



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



QUALIDADE INFORMACIONAL DE REGISTROS ORCID COMO FONTE PARA ESTUDOS MÉTRICOS DA INFORMAÇÃO:

construção de um índice multidimensional e implicações
para a infraestrutura científica brasileira

Robson Lopes de Almeida

<https://orcid.org/0000-0002-8925-049X> | robson.almeida@terceiro.rnp.br
Rede Nacional de Ensino e Pesquisa (RNP)

RESUMO:

O estudo propõe e aplica um conjunto de indicadores para avaliar a qualidade informacional dos registros ORCID de coordenadores de Programas de Pós-Graduação de excelência (Avaliação Quadrienal da CAPES 2021-2024). A partir de coleta automatizada via API pública e estratégia de busca em cascata, foram mensuradas as dimensões de completude, interoperabilidade e consistência estrutural, consolidadas no Índice Geral de Completude ORCID (IGCO). Os resultados revelam silêncio informacional em 10,4% dos perfis e concentração em patamar elevado de completude (IGCO médio = 0,80; 82,1% em alta completude), com subgrupo de baixa aderência que evidencia adoção ainda reativa. Discute-se o papel da integração institucional, no âmbito do Consórcio Brasileiro ORCID, para o fortalecimento da qualidade dos metadados e da autonomia informacional.

Palavras-chave: ORCID. Identificadores persistentes. Infraestrutura de dados. Qualidade informacional. Soberania científica.

EIXO TEMÁTICO:

Bases e Fontes de Dados

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.967

COMO CITAR ESTE ARTIGO:

ALMEIDA, Robson Lopes de. Qualidade informacional de registros orcid como fonte para estudos métricos da informação: construção de um índice multidimensional e implicações para a infraestrutura científica brasileira. *In*: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.967. Disponível em: <https://doi.org/10.22477/x.ebbc.967>.



1 INTRODUÇÃO

A crescente demanda por métricas nos processos de avaliação da ciência tem ampliado o debate sobre a qualidade e a governança das fontes de dados que sustentam o monitoramento da produção acadêmica. Em um cenário de plataforma progressiva da ciência, marcado pela centralidade de bases comerciais internacionais, a consolidação de infraestruturas abertas e interoperáveis torna-se estratégica para o fortalecimento da soberania científica.

Nesse contexto, o ORCID (*Open Researcher and Contributor ID*) consolidou-se como padrão internacional de identificação persistente de pesquisadores, viabilizando a desambiguação de autoria e a integração entre periódicos, agências de fomento e sistemas institucionais. No Brasil, sua adoção como ambiente interoperável vem sendo impulsionada por iniciativas articuladas, em especial o Consórcio Brasileiro ORCID, liderado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e operado pela Rede Nacional de Ensino e Pesquisa (RNP).

Apesar da ampla adesão ao ORCID, permanecem lacunas quanto à mensuração sistemática da maturidade informacional dos registros (densidade, padronização e capacidade de interoperabilidade dos metadados registrados). Diante disso, o presente estudo propõe e aplica um conjunto estruturado de indicadores, consolidado no Índice Geral de Completude ORCID (IGCO), com o objetivo de avaliar a completude e a interoperabilidade dos perfis e discutir suas implicações para os estudos métricos da informação e para a autonomia informacional em infraestruturas abertas.

2 BREVE REVISÃO DA LITERATURA

O ORCID iD consolidou-se como padrão global para desambiguação de autoria, integrando-se a fluxos editoriais, agências de fomento e bases comerciais. Contudo, sua adoção é frequentemente impulsionada por exigências externas: Porter, Umbach e Willis (2025) identificaram, em universidades norte-americanas, que 72% dos pesquisadores com ORCID o criaram sobretudo por demanda de periódicos ou agências, sem atualização sistemática posterior.

Krämer, Momeni e Mayr (2017) demonstram que a ampla adesão ao ORCID não garante cobertura ou integridade dos metadados, evidenciando lacunas relevantes para análises bibliométricas. A literatura converge para três dimensões críticas: completude, interoperabilidade e consistência estrutural. A integração via API de membros, como demonstrado pela Fiocruz (Alves *et al.*, 2023), permite validação e povoamento automatizado, superando a dependência de atualização manual. Permanecem, porém, escassas propostas de métricas sistemáticas sobre a qualidade estrutural dos perfis, lacuna que este estudo busca preencher.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa caracteriza-se como um estudo exploratório-descritivo, de abordagem quantitativa, voltado à avaliação de registros ORCID como fonte para estudos métricos. O objetivo foi propor e operacionalizar indicadores de qualidade estrutural dos perfis.

