26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
ISSN 2177-3688

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

PRESENTE, MAS AUSENTE: UM ÍNDICE MULTIDIMENSIONAL DE COMPLETUDE ORCID

PRESENT BUT ABSENT: *A MULTIDIMENSIONAL INDEX OF ORCID RECORD COMPLETENESS*

Robson Lopes de Almeida – Rede Nacional de Ensino e Pesquisa (RNP)

Modalidade: Trabalho Completo

Resumo: Este trabalho propõe o Índice Geral de Completude ORCID (IGCO), medida multidimensional de completude funcional publicamente observável, e o aplica aos registros ORCID recuperados para um universo de 684 coordenadores de Programas de Pós-Graduação com notas 6 e 7 no resultado parcial da Avaliação Quadrienal CAPES 2021–2024. Como os dados da CAPES não continham os ORCID iDs, os registros foram localizados por meio da API pública do ORCID, mediante estratégia automatizada de busca em cascata. O índice reúne três componentes: Metadados Essenciais, Cobertura de Identificadores Externos e Vínculo Institucional. Foram recuperados 614 registros (89,8%), enquanto 70 coordenadores (10,2%) foram classificados em situação de silêncio informacional. Entre os registros recuperados, o IGCO apresentou média de 0,902. Na decomposição do índice, as médias de seus três componentes foram de 0,984 para Metadados Essenciais, 0,917 para Cobertura de Identificadores Externos e 0,806 para Vínculo Institucional. Este último concentrou a menor completude e a maior dispersão. O contraste exploratório entre especificações sugeriu que as diferenças entre Grandes Áreas dependem da construção da medida. Conclui-se que o IGCO oferece uma linha de base para monitorar a evolução da completude dos registros ORCID e examinar, em estudos futuros, sua relação com iniciativas de integração institucional.
Palavras-chave: ORCID; completude da informação; qualidade de metadados; identificadores persistentes; informação científica.

Abstract: This study proposes the General ORCID Completeness Index (IGCO), a multidimensional measure of publicly observable functional completeness, and applies it to ORCID records retrieved for a population of 684 coordinators of Brazilian graduate programs rated 6 or 7 in the partial results of the CAPES 2021–2024 Quadrennial Evaluation. Because the CAPES

data did not include ORCID iDs, the records were located through the ORCID Public API using an automated cascading search strategy. The index comprises three components: Essential Metadata, External Identifier Coverage, and Institutional Affiliation. A total of 614 records (89.8%) were retrieved, while the remaining 70 coordinators (10.2%) could not be matched under the procedure and were classified as cases of informational silence. Among the retrieved records, the mean IGCO was 0.902. When the index was decomposed into its three components, mean scores were 0.984 for Essential Metadata, 0.917 for External Identifier Coverage, and 0.806 for Institutional Affiliation. The latter showed the lowest completeness and the greatest dispersion. The exploratory comparison suggested that differences across broad fields depend on how the measure is constructed. The IGCO provides a baseline for monitoring changes in ORCID record completeness and examining, in future studies, its relationship with institutional integration initiatives.
Keywords: ORCID; information completeness; metadata quality; persistent identifiers; scientific information.

1 INTRODUÇÃO

A crescente integração de bases e sistemas de informação científica amplia a importância da qualidade dos registros utilizados para identificar pesquisadores, relacionar sua produção e apoiar processos de gestão e avaliação. Em ambientes nos quais essas operações passam a depender de coleta e intercâmbio automatizados, registros incompletos limitam a capacidade de reconhecer entidades, estabelecer relações entre fontes e reutilizar informações com segurança.

No Brasil, essa questão se coloca em um ambiente marcado pela coexistência de sistemas nacionais de informação científica e iniciativas de integração por identificadores persistentes. Nesse contexto, o *Open Researcher and Contributor ID* (ORCID) constitui uma infraestrutura aberta de identificação persistente que permite desambiguar pesquisadores e conectá-los a publicações, afiliações, financiamentos e outras atividades acadêmicas (Haak, 2013).

Ter um ORCID iD, contudo, não assegura que o registro reúna informações suficientes para estabelecer essas conexões. Estudos têm identificado lacunas de preenchimento em informações sobre formação acadêmica, vínculo institucional e outras seções dos registros (Arroyo-Machado *et al*., 2026; Fernández-Marcial; González-Solar; Vale, 2023; Wang; Li; Huang; Yang, 2024).

O problema, portanto, não se restringe à adoção do identificador. Porter, Umbach e Willis (2025) afirmam que exigências de periódicos constituem um dos principais motivos para a obtenção do ORCID iD. Esse resultado reforça a necessidade de distinguir a adoção do

identificador da completude de seu registro. O pesquisador pode, assim, estar formalmente identificado sem que o registro reúna as informações necessárias aos usos pretendidos.

Entre as iniciativas brasileiras de integração, o Consórcio ORCID Brasil tem a Rede Nacional de Ensino e Pesquisa (RNP) como organização responsável por sua operação nacional (Heredia, 2018; RNP, 2026). Nesse contexto, medidas de completude podem fornecer uma linha de base para investigar, posteriormente, sua relação com processos de integração institucional.

