



ORCID COMO FONTE DE DADOS PARA ANÁLISE DA CARREIRA ACADÊMICA NO

BRASIL:

oportunidades e desafios

Ana Maria Carneiro

 <https://orcid.org/0000-0001-6688-1881>.

✉ anamacs@unicamp.br.

🏢 Universidade Estadual de Campinas (Unicamp) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

Daniela Maciel Pinto

 <https://orcid.org/0000-0003-0539-622X>.

✉ daniela.macielp@embrapa.br.

🏢 Empresa Brasileira de Pesquisa Agropecuária (Embrapa) |

ROR: <https://ror.org/0482b5b22> | Campinas, Brasil.

João Gabriel Pedreira de Moura Gomes

 <https://orcid.org/0009-0005-1887-9281>.

✉ j241749@dac.unicamp.br.

🏢 Universidade Estadual de Campinas (UNICAMP) |

ROR: <https://ror.org/04wffgt70> | Campinas, Brasil.

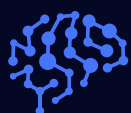
Eixo temático: Bases e Fontes de Dados

Modalidade: Resumo expandido

DOI: 10.22477/ix.ebbc.360

Resumo: A mobilidade na carreira acadêmica tem sido explorada usando uma diversidade de bases de dados secundários, tais como currículos, bases de publicações e, mais recentemente, a base ORCID (*Open Researcher and Contributor ID*). O objetivo deste trabalho é explorar o potencial da base de dados ORCID para analisar aspectos relacionados à carreira acadêmica no Brasil, incluindo a mobilidade entre instituições. Para isso, foram extraídos e organizados todos os registros de autores afiliados a instituições no Brasil. Buscou-se explorar a distribuição dos perfis da carreira docente em termos geográficos e por tipos de instituição, comparando-se com dados do Censo da Educação Superior.

Palavras-Chave: ORCID. Dados secundários. Carreira Acadêmica. Censo da Educação Superior. Brasil



1 INTRODUÇÃO

O que ocorre com um/a acadêmico/a quando muda de instituição em termos de progressão na carreira? E qual(is) a(s) melhor(es) fonte(s) de dados para explorar esta questão? A mobilidade na carreira acadêmica tem sido explorada com bases de dados secundários, tais como bases de publicações acadêmicas (Aman, 2018; Miranda-González *et al.*, 2020); e currículos, baseados em fontes de agências de fomento (Cañibano *et al.*, 2008) e bases nacionais, como os trabalhos baseados na Plataforma Lattes brasileira (Reis *et al.*, 2021).

Alguns estudos recentes têm avançado para o uso do ORCID (*Open Researcher and Contributor ID*) (Zhao *et al.*, 2022; Yan *et al.*, 2020; Gomez *et al.*, 2020; Zhang *et al.*, 2023), base consolidada em 2012, quando passou a ser promovida por editoras e organizações de pesquisa (Youtie *et al.*, 2017). A base permite analisar os currículos, carreiras e produções científicas (Costas *et al.*, 2022). Segundo estatísticas da base, há 8,1 milhões pesquisadores ativos, e o Brasil ocupa o quinto lugar entre os países com mais perfis (ORCID, 2024). O preenchimento da base é feito pelos pesquisadores de forma autodeclarada, mas há também mecanismos que podem alimentar determinados campos automaticamente. O formato da base permite incluir mais tipos de publicações que as bases tradicionais – como trabalhos em eventos, livros, preprints – e informações sobre formação e emprego (Costas *et al.*, 2022). No caso do emprego, traz o período de início e fim, o que permite o estudo longitudinal, fundamental para compreender a mobilidade acadêmica (Yan *et al.*, 2020). Comparada com outras bases, há uma maior presença de acadêmicos da África e América Latina (Costas *et al.*, 2022). Além disso, o foco no indivíduo permite análises empíricas “diretas”, sem a necessidade de proxies baseadas em produtos que muitas vezes trazem um retrato no qual as comunidades de pesquisa do Sul Global são subrepresentadas (Costas *et al.*, 2022).

