ecossistema de publicação científica, embora sua plena eficácia ainda dependa do engajamento ativo do usuário para criar e manter essas conexões.

Tanto OpenAlex como ORCID diferem significativamente do modelo exclusivamente auto-preenchível do Currículo Lattes. Enquanto a OpenAlex aposta na agregação massiva e algorítmica e o ORCID busca um equilíbrio entre

automação e controle do usuário, o Lattes depende da inserção manual e diligência do pesquisador. Essa diversidade de modelos impacta diretamente a recuperação da informação: a OpenAlex reflete o que é disponibilizado no ecossistema global, o ORCID reflete o que é declarado e conectado e o Currículo Lattes pode refletir o que o pesquisador declara (com possíveis omissões ou atrasos).

Essa heterogeneidade nos modelos de coleta e atualização de dados cria um cenário onde é improvável que qualquer uma dessas fontes, isoladamente, ofereça um retrato completo e sem falhas da produção científica. É esperado que existam tanto sobreposições, indicando publicações com boa visibilidade e interoperabilidade via PIDs, quanto singularidades nas fontes. Essas singularidades podem ocorrer por diversas razões: publicações regionais cadastradas no Lattes, publicações capturadas pela OpenAlex que o pesquisador ainda não registrou nas outras fontes, ou registros no ORCID que não foram propagados ou inseridos nas demais plataformas. Portanto, a análise dessas sobreposições e singularidades torna-se fundamental. Ela permite mensurar o grau de cobertura e singularidade entre Lattes, ORCID e OpenAlex e avaliar a eficácia real dos identificadores persistentes como pontes entre esses sistemas no contexto brasileiro. Compreender onde os dados convergem e onde divergem é essencial para identificar as forças e fraquezas de cada fonte e para dimensionar o potencial de ferramentas como a OpenAlex e ORCID na recuperação da literatura científica brasileira.

A literatura científica sobre a temática, utilizada na revisão de literatura do presente estudo, traz estudos que a discutem de uma forma mais ampla, sem adentrar na minúcia da indexação literatura nacional registrada no Currículo Lattes em fontes mais abrangentes, sendo recuperada por PIDs. Por esse motivo, os resultados apresentados não apresentam articulação específica com a literatura, mas são por ela contextualizados nas primeiras sessões do estudo.

A Figura 2 permite observar a distribuição dos DOIs entre as fontes nas diversas grandes áreas, onde fica evidente que o maior percentual de DOIs se encontra na intersecção entre Currículo Lattes e OpenAlex. Ciências Agrárias, Ciências Biológicas e Ciências Exatas e da Terra são as grandes áreas com maior percentual (cerca de 40%). E somando o percentual presente nas três fontes, essas mesmas áreas ultrapassam os 70%, denotando o potencial da OpenAlex para mapeamento das produções dos bolsistas dessas áreas.

Figura 2

Diagramas de Venn proporcionais, para comparação do percentual de produção identificado nas fontes, nas diversas grandes áreas



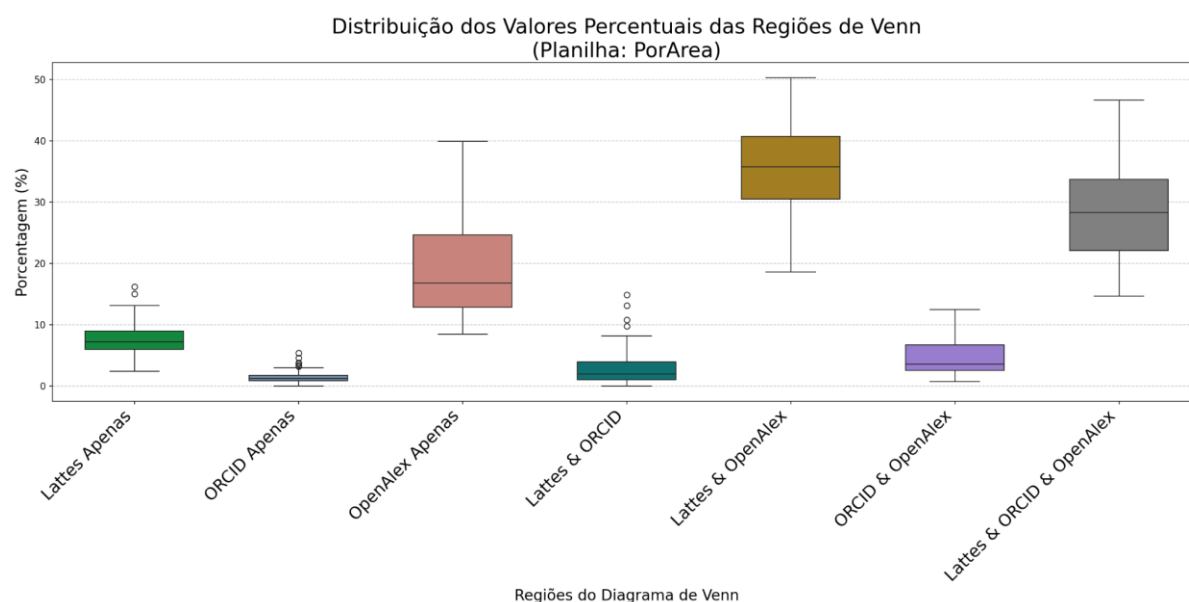
As Humanidades (Ciências Humanas, Ciências Sociais Aplicadas e Linguística Letras e Artes), por sua vez, apresentam os menores percentuais entre as grandes áreas, na intersecção das três fontes, assim como em quase todas as combinações que envolvem o ORCID. Fica latente a limitação do ORCID como fonte nessas áreas, o que também foi sinalizado pelo menor percentual de currículos destes bolsistas com o referido PID. Por outro lado, a exceção se dá na intersecção entre ORCID e OpenAlex, que evidencia os maiores percentuais, assim como os maiores percentuais de DOIs exclusivos na OpenAlex, denotando que a OpenAlex oferece um número significativo de DOIs nessas fontes — principalmente na Linguística, Letras e Artes.