Diante desse cenário, este estudo propõe o Índice Geral de Completude ORCID (IGCO), concebido como medida de completude funcional publicamente observável. O índice reúne três componentes relacionados aos usos examinados: Metadados Essenciais (ME), Cobertura de Identificadores Externos (CIE) e Vínculo Institucional (VI). Sua aplicação empírica concentra-se nos coordenadores de Programas de Pós-Graduação com notas 6 e 7 no resultado parcial da Avaliação Quadrienal CAPES 2021-2024, divulgado em fevereiro de 2026.

O trabalho investiga em que medida os registros ORCID desses coordenadores apresentam os elementos necessários à descrição da produção, à conexão com outras fontes e à identificação do vínculo institucional. Seu objetivo é propor e aplicar uma medida de completude observável e identificar as principais lacunas no universo analisado. O estudo também examina como a escolha dos componentes do índice afeta as diferenças observadas entre as áreas, aspecto relevante porque indicadores de completude podem incorporar características das trajetórias acadêmicas e dos padrões disciplinares.

2 COMPLETUDE COMO DIMENSÃO DE QUALIDADE DA INFORMAÇÃO

A qualidade dos dados é um conceito multidimensional, cuja avaliação depende do contexto em que os dados são utilizados. Wang e Strong (2015) definem dados de qualidade como aqueles adequados ao uso por seus usuários e organizam suas dimensões em quatro categorias: intrínseca, contextual, representacional e de acessibilidade. Nesse modelo, a completude integra a categoria contextual, ao lado da relevância, da atualidade e da quantidade apropriada de dados. A ISO/IEC 25012 adota entendimento convergente ao definir a qualidade dos dados como o grau em que suas características atendem a necessidades estabelecidas ou implícitas em determinadas condições de uso (International Organization for Standardization, 2008).

Entre as dimensões da qualidade da informação, a completude refere-se à presença dos elementos necessários para representar adequadamente uma entidade ou atender a

determinado uso. No campo dos metadados, Bruce e Hillmann (2004) a situam ao lado de critérios como acurácia, consistência, proveniência e atualidade. Király e Büchler (2018) mostram que ela pode ser avaliada por medidas agregadas aplicadas a conjuntos de registros, enquanto Pipino, Lee e Wang (2002) apresentam a razão entre elementos presentes e esperados como uma de suas formas de mensuração. A definição do que constitui um registro completo depende, portanto, dos elementos requeridos para o uso considerado. A presença desses elementos, contudo, não assegura que as informações sejam corretas ou estejam atualizadas.

Para efeito deste trabalho, adota-se a noção de completude funcional publicamente observável. O termo "funcional" indica que a medida está vinculada aos usos definidos no estudo, enquanto "publicamente observável" restringe a análise às informações disponíveis no registro público e acessíveis por meio da API do ORCID. Um registro é considerado mais completo quando apresenta os elementos necessários aos usos analisados: descrição mínima da produção, conexão das obras com outras fontes por meio de identificadores externos e informação sobre vínculo institucional.

Essa delimitação não pressupõe que o ORCID deva reproduzir um currículo acadêmico nem que todos os campos disponíveis precisem estar preenchidos. O IGCO avalia apenas o que está presente no registro público, ou seja, não estima quanto da trajetória acadêmica do pesquisador está representado no ORCID em comparação com uma fonte externa nem verifica a acurácia ou a atualidade das informações registradas.

2.1 ADOÇÃO, COBERTURA E LIMITAÇÕES DO ORCID

Os estudos sobre o ORCID abordam três aspectos relacionados, mas distintos: a adoção do identificador, sua integração a sistemas e fluxos de informação e a completude dos registros. Para fins analíticos, a adoção corresponde à criação ou à existência do ORCID iD; a integração, à incorporação do identificador a metadados, sistemas e processos de informação; e a completude, à disponibilidade das informações necessárias para determinado uso.

O termo cobertura também aparece na literatura com sentidos diferentes, podendo indicar tanto a presença do identificador em bases bibliográficas quanto a proporção da produção representada no registro. A distinção entre esses aspectos evita que a simples existência do identificador seja tomada como evidência de um registro completo ou efetivamente utilizado.

Ao examinarem registros por meio da API pública do ORCID, Eichenlaub e Morgan (2017) identificaram grande volume de perfis com informação mínima e destacaram que a presença

do identificador, sem metadados suficientes, limita funções como desambiguação e interoperabilidade.

Porter (2022), ao comparar publicações associadas aos registros ORCID com a produção atribuída aos mesmos pesquisadores na base Dimensions, aproximou a completude pela cobertura da produção e a utilizou como *proxy* de engajamento. Essa abordagem evidencia que a presença do identificador não equivale ao uso efetivo da infraestrutura, embora operacionalize a completude a partir de uma dimensão distinta da adotada neste trabalho.