A título de exemplo, Zhao *et al.* (2022) exploraram a persistência das colaborações em pesquisa dos cientistas que se movem entre instituições usando uma amostra de 4.343 cientistas norte-americanos e suas publicações na *Web of Science*. O uso da base ORCID permitiu detectar a mobilidade entre instituições, a direção da mobilidade e os países nos quais o cientista trabalhou. Na junção com a *Web of Science* permitiu analisar a produção bibliográfica em co-autoria, as áreas de pesquisa e a temporalidade da colaboração.

O objetivo deste trabalho é explorar a cobertura da base ORCID no Brasil em relação à comunidade acadêmica, focando no caso de professores em instituições de ensino superior (IES). A motivação surgiu no âmbito de um projeto maior que busca analisar os efeitos da mobilidade na carreira acadêmica no Brasil. Neste sentido, buscou-se explorar a distribuição dos perfis da carreira docente em termos geográficos e por tipos de instituição, comparando-se com dados do Censo da Educação Superior 2020. Trata-se de um trabalho pioneiro no Brasil, que logrou obter como resultado o mapeamento da cobertura da base ORCID no Brasil por federação, bem como a diversidade de perfis presentes na base no que diz respeito à baixa mobilidade de emprego.

2 METODOLOGIA

A análise foi realizada a partir da extração, em setembro de 2023, de todos os registros da base ORCID referentes a indivíduos filiados a instituições no Brasil. Foram extraídas as informações de três tabelas (educação, qualificação e emprego). Neste estudo priorizou-se as tabelas emprego e qualificação devido ao foco nas ocupações profissionais (roles). Iniciou-se com a tabela emprego, que continha 449.124 registros (linhas) e 17 variáveis, sendo uma correspondente à chave, ou número do ORCID. Ressalta-se que os dados das demais 16 variáveis não possuíam qualquer padronização, tendo sido encontrado, por exemplo, 4.811 regiões brasileiras. Foram excluídos os registros duplicados (172.880)

Procedeu-se uma análise para tentar padronizar as ocupações e identificar profissionais de ciência e tecnologia, usando dois dicionários em português e inglês, preparados pela equipe, que possuíam palavras relacionadas à docência, pesquisa e ciência. Adicionalmente, recorreu-se a léxicos especializados do Instituto Brasileiro de Geografia e Estatística (IBGE) para a normalização dos dados geográficos (municípios, regiões e estados) do Brasil. Essas operações resultaram na criação de novas variáveis e eliminação de registros não referentes à profissionais de Ciência e Tecnologia (C&T) e ensino superior, restando 76.705 registros, agora distribuídos em 27 variáveis. Com a normalização, surgiram novas duplicatas que foram removidas, resultando em 73.861 registros.

Os próximos passos focam em refinar os dados para encontrar apenas professores em IES. A identificação das IES foi feita de forma manual, resultando em 59.474 registros. As ocupações profissionais foram examinadas e foram selecionadas apenas aqueles referentes à professores, resultando em 40.667 registros. A partir desse conjunto, procedeu-se à identificação dos docentes em exercício no ano de 2020 para fins de comparação com dados do Censo da Educação Superior de 2020 e da tipologia de instituições de ensino superior.

Para a identificação das tipologias das IES utilizou-se a classificação proposta por Schwartzman (2021), cujo banco de dados tivemos acesso no contexto do projeto maior. Este banco utilizou os microdados do Censo da Educação Superior 2020, especificamente da tabela instituições. Neste sentido, os dados disponíveis na base da tipologia referem-se a vínculos (com repetição devido a contratos múltiplos dos docentes com mais de uma IES), enquanto na base ORCID preparada para esta pesquisa os dados referem-se a professores (registros únicos), privilegiando o vínculo mais antigo. Neste sentido, foi necessário ajustar tais dados aos parâmetros estabelecidos pelo Censo da Educação Superior 2020 para docentes em exercício, a partir da aplicação de um fator de correção, formulado conforme segue:

$$\text{Fator de correção} = \frac{\text{Número de Docentes Segundo Censo}}{\text{Número de Docentes Segundo Base Tipologia}}$$