O universo compreendeu 684 coordenadores, após deduplicação de registros administrativos, de Programas de Pós-Graduação *stricto sensu* de excelência (notas

6 e 7 na Avaliação Quadrienal da CAPES 2021–2024). A unidade de análise foi o registro ORCID individual de cada pesquisador. A coleta ocorreu em fevereiro de 2026, por meio de rotinas automatizadas em Python, utilizando a API pública do ORCID (v. 3.0). Foram recuperados dados de produção científica (*works*), identificadores persistentes (DOI e outros PIDs) e vínculos institucionais (*employment*).

Para maximizar a taxa de revocação (*recall*) dos ORCID IDs e mitigar divergências de cadastro, adotou-se uma estratégia de busca em cascata com três etapas sucessivas: (i) busca estruturada completa, com todos os termos do nome; (ii) busca ampla com variantes lexicais e diacríticos; e (iii) busca encurtada, cruzando primeiro e último nome para tratar omissões de nomes do meio. Os registros resgatados nas etapas flexíveis foram submetidos à validação algorítmica de interseção de termos (*token set match*), com similaridade $\geq 85\%$, conjugada à validação de afiliação institucional e presença de produção registrada, de modo a assegurar precisão e reduzir o risco de falsos positivos.

A verificação da qualidade informacional baseou-se em abordagens que defendem a auditoria sistemática de registros ORCID quanto à presença de campos críticos e densidade de metadados (Morgan; Eichenlaub, 2017), bem como à consistência estrutural para análises métricas (Wang *et al.*, 2024).

A definição dos pesos fundamentou-se em critérios de relevância funcional para estudos métricos. Produção (0,40) é a condição primária de utilidade bibliométrica: sem obras cadastradas, o perfil é inutilizável para análises de produção, redes de coautoria ou mapeamento temático. Interoperabilidade (0,35) determina a capacidade de integração automatizada com outras infraestruturas, via DOIs e PIDs. Consistência/Confiança (0,25) desempenha função de validação estrutural: sua ausência penaliza inconsistências, mas sua presença isolada não amplia o

potencial de reuso.

A partir desse referencial, foi proposta a matriz de indicadores (Quadro 1), estruturada em três dimensões: Produção (presença, completude, diversidade e cobertura temporal); Interoperabilidade (densidade de PIDs, metadados e integridade); e Consistência/Confiança (completude e coerência dos vínculos).

Os indicadores DT e CTP são normalizados relativamente ao máximo observado na própria amostra (13 tipos distintos e 56 anos de amplitude, respectivamente), garantindo comparabilidade interna. Essa escolha implica, porém, que um único perfil discrepante possa definir o teto de normalização para toda a amostra, comprimindo os escores dos demais pesquisadores e limitando a interpretação individual dos resultados, limitação discutida na seção seguinte. Cabe registrar ainda que o atributo “Identificador Persistente”, incluído no CP como condição de identificabilidade mínima da obra, gera sobreposição parcial com os indicadores DD e DP da dimensão Interoperabilidade; a separação desses papéis está prevista em uma próxima versão do instrumento.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A aplicação da estratégia de busca em cascata revelou um cenário de desconexão entre registros administrativos nacionais e a infraestrutura global de identificação persistente. Ao final do processo, 89,6% ($n = 613$) dos coordenadores foram recuperados, enquanto 10,4% permaneceram não localizados ($n = 71$), configurando “silêncio informacional” significativo em um segmento reconhecido como de excelência acadêmica.

O resultado confirma as limitações estruturais da reconciliação baseada exclusivamente em nomes, conforme discutido por Krämer, Momeni e Mayr (2017), e reforça a centralidade de identificadores persistentes e de registros estruturalmente auditáveis como condição para a interoperabilidade sistêmica, conforme defendido por Morgan e Eichenlaub (2017). Cabe ressaltar que parte dos perfis localizados apresentou afiliação declarada divergente do vínculo registrado na CAPES, evidenciando a desatualização de metadados institucionais.

Quadro 1 - Matriz de Indicadores de qualidade e interoperabilidade ORCID

DIMENSÃO	INDICADOR (SIGLA)	DEFINIÇÃO OPERACIONAL E BASE DE CÁLCULO	FUNÇÃO ANALÍTICA
1. Produção (peso: 0,40)	Presença de produção (PP)	Verifica se o perfil possui ao menos um trabalho registrado. PP = 1 se total <i>works</i> > 0; caso contrário, não pontua.	Identificar perfis vazios.
	Compleitude da Produção (CP)	CP = média dos escores dos 20 trabalhos mais recentes, onde cada trabalho recebe 1 ponto para cada atributo presente (Título, Tipo, Ano, Identificador Persistente). Escore máximo por obra = 4. CP = (Σ atributos presentes nos 20 trabalhos mais recentes) / (4 × 20)	Avaliar riqueza descritiva mínima dos registros.
	Diversidade Tipológica (DT)	Razão entre a quantidade de tipos de trabalhos distintos (artigo, livro, tese, etc.) DT = n° tipos distintos / n° máximo observado	Medir a representatividade da produção.
	Cobertura Temporal da Produção (CTP)	Amplitude da carreira visível. CTP= (Ano Máx – Ano Min) / maior amplitude observada	Diferenciar perfis históricos de registros pontuais