As limitações dos registros ORCID como fonte para usos métricos também são documentadas na literatura. Wang, Li, Huang e Yang (2024) identificaram informações ausentes, inexatas ou desatualizadas em campos de formação e experiência profissional. Fernández-Marcial, González-Solar e Vale (2023), mesmo em um contexto de elevada adoção, encontraram registros incompletos e inconsistências na informação de afiliação. Mais recentemente, Arroyo-Machado *et al*. (2026) identificaram variações na presença de metadados entre seções dos registros e campos científicos.

Parte dessas limitações está relacionada às condições de adoção e manutenção dos registros. Porter, Umbach e Willis (2025) mostram que a obtenção do ORCID iD pode ser impulsionada por exigências de periódicos, agências de fomento ou sistemas institucionais. Essa forma de adoção, contudo, não assegura, por si só, a atualização posterior do registro. A criação do identificador, portanto, não deve ser confundida com o preenchimento ou a manutenção continuada do registro.

A adoção e a disponibilidade de informações também variam entre campos científicos e contextos institucionais. Porter, Umbach e Willis (2025) encontraram ampla variação disciplinar na adoção do identificador, enquanto Arroyo-Machado *et al.* (2026) identificaram diferenças na presença de metadados entre campos científicos. No Brasil, Carneiro, Pinto e Gomes (2025) também observaram assimetrias na cobertura da base ORCID entre docentes. Em conjunto, essas evidências reforçam a necessidade de distinguir a existência do identificador da completude informacional do registro.

2.2 GOVERNANÇA, PROVENIÊNCIA E INTEGRAÇÃO INSTITUCIONAL

A qualificação dos registros ORCID não depende apenas de ações individuais. Mediante autorização do pesquisador, organizações membros podem utilizar a API de membros do ORCID para registrar ou atualizar informações, associando às afirmações uma origem identificável

(ORCID, [20--a], [20--b]). Esse mecanismo amplia as possibilidades de integração entre sistemas institucionais e registros ORCID e acrescenta uma dimensão de proveniência que não está presente quando se observa apenas se determinado campo está preenchido.

Já em 2017, Eichenlaub e Morgan apontavam educação, atuação do próprio ORCID e maior interoperabilidade entre sistemas como caminhos para ampliar a robustez dos metadados. Mais recentemente, Alves, Silva, Grando e Jorge (2023) registraram o caso da Fiocruz, em que o ORCID foi utilizado para identificação, desambiguação e deduplicação de autores na construção de uma base de produção acadêmica institucional. Em conjunto, esses trabalhos indicam que políticas, sistemas e rotinas de curadoria podem criar condições favoráveis à manutenção e ao reaproveitamento dos registros ORCID (Alves; Silva; Grando; Jorge, 2023; Eichenlaub; Morgan, 2017).

Essa perspectiva é reforçada por Bhatt e Paul (2025), para quem os avanços do ORCID são mais expressivos em ambientes nos quais há políticas institucionais, infraestrutura técnica e conectividade via API. Nesse debate, a noção de *trust markers* refere-se a sinais de proveniência que permitem distinguir informações autodeclaradas daquelas afirmadas por organizações, acrescentando ao registro uma origem identificável e uma camada de responsabilidade institucional. A partir dessa lógica, relatório publicado pela SCImago Lab interpreta afiliações ORCID afirmadas por instituições como sinais de maturidade da infraestrutura digital e da governança da pesquisa, associadas a melhores indicadores de visibilidade e desempenho institucional, sem estabelecer relação causal direta entre adoção do ORCID e excelência em pesquisa (SCImago Research Group, 2026).

Embora relevante para a qualidade informacional, a proveniência das afirmações não integra o IGCO proposto neste estudo, que se restringe à completude funcional publicamente observável, isto é, à presença dos elementos definidos como necessários aos usos examinados. O trabalho tampouco avalia os mecanismos de integração eventualmente adotados pelas instituições dos coordenadores analisados. A ausência de determinados elementos identifica uma lacuna, mas não permite atribuí-la à inexistência de integração institucional. Essa relação permanece como hipótese para estudos posteriores, inclusive em análises que considerem iniciativas associadas ao Consórcio ORCID Brasil.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa caracteriza-se como exploratória e descritiva, de abordagem quantitativa. O universo empírico foi definido a partir da relação de coordenadores dos Programas de Pós-Graduação que receberam notas 6 ou 7 no resultado parcial da Avaliação Quadrienal CAPES 2021–2024, divulgado em fevereiro de 2026 (CAPES, 2026). A relação consultada informava o nome do coordenador, o programa, a instituição e a área de avaliação.

A exportação inicial resultou em 685 ocorrências. Após a consolidação de uma duplicidade nominal referente ao mesmo coordenador associado a duas instituições no mesmo programa, foram identificados 684 coordenadores únicos. Na etapa de recuperação, a unidade de referência é o coordenador; para o cálculo do IGCO, a unidade de análise corresponde ao registro ORCID individual nos casos em que foi possível atribuir uma correspondência. O recorte não pretende representar a comunidade científica brasileira, mas analisar a completude dos registros em um grupo diretamente inserido nos processos formais de avaliação da pós-graduação.