As Ciências da Saúde apresentam o menor percentual exclusivo na OpenAlex, e por outro lado os maiores percentuais exclusivos nas demais fontes. Também chama a atenção o maior percentual apresentado na intersecção entre Lattes e ORCID. Assim, percebe-se ser esta a área em que a OpenAlex menos cobriu.

Fica clara a esperada diversidade entre as grandes áreas, cabendo analisar se no nível mais granular (das áreas específicas), as diferenças são discrepantes. A Figura 3 a seguir permite observar que a maior heterogeneidade entre as áreas se dá na intersecção entre Lattes e OpenAlex, na qual as cinco áreas com maiores percentuais (cerca de 50,3% a 47,8%) são: Probabilidade e Estatística, Biofísica, Engenharia de Minas, Parasitologia e Matemática. No extremo oposto, as cinco áreas com menores percentuais (cerca de 26,5% a 18,6%) são: Engenharia Elétrica, Ciência da Informação, Museologia, Robótica, Mecatrônica e Automação e Biologia Geral. Pode-se notar que o mapeamento exclusivo da produção do Lattes na OpenAlex, para algumas áreas específicas é bem incipiente, em comparação com outras.

Figura 3

Boxplot comparativo do percentual de DOIs das áreas específicas, conforme diferentes combinações das fontes



Nota. Fonte: elaborado pelos autores (2025). *[Descrição da figura]* Boxplots que apresentam a distribuição percentual de DOIs entre áreas específicas, considerando diferentes combinações das fontes Lattes, OpenAlex e ORCID. No eixo horizontal, estão as combinações analisadas (registros exclusivos e interseções entre as bases), e no eixo vertical, os percentuais correspondentes. Cada caixa representa a variação dos valores percentuais das áreas específicas para cada combinação de fontes. A combinação entre Lattes e OpenAlex concentra os maiores valores, indicando maior cobertura conjunta dessas bases. Combinações que envolvem o ORCID apresentam os menores percentuais. Observa-se maior variabilidade nas combinações com OpenAlex e presença de outliers em duas combinações envolvendo Lattes. *[Fim da descrição]*.

Ainda na Figura 3, outra combinação de fontes em que a heterogeneidade se pronuncia é na combinação entre as três fontes, em que podemos destacar as cinco áreas com maiores percentuais (cerca de 46,6% a 38,5%): Robótica, Mecatrônica e Automação, Farmácia, Fisioterapia e Terapia Ocupacional, Educação Física e Biologia Geral. E quanto às cinco áreas com menores percentuais (cerca de 16,0% e 14,7%): Museologia, Desenho Industrial, Geografia, Antropologia e Filosofia. Ou seja, a exigência da presença dos dois PIDs vai implicar num menor percentual para parte das áreas.

E dentre as fontes exclusivas, a OpenAlex é a que apresenta a maior heterogeneidade entre as áreas, tendo as seguintes cinco áreas com maiores percentuais (cerca de 40,0% a 32,7%): Museologia, Filosofia, Direito, Ciência da Informação e Geografia. Por outro lado, as áreas com menores percentuais (cerca de 9,5% a 8,5%) são: Farmácia, Física, Ciência e Tecnologia de Alimentos, Biotecnologia e Odontologia. Nesse caso, pode-se perceber que o grau de singularidade da OpenAlex é significativamente diminuto para parte das áreas.

A diferença entre as áreas específicas é menos pronunciada nas demais combinações de fontes, e como já se observou anteriormente, as demais combinações que envolvem o ORCID são pouco pronunciadas, revelando que esta fonte oferece pouca singularidade em relação às demais fontes. Contudo importa destacar as cinco áreas em que os percentuais de DOIs exclusivos no ORCID são mais pronunciados (cerca de 5,4% a 3,3%): Engenharia Elétrica, Física, Biologia Geral, Medicina e Enfermagem.

E finalmente, as cinco áreas com maior percentual exclusivo no Lattes (cerca de 16,2% a 11,8%) são: Serviço Social, Desenho Industrial, Museologia, Engenharia Nuclear e Enfermagem. E as com menor percentual (cerca de 3,5% a 2,4%) são: Ecologia, Morfologia, Oceanografia, Matemática e Robótica, Mecatrônica e Automação.

Uma análise da heterogeneidade das áreas específicas em cada grande área se faz importante (dados não apresentados). Linguística, Letras e Artes e Ciências Agrárias são as que mostram mais homogeneidade entre suas áreas específicas, em praticamente todas as combinações de fontes. Por outro lado, algumas áreas se destacam por apresentar maior heterogeneidade, como é o caso das Ciências Sociais Aplicadas, em cujas áreas específicas o percentual de DOIs exclusivos na Lattes varia muito, e principalmente na OpenAlex. As Ciências Biológicas, por sua vez, mostram maior heterogeneidade nas combinações entre Lattes e ORCID, e principalmente Lattes e OpenAlex. E as áreas que se destacam por apresentar maior heterogeneidade, em praticamente todas as combinações de fontes, são: Ciências da Saúde, Engenharias e Outras.