DIMENSÃO	INDICADOR (SIGLA)	DEFINIÇÃO OPERACIONAL E BASE DE CÁLCULO	FUNÇÃO ANALÍTICA
2. Interoperabilidade (peso: 0,35)	Densidade de DOIs (DD)	Proporção de trabalhos que possuem um Digital Object Identifier (DOI) válido. DD = total works com DOI válido / total works	Medir a integração com a infraestrutura global.
	Densidade de PIDs (DP)	Proporção de trabalhos com qualquer tipo de identificador persistente (DOI, ISBN, Handle etc.). DP = total works com qualquer PID individual válido (DOI, Handle, ISBN etc.) / total works	Avaliar o potencial de rastreabilidade.
2. Interoperabilidade (peso: 0,35)	Metadados Essenciais (ME)	Porcentagem de trabalhos que contêm o mínimo descritivo obrigatório: Título + Tipo + Ano. ME = total works contendo Título + Tipo + Ano / total works	Verificar a citabilidade mínima do registro.
	Taxa de Duplicidade (TD)	Identifica a repetição de registros (trabalhos duplicados) através de DOIs idênticos. TD = (Total DOIs - DOIs únicos) / Total DOIs	Detectar falhas de curadoria e limpeza de dados.
3. Consistência/ Confiança (peso: 0,25)	Compleitude Institucional (CI)	Verifica se o pesquisador declarou o vínculo empregatício (<i>Employment</i>) mais atual. CI = 1 se vínculo ativo; 0,5 se apenas vínculo passado; 0 se inexistente	Validar vínculo atual do pesquisador.
	Consistência Temporal (CT)	Avalia a lógica cronológica, penalizando os vínculos com datas de fim anteriores ao início. CT = 1 = datas coerentes; 0 = inconsistência detectada	Detectar erro estrutural.
	Perfil Estrutural Mínimo (PEM)	Penaliza se a soma de (Trabalhos + Vínculos) for igual a zero. PEM = 1 se (works + vínculos > 0); 0 se apenas nome	Isolar perfis que contêm apenas o nome (meramente formais).

DIMENSÃO	INDICADOR (SIGLA)	DEFINIÇÃO OPERACIONAL E BASE DE CÁLCULO	FUNÇÃO ANALÍTICA
Índice Final	Índice Geral de Completude de ORCID (IGCO)	Média ponderada das três dimensões anteriores (Produção, Interoperabilidade e Consistência/Confiança) IGCO = (0,40 × Produção) + (0,35 × Interoperabilidade) + (0,25 × Consistência/Confiança) , onde cada dimensão é a média simples dos seus indicadores normalizados (0-1)	Classificar a qualidade global do perfil.

Fonte: dados da pesquisa.

Entre os perfis recuperados ($n = 613$), verificou-se que 503 (82,1%) atingiram o limiar de alta completude ($IGCO \geq 0,75$), indicando uso ativo da plataforma com densidade informacional adequada ao reuso sistemático. Em contrapartida, os 110 perfis de baixa completude (17,9%) distribuem-se em dois subgrupos: o primeiro reúne aqueles sem qualquer obra registrada (os chamados “perfis fantasmas”), desprovidos também de afiliações institucionais ativas, que confirmam a hipótese de adoção passiva descrita por Porter, Umbach e Willis (2025), segundo a qual o identificador é criado apenas para atender requisitos formais, sem manutenção sistemática posterior. O segundo conta com perfis que, embora possuam produção registrada, apresentam gargalos em interoperabilidade, como baixa densidade de DOIs ou ausência de vínculo institucional ativo.

No conjunto analisado, o IGCO apresentou média de 0,80 ($DP = 0,07$) e mediana de 0,83, sugerindo concentração de valores em patamar elevado. A menor dispersão reflete o efeito do protocolo de recuperação: ao exigir presença de produção ou afiliação confirmada como condição de aceite, o método reduziu a inclusão de perfis estruturalmente precários. Cabe ressaltar que o silêncio de 10,4% agrega dois fenômenos não separáveis: pesquisadores sem ORCID cadastrado e aqueles cujo registro existe, mas não foi localizado pela busca; essa distinção exigiria acesso a bases administrativas com ID declarado. Indicadores como $DD = 0,67$ e

CI = 0,72 revelam que, mesmo entre perfis de alta completude, há margem para qualificação.