3.1 RECUPERAÇÃO DOS REGISTROS ORCID

Os dados disponibilizados pela CAPES não continham os ORCID iDs dos coordenadores analisados. Por essa razão, foi necessária uma etapa específica para localizar e atribuir os registros ORCID correspondentes a esses pesquisadores.

Os procedimentos de coleta foram realizados entre 15 e 18 de fevereiro de 2026, em ambiente Google Colaboratory, com Python 3 e acesso à API pública do ORCID, versão 3.0. A busca adotou uma estratégia em cascata, aplicada sucessivamente até que um registro atendesse aos critérios de correspondência definidos. Na primeira etapa, foram utilizados os campos estruturados de prenome e sobrenome. Quando não houve correspondência satisfatória, realizou-se uma busca ampla combinando os termos do nome. Na terceira etapa, a consulta foi simplificada para primeiro e último nomes, com o objetivo de recuperar registros em que nomes intermediários estivessem ausentes ou apresentados de forma diferente.

Para comparar os nomes constantes nos dados da CAPES com aqueles recuperados no ORCID, os nomes foram normalizados, com remoção de acentos e conversão para caixa baixa. A similaridade nominal foi calculada em Python por duas medidas complementares: uma baseada na correspondência entre sequências de caracteres (SequenceMatcher) e outra na

l\sobreposição dos termos nominais com mais de dois caracteres. Para cada par de nomes, adotou-se como escore o maior valor obtido entre as duas medidas.

Na ausência de um conjunto de referência previamente validado que permitisse calibrar empiricamente um ponto de corte, adotou-se 0,85 como limiar heurístico mínimo para a avaliação dos candidatos. O valor funcionou como condição de elegibilidade, e não como critério autônomo de atribuição. Nos 614 casos em que foi possível estabelecer correspondência entre o coordenador e um registro ORCID, o menor escore de similaridade observado foi 0,91.

A atribuição combinou similaridade nominal e informação de afiliação. Candidatos abaixo do limiar mínimo não foram aceitos. Entre os candidatos elegíveis, a coincidência com a instituição informada pela CAPES reforçava a correspondência. Quando a afiliação não permitia essa confirmação, aplicaram-se critérios adicionais: para nomes com três ou mais termos, o candidato somente podia ser aceito quando apresentava similaridade nominal igual ou superior a 0,95 e ao menos uma obra pública no registro; nos demais casos previstos pela rotina, a atribuição era admitida apenas quando a consulta retornava no máximo dois candidatos e a similaridade permanecia acima do limiar mínimo. Em cada consulta, foram examinados os candidatos na ordem retornada pela API, até o limite de 10 resultados.

Quando nenhum registro atendia aos critérios definidos, o caso foi classificado como “silêncio informacional”. A classificação indica apenas que o procedimento não permitiu atribuir um ORCID iD segundo os critérios estabelecidos e não deve ser interpretada como evidência de que o pesquisador não possua ORCID. A inexistência de um conjunto de referência previamente validado impediu a estimativa formal da precisão e da revocação da estratégia de recuperação; a rotina também não foi submetida a validação manual independente.

3.2 CONSTRUÇÃO DO ÍNDICE GERAL DE COMPLETUDE ORCID

O Índice Geral de Completude ORCID (IGCO) mede a completude funcional publicamente observável por meio de três componentes (Quadro 1), cada um associado a um uso do registro: descrever a produção, conectar as obras a outras fontes e informar o vínculo institucional.

O componente Metadados Essenciais (ME) corresponde à proporção de obras públicas que apresentam simultaneamente título, tipo e ano. Cobertura de Identificadores Externos (CIE) é a proporção de obras com ao menos um identificador externo registrado no ORCID. A expressão “identificadores externos” foi adotada porque a coleta registra a presença de *external-id*

nas obras, sem classificar todos os tipos quanto à persistência; o componente indica potencial de conexão e reconciliação, não densidade estrita de identificadores persistentes. Vínculo Institucional (VI) recebe valor 1 quando há ao menos um registro em *Employment* sem data de término, 0,5 quando há somente vínculos encerrados e 0 quando essa seção não contém informação pública. A ausência de data de término funciona como proxy de vínculo atual, sem comprovar que a afiliação esteja de fato vigente.

Quadro 1 - Componentes e dimensões funcionais do IGCO[1]

Dimensão funcional	Indicador	Operacionalização
Descrição da produção	ME	Obras com título + tipo + ano / total de obras públicas consideradas
Conexão com outras fontes	CIE	Obras com ao menos um identificador externo / total de obras públicas consideradas
Vínculo institucional	VI	1 = vínculo sem data de término; 0,5 = somente vínculos encerrados; 0 = ausente

Fonte**:** elaboração própria (2026).

Os três componentes variam de 0 a 1 e recebem o mesmo peso. O IGCO é sua média aritmética: IGCO = (ME + CIE + VI) / 3. A ponderação uniforme atribui igual importância aos três usos considerados, como escolha operacional ainda não validada empiricamente. A sensibilidade à ponderação foi examinada comparando-se os escores obtidos com pesos iguais e com pesos de 0,40 para ME, 0,35 para CIE e 0,25 para VI, por meio da correlação de Spearman.