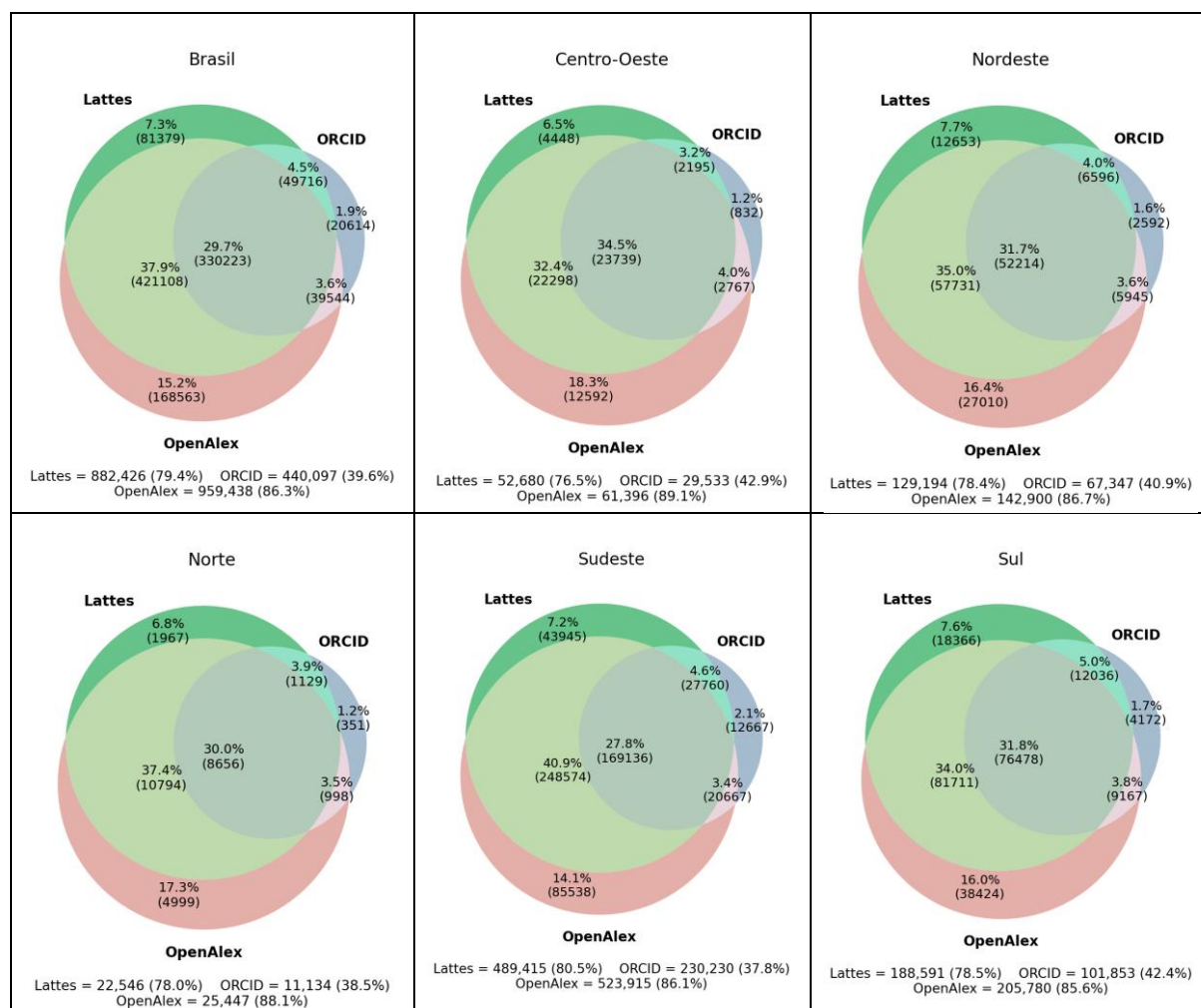
A Figura 4 apresenta a análise comparativa da sobreposição e singularidade das fontes para o conjunto de currículos de bolsistas afiliados a instituições das diferentes regiões do país e do Brasil como um todo. O conjunto completo de currículo dos bolsistas PQ brasileiros mostra que os maiores percentuais de DOIs se concentram na sobreposição entre Lattes e OpenAlex (37,9%) e nas três fontes (29,3%). Isto significa que, após recuperação nas outras fontes, o percentual de DOIs da Lattes passa a ser de 79,0%, e que as demais fontes contribuem com os seguintes percentuais de novos DOIs: 18,8% da OpenAlex e 5,5% da ORCID.

O Sudeste, sendo a região com 56,3% dos pesquisadores e 54,7% da produção científica considerada neste estudo, apresenta o maior percentual de DOIs exclusivamente na ORCID (2,1%) e o maior percentual na sobreposição entre Lattes e OpenAlex. Contudo é a região com menor percentual de singularidade na OpenAlex (14,6%), na sobreposição ORCID e OpenAlex (3,4%) e nas três fontes (27,8%). Ou seja, a região Sudeste é aquela cujos DOIs presentes nos currículos foram menos complementados pela OpenAlex, o que pode se dever ao fato do DOI ser mais frequentemente registrado no Currículo Lattes. Contudo, junto à região Sul, foi a mais beneficiada pela ORCID (em relação à Lattes), mesmo que os currículos do Sudeste tenham sido os que apresentaram menor percentual de ORCID registrado. Pode-se depreender que a porção de currículos com ORCID viabilizou a recuperação de um montante mais significativo de DOIs.

O Centro-Oeste é a segunda região com menor produção (6,2%), ficando atrás apenas da região Norte (2,6%). Na Figura 4 o Centro-Oeste é a que apresenta os maiores percentuais em mais combinações de fontes, o que evidencia que a região cujos DOIs presentes nos currículos foram mais complementados pela OpenAlex: OpenAlex exclusiva (18,3%), na sobreposição entre OpenAlex e ORCID (4,0%) e nas três fontes (34,5%). Por outro lado, é a região com menores percentuais, em todas as demais combinações. A região Norte, por sua vez, destaca-se por apresentar um menor percentual de complementaridade da ORCID, em relação à Lattes (4,7%).