Em seguida, o fator de correção foi aplicado aos dados originais e também aos dados do ORCID, conforme formulação:

$$\text{Base_Original_fator de correção} = \text{Base_Tipologia} \times \text{Fator de Correção}$$

Este ajuste buscou calcular a proporção entre o número total de docentes identificados no ORCID



em relação aos dados oficiais do Censo.

A análise dos dados foi realizada utilizando a linguagem de programação R (versão 4.3.1; 2023-06-16 ucrt) com o ambiente de desenvolvimento Rstudio (versão 2023.6.1.524) e os pacotes "sf", "geobr", "widyr", "igraph", "ggraph", "tidyr", "SnowballC", "stopwords", "tm", "tidytext", "lubridate", "ggthemes", "treemapify", "openxlsx", "rgdal", "sp", "maps", "ggmap", "wordcloud", que possibilitaram a análise espacial, o processamento de texto e a visualização de dados. Complementarmente, os softwares Tableau (versão 2023.2.4.0115.0353), LibreOffice Calc (versão 7.5.5.2) e OpenOffice Calc (versão 4.1.14).

3 RESULTADOS

A análise da base de dados identificou 25.591 registros de docentes únicos, em exercício no ano de 2020 ou posterior, vinculados à IES. Para contextualizar, segundo dados do Censo da Educação Superior de 2020, o Brasil possuía, em 2020, 314.384 docentes no ensino superior em exercício. Assim, os registros na base ORCID correspondem a aproximadamente 8,1% do contingente total de docentes do país.

Complementarmente, para esses indivíduos, buscou-se identificar qualificações distintas registradas na base a partir do cruzamento da tabela "emprego" com a tabela "qualificação", que possuía um total de 190.594 registros e 17 variáveis. Esse cruzamento teve como principal objetivo identificar quantas ocupações distintas um mesmo indivíduo poderia ter, podendo corresponder a vínculos sequenciais ou simultâneos. A grande maioria possui apenas uma ocupação ou um vínculo com uma instituição (Tabela 1), o que indica um percurso acadêmico sem a presença de mobilidade. Uma relação semelhante é percebida quando se busca identificar o número de instituições às quais os indivíduos estiveram vinculados. As IES com mais registros foram Universidade de São Paulo (USP) (1.641), Universidade Federal do Rio de Janeiro (UFRJ) (612) e a Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) (565).

A região sudeste se destaca, com São Paulo (8.712), Rio de Janeiro (3.295) e Minas Gerais (2.973) (Figura 1-A). Comparando-se com os dados do Censo da Educação Superior, a distribuição mostra-se semelhante entre as unidades federativas (Figura 1-B).

Tabela 1 - Quantidade de perfis por indivíduos segundo ocupações e instituições

Nº repetições	Indivíduos por ocupações (roles)	Indivíduos por Instituições
1	25306	24835
2	615	1010
3	35	102
4	4	12
5	-	2
6	-	-
7	-	-