Para examinar como a escolha dos componentes influencia os resultados, foi analisada também uma versão exploratória mais ampla do índice. Essa versão incluía indicadores de presença de produção, completude descritiva das até 20 obras mais recentes, diversidade de tipos de produção, cobertura temporal, densidade de DOI e de identificadores externos e ausência de DOI duplicados. Também considerava o vínculo institucional, a consistência cronológica dos vínculos e a presença mínima de produção ou afiliação. Os indicadores foram organizados em três dimensões, com pesos de 0,40, 0,35 e 0,25. Como essa versão difere do IGCO final tanto nos componentes quanto na organização das dimensões e nos pesos, a comparação permite

[1] Nota explicativa: nos registros sem obras (n = 4), ME e CIE assumem valor zero.

observar o efeito dessas escolhas em conjunto, sem identificar a contribuição de cada uma delas separadamente.

As comparações entre as nove Grandes Áreas foram realizadas pelo teste de Kruskal-Wallis, considerando os 614 registros recuperados. Além de verificar a existência de diferenças estatisticamente significativas, foi calculado o tamanho de efeito, uma medida da magnitude dessas diferenças, utilizando o *eta* quadrado baseado na estatística do teste (Kassambara, [20--]). Esse procedimento foi aplicado ao IGCO final, aos seus componentes e à especificação exploratória.

Para facilitar a descrição dos resultados, os registros foram agrupados em três faixas de completude: baixa (IGCO < 0,60), intermediária (0,60 ≤ IGCO < 0,80) e alta (IGCO ≥ 0,80). Essas faixas foram adotadas apenas para organizar a apresentação dos resultados e não constituem uma classificação validada pela literatura ou estabelecida pelo ORCID.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS: RECUPERABILIDADE DOS REGISTROS

A estratégia em cascata permitiu atribuir registros ORCID a 614 dos 684 coordenadores (89,8%). Os 70 casos restantes (10,2%) foram classificados como silêncio informacional, nos termos definidos na metodologia.

Do total de registros recuperados, 102 foram identificados na busca estruturada completa, 285 na busca ampla e 227 na busca estruturada encurtada. A consulta mais restritiva, portanto, teria deixado de recuperar parcela expressiva dos registros. Quase metade das atribuições (289; 47,1%) decorreu de consultas com mais de um candidato, o que mostra que a desambiguação foi uma etapa relevante do procedimento.

4.1 COMPLETUDE FUNCIONAL DOS REGISTROS

Entre os registros recuperados, o IGCO médio foi de 0,902, com mediana de 1,000 e desvio-padrão de 0,158. A distribuição dos escores revela concentração em patamar elevado de completude: 463 registros (75,4%) situaram-se em alta completude, 138 (22,5%) na faixa intermediária e 13 (2,1%) na baixa.

Os componentes, contudo, apresentaram resultados diferentes. Metadados Essenciais (ME) alcançou média de 0,984 e desvio-padrão de 0,098; Cobertura de Identificadores Externos (CIE), média de 0,917 e desvio-padrão de 0,226; e Vínculo Institucional (VI), média de 0,806 e desvio-padrão de 0,394. Do total de registros recuperados, 117 (19,1%) não continham

informação pública em *Employment* e 4 apresentavam somente vínculos com data de término. O vínculo institucional concentrou, assim, a principal lacuna e a maior variação.

O número de obras públicas registradas variou bastante entre os registros. A mediana foi de 40,5 obras: 25% dos registros tinham até 11 obras e 75%, até 81, com valores variando de 0 a 1.350. Quatro registros não continham obras públicas e, por essa razão, receberam valor zero em ME e CIE. Essa diferença de volume é importante para interpretar os dois componentes. Como ME e CIE são calculados como proporções dentro de cada registro, dois registros podem obter o mesmo escore mesmo que um tenha poucas obras registradas e outro tenha centenas. Esses componentes medem, portanto, a completude relativa das obras disponíveis no ORCID, e não a quantidade de produção registrada nem a extensão da trajetória acadêmica do pesquisador.

O Vínculo Institucional (VI) também deve ser interpretado com cautela. O componente considera como indício de vínculo atual a existência de ao menos um registro em *Employment* sem data de término. Trata-se, portanto, de uma aproximação: o VI não verifica se esse vínculo corresponde à instituição do programa coordenado nem se a ausência de data de término significa, de fato, que o vínculo continua vigente. Da mesma forma, a ausência de informação em *Employment* não permite identificar sua causa. Ela pode estar relacionada à atualização realizada pelo próprio pesquisador, às configurações de visibilidade do registro ou aos mecanismos de integração entre a instituição e o ORCID. Assim, o VI permite identificar uma lacuna de informação institucional, mas não explicar por que ela ocorre.