Figura 4

Diagramas de Venn proporcionais, para comparação do percentual de produção identificado nas fontes, nas diversas regiões do Brasil e total



Nota. Fonte: elaborado pelos autores (2025). [Descrição do gráfico] Diagramas de Venn proporcionais que apresentam a distribuição de produções com DOI entre as fontes Lattes, OpenAlex e ORCID para o Brasil e suas regiões (Centro-Oeste, Nordeste, Norte, Sudeste e Sul). Cada diagrama mostra as proporções de registros exclusivos de cada fonte e suas interseções. Os diagramas também exibem valores percentuais e absolutos, evidenciando diferenças regionais na cobertura das fontes. Observa-se predominância da interseção entre Lattes e OpenAlex em todas as regiões. A OpenAlex também apresenta participação relevante em registros exclusivos, com variações entre as regiões. O ORCID, por sua vez, mantém baixa participação, especialmente contribuição exclusiva. [Fim da descrição].

As regiões Sul e Nordeste são as que apresentam percentual mediano de produção (21,6% e 14,8%, respectivamente) e as com maiores percentuais de DOIs exclusivos na Lattes (cerca de 7,5%). Os currículos de bolsistas do Sul foram os que apresentaram maior percentual de DOIs registrados, fazendo com que a sobreposição entre Lattes e ORCID seja a mais pronunciada entre as regiões (5,0%).

Como se pôde notar, as diferenças entre as regiões não são tão marcadas como observado entre as grandes áreas, restando analisar a heterogeneidade do percentual de DOIs nas diversas combinações de fontes, entre os Estados.

Como observado na Figura 3, sobre as áreas específicas, a Figura 5 mostra pouca variação, quando analisados os box-plots dos Estados. A principal diferença está na distribuição percentual da singularidade da fonte OpenAlex, cuja heterogeneidade é significativamente menor em relação às áreas, tendo Roraima e Rondônia como outliers superiores (34,0% e 29,7%, respectivamente), seguidos de Sergipe, Distrito Federal e Mato Grosso (cerca de 19,7% a 18,9%), por outro lado, os Estados com menores percentuais são Amazonas, Rio de Janeiro, Acre, São Paulo e Amapá (cerca de 14,6% a 13,1%).

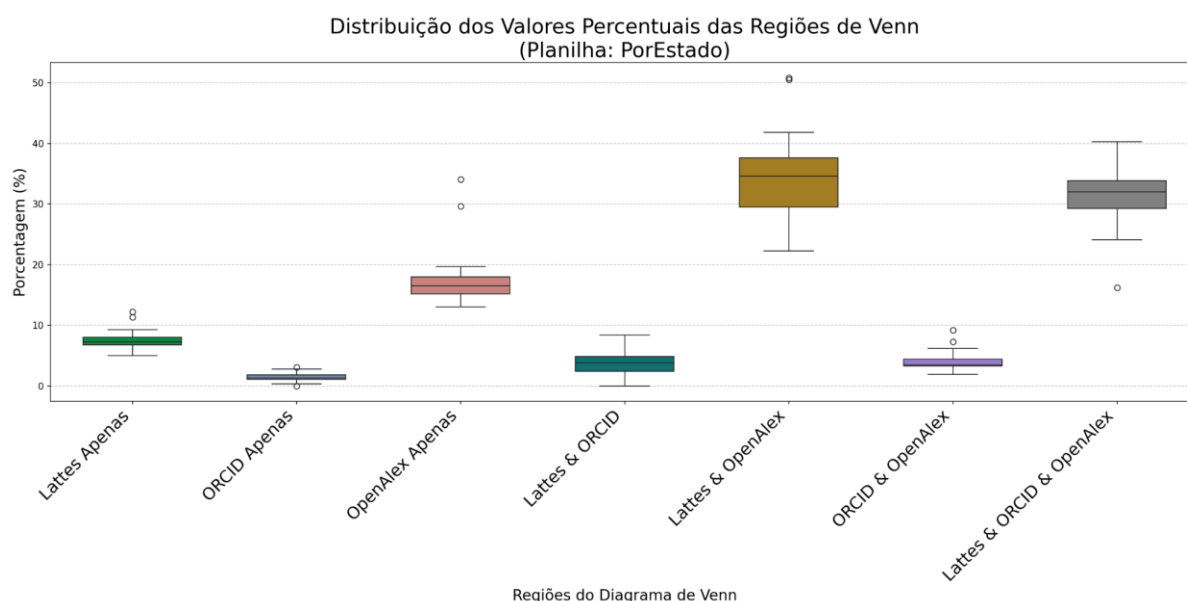
Na sobreposição entre Lattes e OpenAlex os maiores percentuais são de dois outliers superiores (Amapá com 50,8% e Acre com 50,5%), seguidos de Rio de Janeiro, São Paulo e Pará (cerca de 41,8% a 40,7%). Enquanto os Estados com menor percentual (cerca de 26,7% a 22,3%) são: Piauí, Maranhão, Tocantins, Rondônia e Sergipe.

Na sobreposição entre as três fontes, Sergipe, Maranhão, Tocantins, Mato Grosso e Mato Grosso do Sul lideram com os maiores percentuais (cerca de 40,3% a 36,1%). Ao passo que os menores percentuais (cerca de 27,7% a 16,2%) foram observados para Ceará, Acre, Amapá, Rio de Janeiro e Roraima.