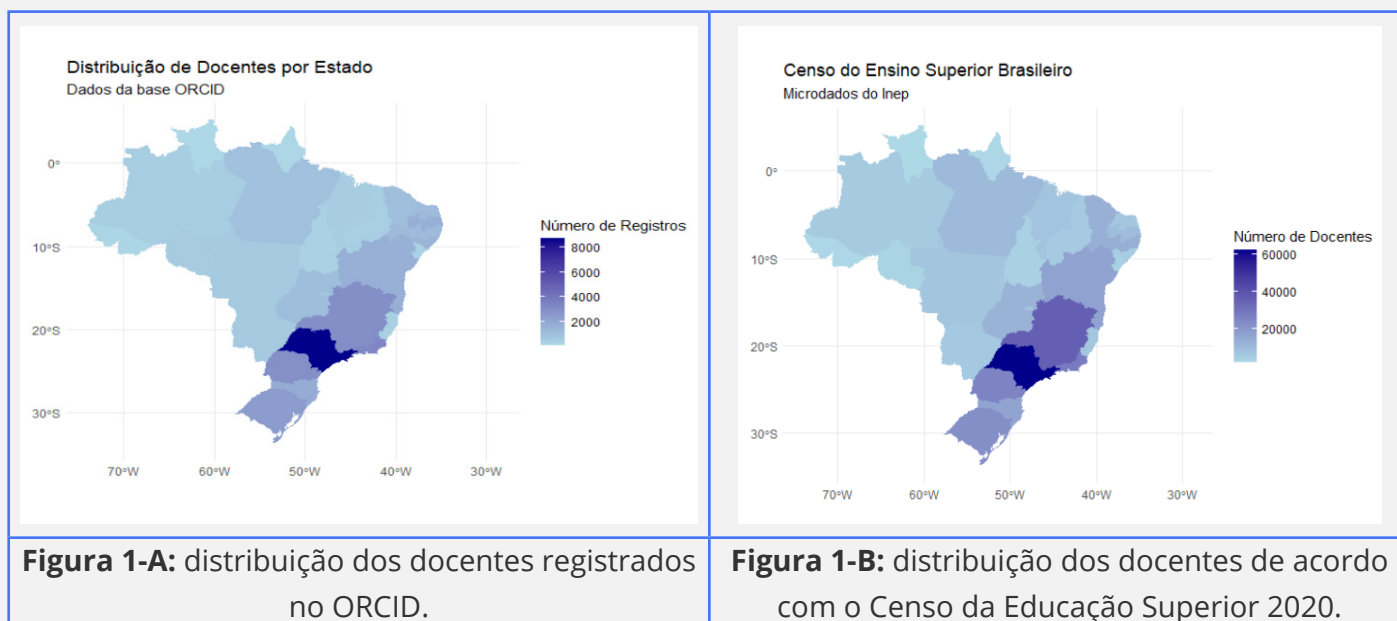
Nº repetições	Indivíduos por ocupações (roles)	Indivíduos por Instituições
8	1	-

Fonte: Elaborado pelos autores (2024).

Ao mapear os diferentes títulos acadêmicos adotados, identificou-se 40 tipos distintos. O título de “docente” é o mais comum, com 8.854 registros, seguido por “associate professor” (3.619 registros) e “professor titular” (2.337 registros). Outros títulos frequentes incluem “assistant professor”, “teacher”, “tutor”, “professor doutor”, “adjunct professor” e “lecturer”. Essa diversidade de títulos indica não apenas diferentes níveis de experiência e responsabilidades acadêmicas, mas também variações nas estruturas de cargos de diferentes instituições de ensino e diferentes papéis dentro do ambiente acadêmico, desde posições mais focadas em pesquisa até aquelas com ênfase no ensino.

A utilização da tipologia de IES de Schwartzman (2021) nos dados do ORCID revelou 551 IES únicas (Tabela 2).

Figura 1 - Mapa da distribuição dos docentes por Unidade da Federação, ORCID e Censo da Educação Superior



Fonte: Elaborado pelos autores (2024).

Tabela 2 - Distribuição de docentes ativos e número de IES, segundo dados do Censo da Educação Superior 2020 e da base ORCID

Tipologia	% Docentes Censo	% Docentes ORCID	% IES Censo	% IES ORCID
1 - Privadas de grande porte	5,7	4	1	2,7
2 - Públicas de grande porte com pós-graduação alta	10,7	32	0,4	31,0