Mais da metade dos registros analisados (53,4%) alcançou o valor máximo do IGCO. Esse comportamento caracteriza um efeito teto: quando muitos casos se concentram no limite superior da escala, o índice perde parte de sua capacidade de distinguir diferenças entre registros já muito completos segundo os critérios adotados. Isso decorre, em parte, da própria construção do IGCO, que se concentra em elementos básicos de completude funcional. Por essa razão, o índice é mais adequado para identificar lacunas e acompanhar a evolução dos registros do que para estabelecer distinções finas entre registros próximos do máximo.

Analisados em conjunto, os resultados mostram que os registros apresentam alta completude funcional, mas não de forma homogênea entre os componentes. Metadados Essenciais (ME) e Cobertura de Identificadores Externos (CIE) alcançaram valores próximos do máximo, enquanto Vínculo Institucional apresentou a menor completude e a maior variação. Um IGCO

elevado, contudo, não significa que o registro represente de forma completa toda a trajetória acadêmica do pesquisador. Significa apenas que, entre as informações publicamente disponíveis no ORCID, estão presentes os elementos definidos como necessários aos três usos examinados neste estudo.

4.2 SENSIBILIDADE DO ÍNDICE E VARIAÇÕES ENTRE GRANDES ÁREAS

No IGCO final (Tabela 1), as médias variaram de 0,854 em Linguística, Letras e Artes a 0,936 em Ciências Exatas e da Terra. A diferença entre as nove Grandes Áreas não foi estatisticamente significativa (Kruskal-Wallis: $H = 12,42$; $gl = 8$; $p = 0,134$; $\eta^2_H = 0,007$). A ausência de significância indica apenas que este desenho não detectou uma diferença global, mas não permite concluir que as áreas sejam equivalentes.

Na análise de sensibilidade à ponderação, os pesos alternativos produziram ordenação praticamente equivalente à dos pesos iguais (ρ de Spearman = 0,995). Já o contraste entre especificações apresentou diferença na magnitude do efeito entre Grandes Áreas: a especificação exploratória mais ampla apresentou tamanho de efeito de 0,106 ($p < 0,001$), enquanto o IGCO final apresentou 0,007 ($p = 0,134$). Como as especificações diferem em vários componentes e na estrutura das dimensões, esse contraste não permite atribuir a mudança a um indicador isolado nem separar o efeito da composição daquele da ponderação. O resultado sugere que as diferenças observadas entre áreas dependem da forma como a medida é construída.

Tabela 1 - IGCO e seus componentes segundo a Grande Área do Conhecimento[2]

Grande Área	N	Rec.	Sil. (%)	ME	CIE	VI	IGCO
Ciências Agrárias	78	75	3,8	0,983	0,956	0,807	0,915
Ciências Biológicas	88	72	18,2	0,984	0,920	0,764	0,889
Ciências da Saúde	112	101	9,8	0,987	0,922	0,713	0,874
Ciências Exatas e da Terra	84	78	7,1	0,977	0,927	0,904	0,936
Ciências Humanas	86	80	7,0	0,990	0,902	0,806	0,900
Ciências Sociais Aplicadas	73	64	12,3	0,989	0,865	0,828	0,894
Engenharias	67	60	10,4	0,994	0,956	0,817	0,922
Linguística, Letras e Artes	41	36	12,2	0,949	0,778	0,833	0,854

[2] Nota explicativa: N = coordenadores no universo; Rec. = registros recuperados; Sil. (%) = silêncio informacional sobre N. ME, CIE, VI e IGCO: médias calculadas sobre os registros recuperados.

Grande Área	N	Rec.	Sil. (%)	ME	CIE	VI	IGCO
Multidisciplinar	55	48	12,7	0,991	0,971	0,844	0,935
Total	**684**	**614**	**10,2**	**0,984**	**0,917**	**0,806**	**0,902**

Fonte: dados da pesquisa (2026).

Esse achado tem uma implicação metodológica importante. Quando um índice de completude inclui extensão temporal ou diversidade da produção, pode também captar a duração das carreiras e os padrões de publicação de cada campo. Com isso, diferenças próprias das trajetórias acadêmicas podem ser interpretadas como lacunas informacionais. Comparações entre áreas exigem, portanto, atenção à construção do indicador.

Entre os componentes do IGCO final, não houve diferença estatisticamente significativa entre áreas em Metadados Essenciais ($p = 0,569$) nem em Vínculo Institucional ($p = 0,174$). A Cobertura de Identificadores Externos apresentou diferença significativa ($H = 19,52$; $gl = 8$; $p = 0,012$), mas o efeito foi pequeno ($\eta^2_H = 0,019$): as médias variaram de 0,778 em Linguística, Letras e Artes a 0,971 na área Multidisciplinar. A variação na cobertura de identificadores externos pode refletir diferenças nos tipos de produção e nos canais editoriais entre campos. Estudos recentes também identificam variação disciplinar na presença de metadados ORCID (Arroyo-Machado *et al.*, 2026). O presente desenho, contudo, não permite separar esses mecanismos. O resultado não expressa qualidade científica; indica pequena variação na capacidade de conectar as obras a outras fontes.