A singularidade da Lattes é marcada pelos que apresentam maior percentual (cerca de 12,2% a 8,5%): Espírito Santo, Roraima, Ceará, Paraíba e Tocantins. E os de menor percentual (cerca de 6,1% a 5,0%): Rio Grande do Norte, Bahia, Sergipe, Distrito Federal e Rondônia.

Figura 5

Box-plot comparativo do percentual de DOIs nos Estados, conforme diferentes combinações das fontes



Nota. Fonte: elaborado pelos autores (2025). *[Descrição da figura]* Boxplots que apresentam a distribuição dos percentuais de DOIs nos Estados brasileiros, considerando diferentes combinações das fontes Lattes, OpenAlex e ORCID. No eixo horizontal, estão as combinações analisadas (registros exclusivos e interseções entre as bases), e no eixo vertical, os percentuais correspondentes. Cada caixa representa a variação dos valores percentuais dos Estados para cada combinação de fontes. A combinação entre Lattes e OpenAlex concentra os maiores percentuais. As combinações envolvendo apenas o ORCID apresentam os menores valores, com variação moderada entre os estados. *[Fim da descrição]*.

Observando a heterogeneidade entre os Estados (dados não apresentados), o Norte está na liderança em quase todas as combinações de fontes, revelando grande diferença entre seus estados. Só ficando atrás do Sudeste, quando considerada a exclusividade na Lattes, o que é causado pelos extremos: Espírito Santo (com 12,2%) e Rio de Janeiro (7,9%). Outra região com maior heterogeneidade entre seus Estados é a Nordeste. Enquanto a Sul é a mais homogênea na maioria das combinações de fontes, perdendo apenas para a Sudeste, tanto na exclusividade da OpenAlex quanto na sobreposição entre OpenAlex e ORCID.

5 Considerações

O estudo tem respaldo teórico pertinente à área de Ciência da Informação, ao apontar fragilidades e potencialidades da infraestrutura de comunicação científica aberta, que pode proporcionar visibilidade e equidade do conhecimento gerado pela pesquisa desenvolvida no Brasil. Nesse âmbito, fontes descentralizadas e regionais se contrapõem às bases de indexação comerciais que marginalizam a produção científica em português, como se observa na ciência brasileira e de diversos países não anglófonos. Além disso, os PIDs são elementos essenciais para a padronização e interoperabilidade da informação científica, servindo para a inclusão e rastreamento do

impacto da pesquisa. Em tal perspectiva, destaca-se como a tecnologia, a política de metadados e os sistemas de publicação influenciam a circulação e o acesso ao conhecimento gerado no país.

Discutiu-se a recuperação da produção intelectual enquanto um desafio para a comunidade acadêmico-científica, seja pela inexistência de fontes informacionais que indexem todo o conteúdo publicado, ou pelas dificuldades em sua recuperação inequívoca. Tratando-se de processos de recuperação que envolvem produções em nível nacional, os desafios se intensificam, sobretudo considerando a baixa representatividade de indexação de conteúdos em determinadas fontes. Assim, a recuperação inequívoca é assegurada quando utilizados os PIDs, que representam as produções intelectuais e alguns de seus elementos descritivos.

Nesse contexto, além de considerar a relevância do Currículo Lattes na comunidade científica brasileira, a amplitude de cobertura da OpenAlex e a relevância dos identificadores DOI e ORCID, a discussão proposta pela pesquisa mostra que a diferença entre os modelos de coleta e atualização de dados do Currículo Lattes, OpenAlex e ORCID faz com que seja improvável que apenas uma dessas fontes indexe toda a produção científica nacional. Sendo fundamental a mensuração da sobreposição e singularidade de cobertura da produção intelectual entre essas fontes, utilizando os PIDs como elementos para efetivar a análise.

Entretanto, a mencionada sobreposição e singularidade da cobertura possui suas particularidades a depender das áreas de conhecimento e também da localização geográfica da afiliação institucional do bolsista PQ. Portanto, a análise privilegiou a exploração dos contextos particulares de cada área e localização, assim como seus níveis de heterogeneidade.

Uma vez que o estudo teve como objetivo analisar a recuperação da produção científica brasileira de bolsistas PQ do CNPq na OpenAlex e ORCID, por meio dos identificadores DOI e ORCID, os resultados, considerando o conjunto de seus currículos, mostraram que a OpenAlex é a fonte com maior percentual de complementaridade à Lattes, contribuindo com um percentual de 18,8% novos DOIs, enquanto ORCID contribuiu com apenas 5,5%. O modelo de atualização da ORCID, conforme discutido, exige participação ativa dos pesquisadores, demandando sua ação no preenchimento ou na autorização para inserção automática de novas produções por fontes terceiras, fato que naturalmente pode implicar uma cobertura mais tímida justificando seu baixo percentual de complementaridade. A OpenAlex, enquanto índice agregador de diferentes fontes, mostrou cobertura superior, por não demandar ação do pesquisador.

As grandes áreas mostraram que o maior percentual de DOIs das produções dos bolsistas PQ está na intersecção entre Lattes e OpenAlex, evidenciando o potencial da OpenAlex na cobertura da produção nas áreas de Ciências Agrárias, Ciências Biológicas e Ciências Exatas e da Terra. As Humanidades apresentaram o menor percentual de bolsistas com o identificador ORCID e a baixa representatividade de sua produção nessa fonte informacional. Embora apresentem o menor percentual na intersecção entre as três fontes, é significativo o número de produções exclusivas na OpenAlex, sobretudo na Linguística, Letras e Artes. Por sua vez, as Ciências da Saúde foram a área em que a OpenAlex menos apresentou complementaridade na cobertura, mas é a área com maior percentual exclusivo em Lattes e ORCID, apresentando o maior número de produções presentes concomitantemente nas duas fontes.

