

PADRONIZADORES DE AUTORES: uso em base de dados

Ana Gabriela Clipes Ferreira

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

anaclipes@ufrgs.br

1 INTRODUÇÃO

Após a escolha do método para a coleta de dados nos estudos métricos, na etapa de tratamento destes, independente do tipo de base, há uma etapa essencial para a certificação da qualidade dos mesmos: a revisão e limpeza das informações coletadas, através de eliminação de registros duplos ou uniformização das autoridades, sejam institucionais ou autorais. Esta etapa despende até os dias atuais um desafio para os pesquisadores, como observado por Mugnaini (2006), bem como para a estratégia metodológica adotada no estudo, visando sempre minimizar os limites da pesquisa.

A padronização do nome de autores e entidades também demanda pesquisa e atenção na etapa de descrição dos autores na editoração dos artigos de periódicos, uma vez que os próprios autores podem se identificar de maneira incompleta ou diferente em algumas ocasiões, bem como pode haver acréscimo ou retirada de nomes do meio, alterações no sobrenome de autoras após mudança de estado civil, sobrenomes latinos (por vezes os sobrenomes espanhóis são confundidos com os brasileiros) e ambiguidades, em especial nos sobrenomes orientais, conforme pesquisa de Youtie et al. (2017), e finalmente os homônimos. Consequentemente, no trabalho cotidiano das bibliotecas e indexação das bases de dados essa inconsistência de autoridades gera recuperação da informação ineficiente ou equivocada. Assim, este trabalho busca identificar as principais características dos códigos padronizadores de nomes e como eles podem contribuir para os estudos métricos.



2 CÓDIGOS DE IDENTIFICAÇÃO DE AUTORES

Os identificadores de pesquisadores através de códigos são uma iniciativa similar ao Digital Object Identifier (DOI[®]), serviço da Crossref, porém visam identificar o produtor da informação (autor) e não o produto (artigo, livro, capítulo, entre outros objetos digitais). A persistência do número e ser único são características em comum. Porém, os identificadores de autores são gratuitos quando obtidos de maneira individual, enquanto o DOI é obtido através de assinatura.

Os mais conhecidos, difundidos e utilizados pela comunidade acadêmica são o Open Researcher and Contributor ID (ORCID) (https://orcid.org/content/about-orcid?locale_v3=pt), empresa sem fins lucrativos, e o ResearcherID (<https://clarivate.com/products/researcherid/>), da Clarivate Analytics. Pode-se inferir que o ORCID é o padrão mais utilizado no Brasil, visto que além de estar integrado com o ResearcherID e com o Currículo Lattes, foi adotado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) como um dos pré-requisitos para a inscrição em programas internacionais (CAPES, 2017), visando aperfeiçoar a avaliação e seleção destes.

3 METODOLOGIA

Estudo exploratório, tendo como método a busca de autores de diferentes maneiras e comparação dos resultados a fim de avaliar a funcionalidade dos identificadores e padronização das informações na base selecionada. Será utilizada a base de dados Web of Science (WoS) para a busca por autor por esta interagir com os identificadores de autores e por ser amplamente utilizada nos estudos métricos. Não pela sua cobertura de áreas ou regional, já que não inclui totalmente algumas áreas do conhecimento e também não abrange as regiões da mesma forma, mas pelos recursos oferecidos. O padronizador definido foi o ORCID, pois este agrega também o ResearcherID e está presente em um maior número de recursos no Brasil atualmente, além de ser uma iniciativa gratuita quando realizada a adesão do pesquisador de maneira individual.



As estratégias de buscas foram: primeira etapa, identificar os cinco autores mais produtivos em artigos científicos da Universidade Federal do Rio Grande do Sul (UFRGS) na WoS através da busca avançada, preenchendo o campo da seguinte maneira: OG=Universidade Federal do Rio Grande do Sul e pela sigla UFRGS. Em seguida, (a) o nome completo do autor em ordem indireta, de acordo com o Currículo Lattes, (b) os sobrenomes e prenomes abreviados conforme o padrão WoS e, (c) finalmente pelo ORCID, preenchendo o campo de busca avançada com AI=0000-0000-0000-0000, substituindo os zeros pelo respectivo identificador do pesquisador. A coleta foi realizada em 15 janeiro de 2018, sem delimitador de período temporal (ou seja, na base os anos de 1945 até 2018) e incluindo todos os idiomas. Foi recuperado e informado também o número de trabalhos registrados na base do ORCID, após consulta pelo ID do pesquisador. Por fim, os resultados foram organizados em um quadro para melhor visualização.

4 RESULTADOS

O número total de artigos com a busca pelo nome completo da instituição foi de 33.365 artigos. Porém, quando é utilizada a sigla, o número diminui consideravelmente para 2.014 artigos. Dessa forma, essa busca foi desconsiderada das análises, uma vez que não traz resultados fidedignos como os da primeira alternativa, provavelmente a forma mais utilizada. Os cinco autores mais produtivos são, nessa ordem: Moacir Wajner, Diogo Onofre de Souza, David Driemeier, Francisco Mauro Salzano e Ivan Izquierdo. Os resultados estão organizados no Quadro 1.