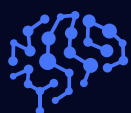


Tipologia	% Docentes Censo	% Docentes ORCID	% IES Censo	% IES ORCID
3 - Públicas especializadas com pós-graduação alta	9,2	18	1	7,6
4 - Privadas comunitárias e diferenciadas	8,6	8	3	10,9
5 - Universidades e Centros Universitários privados	18,5	8	10	14,2
6 - Outras instituições públicas	23,5	27	5	22,9
7 - Faculdades privadas isoladas	20,4	1	53	5,8
8 - Instituições de educação tecnológica	3,3	1	9	1,8
9 - Instituições de Pesquisa e Pós-Graduação	0,1	1	3	2,2
Sem informação	0,0	0	14	0,9
TOTAL	100 N = 314.383	100 N = 25.961	100 N = 2.945	100 N = 551

Fonte: Elaborado pelos autores (2024).

4 CONSIDERAÇÕES FINAIS

A base ORCID coloca-se como uma grande promessa para estudos cientométricos, mas há poucos trabalhos que exploram este potencial (Costas *et al.*, 2022). O quadro tem mudado rapidamente e se nota cobertura crescente no Sul Global e no Brasil (Youtie *et al.*, 2017; Costas *et al.*, 2022). O uso da base deve ser feito tendo em vista suas limitações. As principais apontadas pela literatura são a maior cobertura de cientistas que se formaram mais recentemente e são mais produtivos (Gomez *et al.*, 2020); a não padronização dos dados (Costas *et al.*, 2022; Gomez *et al.*, 2020); o nível de preenchimento variado por campo (Costas *et al.*, 2022); e o fato de que a informação não é verificada, o que pode resultar em dados inacurados ou incorretos (Costas *et al.*, 2022). Além disso, há preocupações éticas com o status de uso compulsório por parte de periódicos e organizações de pesquisa (Costas *et al.*, 2022). Conforme nossos resultados, no caso brasileiro, há uma boa representação em termos geográficos na comparação com os dados do Censo da Educação Superior. Entretanto, há uma sobrerepresentação na base ORCID de vínculos com IES mais intensivas em pesquisa (tipos 2, 9 e 4), uma subrepresentação das instituições privadas (tipos 1, 5, 7), das instituições vocacionais (tipo 8) e das IES públicas mais focadas em ensino (tipo 6). Desta forma, a base parece refletir mais os professores que atuam mais intensamente na pesquisa e pós-graduação, além de estarem vinculados a IES que provavelmente têm tornado



compulsório o registro no ORCID. Pode indicar, ainda, a maior cobertura de profissionais com maior circulação internacional, o que demandaria a inclusão do CV (Curriculum Vitae) no ORCID, que dá uma maior visibilidade neste contexto.

Tendo em vista o potencial de crescimento da cobertura, abre-se uma agenda de pesquisa para estudos sobre a carreira acadêmica no Brasil, incluindo a análise de mobilidade entre instituições, a progressão na carreira e diferenças entre áreas, com a complementação de dados de publicações. E também a comparação com outras bases nacionais, como a Plataforma Lattes. Esta é reconhecida com uma base de dados de cobertura praticamente censitária em relação à academia, diferente de outros países. A base ORCID, apesar da menor cobertura, apresenta uma oportunidade de realizar comparações internacionais, bem como acompanhar os cientistas que saem do Brasil e deixam de atualizar o Lattes.

AGRADECIMENTOS

Este trabalho conta com financiamento da Fapesp (Pesquisa da pesquisa e da inovação: indicadores, métodos e evidências de impactos - Processo n.: 21/15091-8). Os autores agradecem a equipe do projeto maior que apoiaram a realização da pesquisa.

REFERÊNCIAS

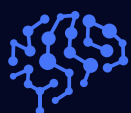
AMAN, V. A new bibliometric approach to measure knowledge transfer of internationally mobile scientists. **Scientometrics**, [S. l.], v. 117, n. 1, p. 227-247, 2018. DOI: <https://doi.org/10.1007/s11192-018-2864-x>. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2864-x>. Acesso em: 25 mar. 2024.