5 CONSIDERAÇÕES FINAIS

Este estudo propôs o Índice Geral de Completude ORCID (IGCO) e o aplicou aos registros ORCID recuperados em um universo de 684 coordenadores de Programas de Pós-Graduação com notas 6 e 7 no resultado parcial da Avaliação Quadrienal CAPES 2021–2024. O índice reúne três componentes ligados a usos concretos do registro: descrição básica da produção, conexão das obras com outras fontes e informação de vínculo institucional.

Foram recuperados 614 registros; 70 coordenadores (10,2%) foram classificados em situação de silêncio informacional. Entre os registros recuperados, o IGCO apresentou média de 0,902. Na decomposição do índice, ME, CIE e VI apresentaram médias de 0,984, 0,917 e 0,806, respectivamente. Este último apresentou a menor média e a maior dispersão, configurando a principal lacuna no universo analisado.

Como resultado metodológico, o contraste exploratório entre especificações sugere que a escolha dos componentes e de sua agregação pode alterar as diferenças observadas entre áreas. A especificação mais ampla, que incluía elementos ligados à extensão temporal e à diversidade da produção, apresentou tamanho de efeito maior do que o IGCO final. O desenho não permite isolar a contribuição desses elementos nem demonstrar que a diferença decorra da senioridade dos pesquisadores. A interpretação deve, portanto, permanecer restrita à comparação entre as duas especificações.

O IGCO apresenta limites que precisam ser considerados em sua interpretação. A análise se restringe às informações públicas disponíveis no ORCID e não verifica a acurácia do conteúdo registrado; além disso, a recuperação dos identificadores depende de regras heurísticas de correspondência, sem conjunto de referência previamente validado ou validação manual independente. Como essas regras consideram informações de afiliação e, em parte dos casos, a presença de obras, a recuperação pode favorecer registros mais preenchidos. A magnitude e a direção desse possível viés não foram estimadas. O índice também apresenta efeito teto em 53,4% dos registros. Por sua vez, ME e CIE são calculados como proporções internas e, portanto, não incorporam o volume de obras associado a cada perfil. Essas limitações circunscrevem o alcance da medida e reforçam seu uso como instrumento de diagnóstico e linha de base, e não como recurso para classificar pesquisadores, programas, instituições ou áreas.

O estudo também distingue três níveis: existência do identificador, completude do registro e proveniência das informações. O IGCO opera no segundo nível. Ele permite observar se determinados elementos estão disponíveis, mas não informa se foram inseridos pelo pesquisador ou afirmados por uma organização. A proveniência deve ser examinada como dimensão complementar quando o objetivo envolver confiança e responsabilidade institucional sobre os dados.

Uma questão para estudos futuros é verificar se, e em que medida, processos de integração institucional se traduzem, ao longo do tempo, em registros mais completos e com proveniência mais bem definida. Essa perspectiva encontra paralelo em Porter (2022), que aponta que ambientes nacionais com sistemas maduros de informação sobre pesquisa oferecem melhores condições para a manutenção dos registros ORCID. Nesse sentido, o IGCO oferece uma proposta de medida para apoiar estudos métricos da informação e acompanhar processos de

qualificação de infraestruturas abertas de ciência, permitindo observar se a presença formal dos pesquisadores na base do ORCID se converte, de fato, em presença informacional efetiva.

REFERÊNCIAS

ALVES, Marcos Wesley Soares; SILVA, Waldeyr Mendes Cordeiro da; GRANDO, Rafaela Lora; JORGE, Vanessa de Arruda. Uso do ORCID como indicador persistente na construção do banco de dados de produção acadêmica da Fundação Oswaldo Cruz. **BiblioCanto**, Natal, v. 9, n. 2, p. 1-11, 2023. Disponível em: https://doi.org/10.21680/2447-7842.2023v9n2ID33661. Acesso em: 20 out. 2026.

ARROYO-MACHADO, Wenceslao; VARGAS-QUESADA, Benjamín; MUÑOZ-ÉCIJA, Teresa; CHINCHILLA-RODRÍGUEZ, Zaida. Exploring ORCID adoption and metadata presence in Spain's research landscape. **Scientometrics**, [*s. l.*], v. 131, n. 2, p. 805-823, fev. 2026. Disponível em: https://doi.org/10.3233/ISU-230218. Acesso em: 20 out. 2026.

BHATT, Rakesh Kumar; PAUL, Snehasish. The role of ORCID in advancing authorship, research visibility, and institutional integration in the digital knowledge ecosystem. **Annals of Library and Information Studies**,[*s. l.*], v. 72, n. 3, p. 315-322, 2025. Disponível em: https://doi.org/10.56042/alis.v72i3.20362. Acesso em: 20 out. 2026.

BRUCE, Thomas R.; HILLMANN, Diane I. The continuum of metadata quality: defining, expressing, exploiting. *In*: HILLMANN, Diane I.; WESTBROOKS, Elaine L. (ed.). **Metadata in practice**. Chicago: American Library Association, 2004. p. 238-256.

CAPES. **Resultado parcial da Avaliação Quadrienal 2021-2024**. Brasília, DF: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, 2026. Disponível em: https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/avaliacao-quadrienal/resultado-parcial-da-avaliacao-quadrienal-2021-2024. Acesso em: 18 fev. 2026.

