



COBERTURA DE REVISTAS, ARTIGOS E CITAÇÕES EM BASES DE DADOS:

uma análise a partir do Portal de Periódicos da UFG

Érika Demachki¹
João de Melo Maricato¹

Resumo: Analisa comparativamente a cobertura de artigos com DOI do Portal de Periódicos da Universidade Federal de Goiás nas bases de dados Crossref, Scopus e Web of Science (WoS), em busca de identificar quais bases de dados cobrem melhor as revistas, artigos e citações. A Crossref indexa 100% das publicações, uma vez que todas possuem número de DOI, enquanto a Scopus cobre 10,9% e a WoS 3,3%. A cobertura de citação foi de 19,9% para a Crossref, 7,3% pela Scopus e 1,4% pela WoS, indicando a Crossref como a melhor base em termos de indexação e citação.

Palavras-Chave: Citação. Cobertura. Bibliometria. Bases de dados. Revistas científicas. Indexação.

1 INTRODUÇÃO

A bibliometria estuda aspectos quantitativos da produção, disseminação e uso da informação registrada (SUTCLIFFE, 1992). Dentre as técnicas, a análise de citação se tornou uma métrica bem consolidada para medição do impacto acadêmico. Atualmente diversos índices e indicadores elaborados a partir de citações são utilizados na avaliação da qualidade de revistas e artigos científicos.

A maioria dos estudos internacionais que utilizam indicadores bibliométricos coletam os dados de citações de bases de dados como *Web of Science* (WoS) e Scopus. Entretanto, à medida que outras bases de dados que coletam citações foram surgindo e se consolidando, diversos estudos passaram a analisar comparativamente as suas coberturas, sendo identificado que, muitas vezes, estas bases de dados são complementares, não havendo uma que contemple a totalidade dos dados (BAR-ILAN, 2010; NORIS; OPPENHEIM, 2007, MARTÍN-MARTÍN *et al.*, 2021).

Entretanto, mesmo que a WoS e a Scopus sejam consideradas pela comunidade acadêmica como as duas fontes mais confiáveis e competentes de dados bibliométricos, sua representatividade vem sendo questionada por diversas pesquisas (MELO; TRINCA;

¹ Universidade de Brasília (UnB)

MARICATO, 2021; TENNANT, 2020; VERA-BACETA *et al.*, 2019) que mostram que a maior parte dos artigos indexados por essas duas bases são provenientes de alguns poucos países ocidentais de língua inglesa. No caso da produção científica brasileira, onde os periódicos do país são responsáveis por publicar parte importante da ciência que desenvolve (PACKER, 2011), constata-se baixa cobertura das revistas e artigos indexados em bases de dados bibliográficas internacionais como a WoS e Scopus (MELO; TRINCA; MARICATO, 2021).

Mesmo com as recentes críticas que a WoS e Scopus vêm recebendo, ainda há uma dependência dos métodos de medição e avaliação da qualidade de pesquisa utilizados por estas bases, vistas como imparciais e de representatividade limitada (TENNANT, 2020). Estas continuam sendo utilizadas como parâmetro para definição de qualidade de periódicos, o que prejudica a avaliação de qualidade de periódicos brasileiros (e de outros países que tem sub-representação de revistas e artigos indexados nestas bases). Desse modo, há a necessidade de desenvolvimento de estudos bibliométricos aplicados ao contexto brasileiro (e regionais), de modo a serem investigadas outras fontes de dados e respectivas coberturas, para que sejam identificadas as que representem com maior fidelidade a cobertura de suas revistas, artigos e citações.

Nesse contexto, a Crossref realiza o registro de identificadores para objetos digitais, com o objetivo de torná-los fáceis de serem encontrados, citados, referenciados acessados e reutilizados. A sua base se mostra uma fonte de dados bibliométricos relevante para estudo da ciência brasileira. A Crossref disponibiliza, através de sua API, acesso automatizado aos indicadores disponíveis publicamente, possibilitando a coleta de dados bibliométricos a partir de metadados de relativa qualidade e confiabilidade, caracterizando-a como uma fonte de dados transparente. Uma vez que as instituições de ensino superior (IES) são responsáveis pela publicação de grande parte dos periódicos nacionais, este trabalho visa analisar comparativamente a cobertura de publicações, artigos e citações do Portal de Periódicos da Universidade Federal de Goiás (UFG) nas bases de dados WoS, Scopus e Crossref. Com isso busca-se responder a seguinte questão: quais as bases de dados que melhor cobrem revistas, artigos e citações do portal de periódicos da UFG, fornecendo indícios de que são mais adequadas para a avaliação da ciência desenvolvida no Brasil?

A partir deste questionamento, o objetivo deste trabalho é identificar quais as melhores bases de dados em termos de cobertura para revistas, artigos e citações de publicações do Portal de Periódicos da UFG, entre as bases WoS, Scopus e Crossref. A partir deste objetivo, os

seguintes objetivos específicos foram traçados: identificar a cobertura das bases de dados para revistas do portal; identificar a cobertura das bases de dados para artigos do portal; identificar a cobertura das bases de dados para citações do portal; e, investigar as correlações entre os indicadores de citação, nas três bases, dos artigos do portal.

2 METODOLOGIA

Este estudo é caracterizado como de natureza exploratória e descritiva, com abordagem metodológica quali-quantitativa. As bases de dados bibliométricas selecionadas para análise das coberturas de citação foram a WoS, Scopus e Crossref. A Crossref foi selecionada por disponibilizar acesso automatizado aos indicadores de citação, possibilitando uma coleta de dados transparente e consistente a partir de metadados de qualidade e confiáveis (diferentemente de outras fontes de dados bibliométricos, como o Google Scholar).