A diferença em relação à cobertura evidencia as particularidades entre as áreas do conhecimento, suas características de publicação de seus resultados de pesquisa e, com isso, apresenta comportamentos diferentes em relação ao envolvimento com determinadas fontes informacionais. Como ficou evidente o caso das Humanidades, com menor número de bolsistas com ORCID e, conseqüentemente, menor produção nessa fonte; além de ser interessante notar seu potencial de singularidade na OpenAlex, tornando-a relevante para maximizar a cobertura da produção nacional nos estudos de tais áreas do conhecimento.

Em relação às regiões do país, observou-se que: a região Sudeste, enquanto região que possui mais da metade dos pesquisadores e produções consideradas neste estudo, é aquela cujos DOIs presentes nos currículos foram menos complementados pela OpenAlex; a região Sul, que apresentou o maior percentual de DOIs registrados, foi a mais complementada pela ORCID; o Centro-Oeste, enquanto segunda região com menor produção, foi a região mais complementada pela OpenAlex; a região Norte, região com menor produção, apresentou a menor complementaridade a partir da ORCID; e as regiões Sul e Nordeste, com um percentual mediano de produções, apresentaram a maior quantidade de DOIs exclusivos no Currículo Lattes.

Em seu olhar macro, os resultados mostraram que as diferenças em relação à cobertura das fontes, e suas singularidades, foram mais evidentes entre as grandes áreas do conhecimento do que entre as regiões da afiliação dos bolsistas PQ, sugerindo que as particularidades da comunicação científica e o ecossistema

editorial/informacional divergem mais entre áreas do que entre regiões. Contudo, é importante destacar que, por se tratar dos currículos de bolsistas deste nível acadêmico, além da já conhecida maior atualidade das informações declaradas e produtividade em pesquisa, espera-se que apresentem maior maturidade acadêmico-científica, incluindo a consciência das práticas recentes de comunicação científica, que incluem a utilização dos identificadores persistentes. Ao mesmo tempo, as práticas destes pesquisadores acabam servindo de modelo para sua comunidade científica, incluindo orientandos e gerações futuras.

A sobreposição das produções entre as fontes evidencia o potencial de univocidade dos PIDs, tornando possível que pesquisas como esta sejam realizadas de maneira objetiva, sem que as variações nos títulos ou nomes de autor interfiram na qualidade dos resultados.

O estudo se limita à produção científica dos bolsistas PQ, com ORCID declarado, na última década, ou seja, se trata de um recorte da produção nacional. Nesse sentido, há possibilidade de que novos estudos, considerando outra população como alvo de análise, possam contribuir com outros resultados à temática, em cenário nacional, além do potencial comparativo evolutivo em aspectos como, por exemplo, o ORCID declarado nos currículos ou a modernização do método de preenchimento do currículo.

A estrutura da Plataforma Lattes também é um aspecto a ser considerado no contexto da presente pesquisa, uma vez que a inserção do identificador persistente DOI nas produções é manual, sujeita a erros, e limitada àquelas de caráter bibliográfico. Uma vez que os identificadores foram utilizados como chave de recuperação, é esperado que os erros e limitações ocasionadas pela estrutura da plataforma interfiram no correto pareamento das publicações com as fontes ORCID e OpenAlex, impactando nos resultados das análises realizadas.

Dessa forma, assim como os resultados, as limitações do estudo também apontam para a necessidade de aprimoramento de fontes informacionais, para que possibilitem uma maior interoperabilidade de dados e menor incidência de erros ocasionados por preenchimentos manuais de PIDs, além do amplo incentivo e adoção dos identificadores na comunidade científica, tendo em vista seu inequívoco potencial de revocação, fazendo com que necessidades informacionais de diferentes áreas do conhecimento, instituições ou pesquisadores sejam sanadas de maneira propositiva.

Referências

- Akness, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus e Web of Science. *Journal of Data and Information Science*, 4(1), 1-21. <https://www.i-jdis.com/EN/10.2478/jdis-2019-0001>
- Alperin, J. P., Portenoy, J., Demes, K., Larivière, V., & Hustein, S. (2024). An analysis of the suitability of OpenAlex for bibliometric analyses. [Preprint] *arXiv*. <https://arxiv.org/abs/2404.17663>
- Asubiaro, T., Onaolapo, S., & Mills, D. (2024). Regional disparities in Web of Science and Scopus journal coverage. *Scientometrics*, 129, 1469-1491. <https://link.springer.com/article/10.1007/s11192-024-04948-x>
- Boudry, C., & Durant-Barthez, M. (2020). Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *PLoS ONE*, 15(9), Article e0238583. <https://doi.org/10.1371/journal.pone.0238583>
- Castro, R. (2011). Indexação de revistas científicas em bases de dados. In D. A. Población, G. P. Witter, L. M. S. V. C. Ramos, & V. M. B. O. Funaro (Orgs.), *Revistas científicas: Dos processos tradicionais às perspectivas alternativas de comunicação* (pp. 101-216). Ateliê Editorial.
- Cebrián, G., Borrego, A., & Abadal, E. (2025). OpenAlex y Crossref como fuentes de datos bibliográficas alternativas a Web of Science y Scopus em ciencias de la salud. *Revista Española de Documentación Científica*, 48(1), 1649. <https://doi.org/10.3989/redc.2025.1.1649>
- Céspedes, L., Kozłowski, D., Pradier, C., Sainte-Marie, M., Shokida, N. S., Benz, P., Poitras, C., Ninkvov, A. B., Ebrahimi, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2025). Evaluating the linguistic coverage of

- OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*, 76(6), 884-895. <https://doi.org/10.1002/asi.24979>
- Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., Van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected research: The potential of the PID Graph. *Patterns*, 2(2), Article e100180. <https://doi.org/10.1016/j.patter.2020.100180>
- Culbert, J. K., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, 130, 2475-2492. <https://doi.org/10.1007/s11192-025-05293-3>
- Danesh, F., & Mardani-Nejad, A. (2022). A historical overview of bibliometrics. In R. Ball (Ed.), *Handbook bibliometrics* (pp. 7-18). De Gruyter.
- Glänzel, W., & Shoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management*, 35(1), 31-44. [https://doi.org/10.1016/S0306-4573\(98\)00028-4](https://doi.org/10.1016/S0306-4573(98)00028-4)
- Houghton, F., & Foster, A. E. (2024). Resistance and collaboration within the panoptivarsity: Differing views of ORCID in Ireland. *Journal of Documentation*, 81(1), 195-210. <https://doi.org/10.1108/JD-07-2024-0184>
- Jiao, C., Li, K., & Fang, Z. (2023). How are exclusively data journals indexed in major scholarly databases? An examination of four databases. *Scientific Data*, 10, Article e737. <https://doi.org/10.1038/s41597-023-02625-x>
- Krüger, A. K. (2020). Quantification 2.0? Bibliometric infrastructures in academic evaluation. *Politics and Governance*, 8(2), 58-67. <https://www.cogitatiopress.com/politicsandgovernance/article/view/2575>
- Madi, N. R., Rodrigues, I., & Faria, L. I. L. (2026). O Openalex como alternativa na avaliação da ciência brasileira: Comparação com Web of Science e Scopus. *Ciência da Informação Express*, 7, Article e173. <https://doi.org/10.60144/v7i.2026.173>
- Maricato, J. M., Mazoni, A., Mugnaini, R., Packer, A. L., & Costas R. (2023). SciELO as an open scientometric research infrastructure: General discussion of coverage in OpenAlex, WoS, Scopus and Dimensions. In *27th International Conference on Science, Technology and Innovation Indicators* [Conference]. Leiden, Stadsgehoorzaal. <https://dapp.orvium.io/deposits/6442c231903ef57acd6dc640/view>
- Meadows, A. J. (1999). *A comunicação científica*. Briquet de Lemos.
- Mena-Chalco, J. (2025). Quantos CVs Lattes foram esquecidos pelos seus autores? *Linkedin*. <https://www.linkedin.com/pulse/quantos-cvs-lattes-foram-esquecidos-pelos-seus-jes%C3%BAs-p-mena-chalco-vts4f/?trackingId=9lp3YYr7UCiFysPOrpANuQ%3D%3D>
- Mena-Chalco, J., & Cesar R. M. C., Jr. (2009). scriptLattes: An open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, 15, 31-39. <https://doi.org/10.1007/BF03194511>
- Mezquita, B., Martín-Delgado, L., Wennberg-Capellades, L., & Borrego, A. A. (2025). A comparison of OpenAlex with Scopus and Web of Science for tracking scholarly nursing literature. *Sage Open Nursing*, 11, Article 23779608251361012. <https://doi.org/10.1177/2377960825136101>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106, 213-228. <https://link.springer.com/article/10.1007/s11192-015-1765-5>
- Mugnaini, R., Damasceno, R. J. P., Digiampietri, L. A., & Mena-Chalco, J. P. (2019). Panorama da produção científica do Brasil além da indexação: Uma análise exploratória da comunicação em periódicos. *TransInformação*, 31, Artigo e190033. <https://www.scielo.br/j/tinf/a/gWy9yV67t3WHRmWdG8DP7qS/?lang=pt>
- Mugnaini, R., Fraumann, G., Tuesta, E. F., & Packer, A. L. (2020). Openness trends in Brazilian citation data: Factors related to the use of DOIs. *Scientometrics*, 126, 2523-2556. <https://doi.org/10.1007/s11192-020-03663-7>
- Murillo, D., & Zapata, R. (2023). *Evaluación de datos de instituciones panameñas a través de la versión alpha de la plataforma OpenAlex*. Repositorio institucional de documentos digitales de acceso abierto de la

- Universidad Tecnológica de Panamá. <https://ridda2.utp.ac.pa/handle/123456789/18055?locale-attribute=es>
- Murillo-Gonzalez, D., & López, S. (2023). Persistent identifiers in the traceability of scientific research using OpenAlex. In *2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)* [Congreso]. Cali, Colombia. <https://ieeexplore.ieee.org/document/10366365>
- Ortega, J. L., & Delgado-Quirós, L. (2024). The indexation of retracted literature in seven principal scholarly databases: A coverage comparison of dimensions, OpenAlex, PubMed, Scilit, Scopus, The Lens and Web of Science. *Scientometrics*, 129(7), 3769-3785. <https://doi.org/10.1007/s11192-024-05034-y>
- Packer, A. L., & Memeghini, R. (2006). Visibilidade da produção científica. In D. A. Población, G. P. Witter, & J. F. M. Silva (Orgs.), *Comunicação e produção científica: Contexto, indicadores e avaliação* (pp. p. 237-259). Angellara.
- Pranckute, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1), 12. <https://www.mdpi.com/2304-6775/9/1/12>
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. In *26th International Conference on Science and Technology Indicators* [Conference]. Granada, Spain. <https://arxiv.org/abs/2205.01833>
- Rodrigues, W. B., Santos, T. V., & Mugnaini, R. (2025). O uso de identificadores persistentes na recuperação da produção científica brasileira: Análise de cobertura do Currículo Lattes na OpenAlex e ORCID [Conjunto de dados]. *Open Science Framework (OSF)*. <https://doi.org/10.17605/OSF.IO/7NBQ6>
- Santos, T. V. (2021). *Identificadores persistentes: Aplicabilidade na organização e acesso à informação científica*. [Dissertação de Mestrado, Universidade de São Paulo]. Biblioteca Digital de Teses e Dissertações da USP. <https://doi.org/10.11606/D.27.2021.tde-02052022-115537>
- Schnieders, K., Mierz, S., Boccalini, S., Meyer Zu Westerhausen, W., Hauschke, C., Hagemann-Wilholt, S., & Schulze, S. (2022). ORCID coverage in research institutions: Readiness for partially automated research reporting. *Frontiers in Research Metrics and Analytics*, 7, 1-13. <https://doi.org/10.3389/frma.2022.1010504>
- Sidone, O. J. G., Haddad, E. A., & Mena-Chalco, J. P. (2016). A ciência nas regiões brasileiras: Evolução da produção e das redes de colaboração científica. *Transinformação*, 28(1), 15-31. <https://doi.org/10.1590/2318-08892016002800002>
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126, 5113-5142. <https://doi.org/10.1007/s11192-021-03948-5>
- Turgel, I. D., & Chernova, O. A. (2024). Open science alternatives to Scopus and the Web of Science: A case study in regional resilience. *Publications*, 12(4), Article 43. <https://doi.org/10.3390/publications12040043>
- Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, 60(4), Article 103385. <https://doi.org/10.1016/j.ipm.2023.103385>