CARNEIRO, Ana Maria; PINTO, Daniela Maciel; GOMES, João Gabriel Pedreira de Moura. A base ORCID e a representação da carreira docente no Brasil: potencial, cobertura e desafios. Biblios, n. esp., e014, 2025. Disponível em: https://doi.org/10.5195/biblios.2025.1267. Acesso em: 20 out. 2026.

EICHENLAUB, Naomi; MORGAN, Marina. ORCID: using API calls to assess metadata completeness. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2017, Virginia, EUA. **Anais** [...]. Virginia, EUA: DCMI, 2017. p. 104-107. Disponível em: https://doi.org/10.23106/dcmi.952137926. Acesso em: 20 out. 2026.

FERNÁNDEZ-MARCIAL, Viviana; GONZÁLEZ-SOLAR, Llarina; VALE, Ana. Is ORCID your ID? A case study at the Faculty of Arts and Humanities of the University of Porto. **Learned Publishing**, [*s. l.*], v. 36, n. 4, p. 564-576, jul. 2023. Disponível em: https://doi.org/10.1002/leap.1562. Acesso em: 20 out. 2026.

HAAK, Laurel L. ORCID: connecting researchers and scholars with their works. **Insights**, [*s. l.*], v. 26, n. 3, p. 239-243, nov. 2013. Disponível em: https://doi.org/10.1629/2048-7754.103. Acesso em: 20 out. 2026.

HEREDIA, Ana. The ORCID Brazil Consortium: building a community. **ORCID Blog**, [*s. l.*], 17 maio 2018. Disponível em: https://info.orcid.org/the-orcid-brazil-consortium-building-a-community. Acesso em: 1 set. 2026.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 25012:2008**: software engineering - software product quality requirements and evaluation (SQuaRE) - data quality model. Geneva: ISO, 2008.

KASSAMBARA, Alboukadel. **Kruskal-Wallis effect size**. [*S. l.*: *s. n.*], [20--]. Disponível em: https://rpkgs.datanovia.com/rstatix/reference/kruskal_effsize.html. Acesso em: 10 set. 2026.

KIRÁLY, Péter; BÜCHLER, Marco. Measuring completeness as metadata quality metric in Europeana. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA, 2018, Seattle. **Proceedings** [...]. Seattle: IEEE, 2018. p. 2711-2720. DOI: 10.1109/BigData.2018.8622487.

ORCID. **Member API**. [*S. l.*], [20--a]. Disponível em: https://info.orcid.org/what-is-orcid/services/member-api/. Acesso em: 2 set. 2026.

ORCID. **ORCID record schema**. [*S. l.*], [20--b]. Disponível em: https://info.orcid.org/documentation/integration-guide/orcid-record/. Acesso em: 2 set. 2026.

PIPINO, Leo L.; LEE, Yang W.; WANG, Richard Y. Data quality assessment. **Communications of the ACM**, [*s. l.*], v. 45, n. 4, p. 211-218, 2002. Disponível em: https://doi.org/10.1145/505248.506010. Acesso em: 20 out. 2026.

PORTER, Simon J. Measuring research information citizenship across ORCID practice. **Frontiers in Research Metrics and Analytics**, [*s. l.*], v. 7, p. 1-15, mar. 2022. Disponível em: https://doi.org/10.3389/frma.2022.779097. Acesso em: 20 out. 2026.

PORTER, Stephen R.; UMBACH, Paul D.; WILLIS, Chris. Understanding ORCID adoption among academic researchers. **Scientometrics**, [*s. l.*], v. 130, p. 2783-2797, 2025. Disponível em: https://doi.org/10.1007/s11192-025-05300-7. Acesso em: 20 out. 2026.

RNP. **RNP apresenta avanços do Consórcio ORCID Brasil em encontro internacional no Canadá**. Brasília, DF, 13 mar. 2026. Disponível em: https://www.rnp.br/2026/03/13/rnp-apresenta-avancos-do-consorcio-orcid-brasil-em-encontro-internacional-no-canada/. Acesso em: 1 set. 2026.

SCIMAGO RESEARCH GROUP. **Institution-asserted ORCID affiliations as indicators of research performance and institutional standing**. [*S. l.*]: SCImago Lab, 2026. Disponível em: https://www.scimagolab.com/wp-content/uploads/2026/02/ORCID-Report-III_Institutional-Trust-Markers.pdf. Acesso em: 3 set. 2026.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, [*s. l.*], v. 12, n. 4, p. 5-33, dez. 2015. Disponível em: https://doi.org/10.1080/07421222.1996.11518099. Acesso em: 20 out. 2026.

WANG, Tiange; LI, Zixuan; HUANG, Shan; YANG, Bo. Is ORCID a reliable source for CV analysis? Exploring the data availability of ORCID academic profiles. **Scientometrics**, [*s. l.*], v. 129, p. 1637-1662, fev. 2024. Disponível em: https://doi.org/10.1007/s11192-024-04944-1. Acesso em: 20 out. 2026.