A análise por dimensão revela que a Interoperabilidade apresentou desempenho médio superior (Score_Interop = 0,80), impulsionada pela alta cobertura de metadados essenciais (ME = 0,79) e densidade de PIDs (DP = 0,74). A dimensão Produção foi influenciada negativamente pelos indicadores DT = 0,15 e CTP = 0,25, valores que não refletem baixa qualidade produtiva, mas o efeito da normalização relativa, cujo denominador é definido pelo máximo da amostra (13 tipos distintos e 56 anos de amplitude). Um único perfil discrepante pode comprimir os escores de toda a amostra, o que reforça a necessidade de tetos fixos em aplicações comparativas futuras.

Embora a distribuição seja concentrada em patamar elevado, a coexistência dos dois subgrupos, perfis altamente interoperáveis e densos em metadados de um lado e perfis fantasmas ou estruturalmente precários de outro, evidencia que a maturidade informacional não é homogênea. Em contextos de automação e plataforma da ciência, tal assimetria reforça que o papel do ORCID iD como componente crítico da infraestrutura científica depende menos da adesão formal e mais da qualidade estrutural dos metadados.

5 CONSIDERAÇÕES FINAIS

A auditoria dos registros ORCID dos coordenadores de Programas de Pós-Graduação de excelência revelou gargalos estruturais tanto na recuperabilidade quanto na completude das informações. A adoção de uma busca em cascata, aliada à validação algorítmica baseada em similaridade nominal e na presença de produção

registrada, evidenciou a limitação das buscas estritamente literais. Além disso, o silêncio informacional de 10,4%, mesmo após a aplicação de estratégias mais flexíveis, indica restrições na interoperabilidade entre sistemas administrativos nacionais e a infraestrutura global de identificação.

A aplicação do IGCO revelou concentração em patamar elevado, com assimetria qualitativa entre os subgrupos. Embora 82,1% atinjam alta completude, a persistência de perfis fantasmas, identificadores criados sem preenchimento subsequente, evidencia adoção ainda reativa, não atribuível exclusivamente ao pesquisador. A normalização relativa de DT e CTP penaliza áreas monodisciplinares e perfis sem retroalimentação histórica, refletindo menor visibilidade informacional, não baixa qualidade científica; refinamento por tetos fixos está previsto em versão posterior.

A dependência de atualização manual configura fragilidade estrutural que, considerando o perfil de excelência desta amostra, tende a ser mais acentuada em estratos menos expostos da pós-graduação. A consolidação do ORCID exige integração via API de membros para que registros validados sejam automaticamente povoados, arranjo que o Consórcio Brasileiro ORCID está posicionado para viabilizar, elevando a qualidade dos metadados e fortalecendo infraestruturas abertas.

Cabe registrar, por fim, que o IGCO apresenta validade aparente, com componentes justificados por critérios funcionais e ancorados na literatura, mas ainda não foi submetido a validação externa. Essa validação está prevista como desdobramento da pesquisa no âmbito do Consórcio Brasileiro ORCID.

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. DOI: 10.21680/2447-7842.2023v9n2ID33661. Disponível em: <https://periodicos.ufrn.br/bibliocanto/article/view/33661>. Acesso em: 13 jan. 2026.

KRÄMER, Thomas; MOMENI, Fakhri; MAYR, Philipp. Coverage of Author Identifiers in Web of Science and Scopus. **arXiv**:1703.01319, 2017. DOI:10.48550/arXiv.1703.01319. Disponível em: <https://arxiv.org/abs/1703.01319>. Acesso em: 13 jan. 2026.

MORGAN, Marina; EICHENLAUB, Naomi. ORCID: Using API Calls to Assess Metadata Completeness. In: **INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS**, 2017. Proceedings... p. 104–107. DOI: 10.23106/dcmi.952137926. Disponível em: <https://dcpapers.dublincore.org/article/952137926>. Acesso em: 13 jan. 2026.

PORTER, Stephen R.; UMBACH, Paul D.; WILLIS, Chris. Understanding ORCID adoption among academic researchers. **Scientometrics**, v. 130, p. 2783–2797, 2025. DOI: 10.1007/s11192-025-05300-7. Disponível em: <https://link.springer.com/article/10.1007/s11192-025-05300-7>. Acesso em: 02 fev. 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**, v. 129, p. 1637–1662, 2024. DOI: 10.1007/s11192-024-04944-1. Disponível em: <https://link.springer.com/article/10.1007/s11192-024-04944-1>. Acesso em: 10 fev. 2026.