Dados de publicação

Wellington Barbosa Rodrigues

Doutor em Ciência da Computação

Universidade Federal do ABC, Santo André, SP, Brasil

Universidade de São Paulo, São Paulo, SP, Brasil

wellington.rodrigues@ufabc.edu.br

<https://orcid.org/0000-0002-6465-6012>

Doutor e mestre em Ciência da Computação e graduado em Sistemas de Informação. Pesquisador Colaborador na UFABC e Professor Colaborador do Departamento de Informação e Cultura da ECA/USP.

Thamyres Vieira dos Santos

Doutoranda em Ciência da Informação

Universidade de São Paulo, São Paulo, SP, Brasil
Hospital Israelita Albert Einstein, São Paulo, SP, Brasil
thamyres.vieira@hotmail.com
<https://orcid.org/0000-0003-3644-8600>

Doutoranda em Ciência da Informação pela Universidade de São Paulo (USP). Bacharel em Biblioteconomia e mestre em Ciência da Informação também pela USP. Atualmente é bibliotecária no Hospital Israelita Albert Einstein.

Rogério Mugnaini

Doutor em Ciência da Informação

Universidade de São Paulo, São Paulo, SP, Brasil
mugnaini@usp.br
<https://orcid.org/0000-0001-9334-3448>

Estatístico, mestre e doutor em Ciência da Informação. Professor Associado do Departamento de Informação e Cultura da Escola de Comunicações e Artes da Universidade de São Paulo, onde coordena o CiMetrias (Grupo de Pesquisa em Métricas da Ciência e Tecnologia).

Endereço para correspondência do autor principal

Avenida Professor Lúcio Martins Rodrigues, 443, 05508-020, São Paulo, SP, Brasil.

Originalidade

Declaro que o texto é original e não foi enviado para nenhuma outra publicação.

Preprint

O manuscrito não foi submetido a plataforma de preprint.

Informações sobre o trabalho

O trabalho resulta de um projeto de pesquisa coordenado por Rogério Mugnaini, intitulado “Relevância acadêmica de periódicos científicos no contexto nacional: estudo bibliométrico de países e bases de dados”, relacionado à Bolsa de Produtividade (CNPq) e desenvolvido pelo grupo CiMetrias, da Universidade de São Paulo.

Agradecimentos

Agradecemos ao CNPq, pelo financiamento da Bolsa de Produtividade (CNPq) de Rogério Mugnaini, processo 308800/2022-2; e CAPES - Código de Financiamento 001.

Contribuição dos autores

Concepção e preparação do manuscrito: W. B. Rodrigues, T. V. Santos, R. Mugnaini

Coleta dos dados: W. B. Rodrigues

Análise dos dados: W. B. Rodrigues, R. Mugnaini

Discussão dos resultados: W. B. Rodrigues, T. V. Santos, R. Mugnaini

Escrita - Rascunho original: W. B. Rodrigues, T. V. Santos, R. Mugnaini

Escrita - Revisão e aprovação: R. Mugnaini.

Uso de inteligência artificial

Não se aplica.

Financiamento

Bolsa de Produtividade (CNPq) de Rogério Mugnaini, processo 308800/2022-2; e CAPES - Código de Financiamento 001.

Permissão para usar imagens

Não se aplica.

Aprovação do Comitê de Ética em Pesquisa

Não se aplica.

Conflito de interesses

Os autores declaram que não há conflito de interesses.

Declaração de disponibilidade de dados

Os dados estão disponibilizados em quatro arquivos contendo resultados das análises em formato .xlsx e um arquivo principal em formato .csv, depositados em:

Rodrigues, W. B., Santos, T. V., & Mugnaini, R. (2025). O uso de identificadores persistentes na recuperação da produção científica brasileira: análise de cobertura do Currículo Lattes na OpenAlex e ORCID [Conjunto de dados]. *Open Science Framework (OSF)*. <https://doi.org/10.17605/OSF.IO/7NBQ6>

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

Karen Santos d'Amorim e Lúcia da Silveira.

Histórico

Recebido: 30-05-2025 – Aprovado: 14-04-2026 – Publicado em: 01-07-2026



Os artigos neste periódico estão licenciados sob uma Licença Creative Commons Atribuição 4.0 Estados Unidos.



This journal is published by [Pitt Open Library Publishing](https://pittopenlibrarypublishing.org/).