QUADRO 1 – AUTORES MAIS PRODUTIVOS DA UFRGS EM ARTIGOS NA WOS E OS RESULTADOS PELA BUSCA COM DIFERENTES ESTRATÉGIAS DE BUSCAS

Busca WoS 1		Busca WoS 2		Busca WoS 3		Artigos na Base	
						WoS	ORCID
WAJNER M	387	Wajner Moacir	187	0000-0001-6372-1807	0	387	0
SOUZA DO	346	Souza Diogo Onofre	84	0000-0002-4322-0404	314	346	137
DRIEMEIER D	320	Driemeier David	226	0000-0003-3766-0654	315	320	1
SALZANO FM	315	Salzano Francisco Mauro	17	0000-0002-4753-3184	0	315	0
IZQUIERDO I	278	Izquierdo Ivan	22	0000-0002-7208-0869	0	277	0

Fonte: Dados da pesquisa.

O ORCID, como alternativa de busca na WoS, ainda não é um campo eficiente na base, porém apresenta potencial para tanto. Já os resultados no banco de dados da ORCID demandam o preenchimento da maioria das informações pelos próprios pesquisadores, assim como o Currículo Lattes. Somente assim a base poderá estar integrada a outros sistemas de maneira mais eficiente. O resultado na WoS é sempre significativamente menor quando é realizada a pesquisa fora do padrão da base. A consistência se manteve nos demais quatro autores, com apenas um item a mais no último autor, quando utilizado o padrão WoS (Busca WoS1 e artigos na WoS). Nas buscas 2 e 3, bem como nos artigos na base do ORCID, é possível observar a divergência dos resultados.

5 CONSIDERAÇÕES FINAIS

Foi um estudo inicial, com objetivo de testar as funcionalidades e interações do ORCID com a WoS. A proposta do uso de identificadores é válida pelas vantagens já apresentadas e, uma vez consolidado no meio acadêmico, como tem acontecido nos últimos anos, a médio e curto prazo poderá proporcionar melhorias nos resultados das buscas por autor nas revistas científicas e bases de dados. O Open Journal Systems (OJS) em versões atualizadas possui metadados para preenchimento do ORCID, em-



bora não obrigatório, e os periódicos nacionais já começam a recomendar a inserção do identificador de autor no momento da submissão de artigos.

Na Scientific Electronic Library Online (SciELO), o ORCID também já é o identificador de autores utilizado em parte das revistas indexadas na biblioteca, sendo que o uso será obrigatório a partir de 2019 (PACKER et al., 2018). Nas bases de dados em geral, a padronização de autores através de identificadores poderá ampliar as alternativas de buscas, a exemplo da WoS, e padronizar as autoridades no seu catálogo. A persistência do número permite que as ambiguidades, omissões ou alterações de nome na busca de algum dado não afete os resultados.

Embora pareça trivial apresentar a consistência da recuperação da informação de uma base de dados amplamente utilizada nos estudos métricos, os resultados reforçaram a importância da credibilidade de um catálogo de autoridades com políticas bem definidas. Ao realizar a pesquisa por autor com um padrão diferente da WoS, os resultados foram muito diferentes, com baixa recuperação de itens, e a busca com o ORCID se mostrou ineficiente. Sugere-se, assim, o uso de analíticas de autores ou outro recurso que possibilite melhorar a recuperação e integrar as informações.

Um benefício do uso dos identificadores que pode-se inferir é o auxílio na elaboração de rankings institucionais e na avaliação individual de pesquisadores e de departamentos, dado que uma base de dados consistente fornecerá dados mais aproximados da realidade, com menos vieses. Ainda poderá minimizar as possibilidades de erros ou equívocos na recuperação da informação, sendo este um dos intuitos dos padronizadores de autores de uma forma em geral, com pouca ou nenhuma necessidade de limpeza de dados nas pesquisas.

REFERÊNCIAS

CAPES. **CAPES adota identificador digital para inscrição em programas internacionais**, 2017. Disponível em: <<http://www.capes.gov.br/sala-de-imprensa/noticias/8538-capes-adota-identificador-digital-para-inscricao-em-programas-internacionais>>. Acesso em: 27 dez. 2017.



MUGNAINI, R. **Caminhos para adequação da avaliação da produção científica brasileira: impacto nacional versus internacional**. 2006. 254 f. Tese (Doutorado em Cultura e Informação) - Universidade de São Paulo, São Paulo, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-11052007-091052/>>. Acesso em: 01 dez. 2017.

PACKER, A. L. et al. Los criterios de Indexación de SciELO se alinean con la comunicación en la ciencia abierta. **SciELO en Perspectiva**, 2018. Disponível em: <<http://blog.scielo.org/es/2018/01/10/los-criterios-de-indexacion-de-scielo-se-alinean-con-la-comunicacion-en-la-ciencia-abierta/>>. Acesso em: 14 jan. 2018.

YOUTIE, J. et al. Tracking researchers and their outputs: new insights from ORCID^s. **Scientometrics**, v. 113, n. 1, p. 437-453, 2017. <http://dx.doi.org/10.1007/s11192-017-2473-0>