CAÑIBANO, C. *et al.* Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal programme in Spain. **Research Evaluation**, [S. l.], v. 17, n. 1, p. 17-31, 2008. DOI: <https://doi.org/10.3152/095820208X292797>. Disponível em: <https://academic.oup.com/rev/article-abstract/17/1/17/1529038?redirectedFrom=fulltext>. Acesso em: 25 mar.2024.

COSTAS, R. *et al.* Could ORCID play a key role in meta-research? discussing new analytical possibilities to study the dynamics of science and scientists. **SocArXiv**, [S. l.], preprint, 2022. Disponível em: <https://osf.io/preprints/socarxiv/sjck6>. Acesso em: 25 mar.2024.

GOMEZ, C. J. *et al.* Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID. J. **Informetrics**, [S. l.], v. 14, n. 3, 2020. DOI: <https://doi.org/10.1016/j.joi.2020.101044>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157719304389?via%3Dihub>. Acesso em: 25 mar. 2024.

MIRANDA-GONZÁLEZ, A. *et al.* Scholarly migration within Mexico: analyzing internal migration among researchers using Scopus longitudinal bibliometric data. **EPJ Data Science**, [S. l.], v. 9, n. 1, p. 34, 2020. DOI: <https://doi.org/10.1140/epjds/s13688-020-00252-9>. Disponível em em: <https://epjdatascience>.



springeropen.com/articles/10.1140/epjds/s13688-020-00252-9. Acesso em: 25 mar. 2024.

REIS, J. E. D. *et al.* Impact of teacher education abroad in international co-authorship: a study of Federal University of São Carlos's scientific production indexed in the Web of Science. **Transinformação**, [S. l.], v. 33, 2021. DOI: <https://doi.org/10.1590/2318-0889202133e200061>. Disponível em: <https://www.scielo.br/j/tinf/a/ZDdpxb3FyYZJVCvf8n8kSSh/?lang=en>. Acesso em: mar. 2024.

SCHWARTZMAN, S. *et al.* Por uma tipologia do ensino superior brasileiro: teste de conceito. **Estudos Avançados**, [S. l.], v.35, n.101, p.153-86, jan. 2021. DOI: <https://doi.org/10.1590/s0103-4014.2021.35101.011>. Disponível em: <https://www.scielo.br/j/ea/a/D7bZJtnnXyjCPQ3g9T8CzpK/?lang=pt>. Acesso em 25 mar. 2024.

YAN, E. *et al.* Analyzing academic mobility of US professors based on ORCID data and the Carnegie Classification. **Quantitative Science Studies**, [S. l.], v. 1, n. 4, p. 1451-1467, 2020. DOI: https://doi.org/10.1162/qss_a_00088. Disponível em: <https://direct.mit.edu/qss/article/1/4/1451/96122/Analyzing-academic-mobility-of-U-S-professors>. Acesso em: 24 mar. 2024.

YOUTIE, J. *et al.* Tracking researchers and their outputs: new insights from ORCIDs. **Scientometrics**, [S. l.], v. 113, p. 437-453, 2017. DOI: <https://doi.org/10.1007/s11192-017-2473-0>. Disponível em: <https://link.springer.com/article/10.1007/s11192-017-2473-0>. Acesso em: 24 mar. 2024.

ZHANG, L. *et al.* Continued collaboration shortens the transition period of scientists who move to another institution. **Scientometrics**, [S. l.], v. 128, n. 3, p. 1765-1784, 2023. DOI: <https://doi.org/10.1007/s11192-022-04617-x>. Disponível em: <https://link.springer.com/article/10.1007/s11192-022-04617-x>. Acesso em: 25 mar. 2024.

ZHAO, Z. *et al.* Does the mobility of scientists disrupt their collaboration stability? **Journal of Information Science**, [S. l.], v. 48, n. 2, p. 199-209, 2022. DOI: <https://doi.org/10.1177/0165551520948744>. Disponível em: <https://journals.sagepub.com/doi/10.1177/0165551520948744>. Acesso em: 24 mar. 2024.