A coleta de dados deu-se em diferentes etapas, a partir da identificação inicial dos números de DOI registrados junto à Crossref para publicações de revistas científicas hospedadas no Portal de Periódicos da UFG, extraídos da *REST API* do Crossref (2020). Foram considerados DOIs atribuídos até outubro de 2020, totalizando 12.248 publicações com DOI de 32 revistas do portal. A lista de DOIs foi utilizada para coletar os números de artigos indexados e de citação das bases WoS e Scopus por meio da interface de busca avançada. Para a base Crossref os dados foram coletados através de sua API, utilizando-se a ferramenta Odisseia Metrics (RAMOS; MARICATO, 2020). As coletas dos indicadores de citação foram feitas em janeiro de 2022.

Os dados coletados destas três bases foram cruzados por meio do DOI, comum a todos os arquivos, para integrar todos os dados coletados em uma única planilha, possibilitando análises comparativas. Foram calculadas porcentagens, médias e medianas dos dados, com vistas a atender os objetivos específicos de identificar a cobertura das bases de dados para revistas, artigos e citações do portal da UFG, bem como as correlações de Spearman, em busca de atender o objetivo específico de investigar as correlações entre os indicadores de citação dos artigos do portal, verificando se as bases possuem dinâmicas de citação complementares.

3 RESULTADOS E DISCUSSÕES

Quanto à cobertura em nível de revista, somente oito (21%) das 32 revistas do Portal de Periódicos da UFG estão indexadas em pelo menos uma das duas fontes de dados analisadas. Considerando que todas as revistas do portal estão indexadas na Crossref, a Tabela 1 ilustra a cobertura de revistas que estão indexadas na WoS e/ou na Scopus.

Tabela 1 – Revistas e artigos do portal de periódicos da UFG indexadas na WoS e/ou Scopus

Revistas cobertas	Artigos totais	Artigos Scopus	% Scopus	Artigos WoS	% WoS
Ciência Animal Brasileira	823	530	64,4	0	0
Journal of Tropical Pathology	713	35	4,9	0	0
Pesquisa Agropecuária Tropical	689	515	74,7	145	21
Boletim Goiano de Geografia	628	0	0	42	6,7
Ateliê Geográfico	518	64	12,4	91	17,6
Sociedade e Cultura	513	117	22,8	43	8,4
Geoambiente On-Line	364	0	0	43	11,8
Música Hodie	119	74	62,2	45	37,8

Fonte: dados de pesquisa.

Das oito revistas listadas na Tabela 1, seis (16%) estão indexadas na WoS e seis (16%) na Scopus. As revistas com mais artigos publicados são a *Ciência Animal Brasileira*, da área de Ciências Agrárias, e a *Journal of Tropical Pathology*, de Ciências da Saúde, ambas indexadas apenas na Scopus e Crossref. A maior cobertura de indexação na Scopus é para a revista *Pesquisa Agropecuária Tropical*, da área de Ciências Agrárias, com 74,7% dos artigos indexados na base. Já na WoS, a maior cobertura de indexação dos artigos (37,8%), ocorre na revista *Música Hodie*, da área de Linguística, Letras e Artes, onde há cobertura de 45 dos 119 dos seus artigos.

A partir da análise das coberturas da Tabela 1, é possível perceber que as bases Scopus e WoS cobrem poucas revistas do Portal de Periódicos da UFG e, das poucas que são indexadas nestas duas bases, nenhuma possui uma cobertura integral de suas publicações. Portanto, a análise apenas da cobertura de revistas em bases de dados traz informações insuficientes para visualizar a sub-representação da ciência brasileira. Isso porque, além de cobrirem um número relativamente baixo de revistas, também não cobrem a totalidade de artigos publicados por elas. Das revistas indexadas pela base de dados Scopus, a cobertura média dos artigos é de aproximadamente 40%, e da WoS é ainda menor, 17,2%. O maior percentual de cobertura é na Scopus, de 74,7%, e o menor de 4,9% nesta mesma base (para a revista *Journal of Tropical Pathology*). As coberturas de artigos na WoS são, ainda menores, variando de 37,8% a 6,6%.

A cobertura de indexação dos artigos nas bases de dados consideradas por ser vista na Tabela 2. Naturalmente, 100% das publicações encontram-se indexadas na Crossref, pois todos os

artigos possuem número de DOI registrado na base, enquanto a cobertura na Scopus é de 10,9% dos artigos, e na WoS de apenas 3,3% deles.

Tabela 2 – Cobertura, média e mediana por fonte de dados para artigos do portal de periódicos da UFG indexados em cada fonte

Base de Dados	Artigos cobertos	% Cobertura de indexação	Total de citações	MeCAI	MdCAI
Crossref	12248	100	6134	0,5	0
Scopus	1336	10,9	4474	3,3	1
Web of Science	409	3,3	429	1	0

Nota: MeCAI = Média de citações por artigo indexado; MdCAI = Mediana de citações por artigo indexado.
Fonte: dados de pesquisa.

Ao analisarmos as médias e medianas de citações por publicação, é possível observar que a Scopus possui as maiores entre as publicações indexadas na base (Tabela 2), indicando um maior número de citações por publicação para artigos indexados por esta base, seguida pela WoS e pela Crossref, respectivamente. As maiores médias de citação por artigo nas bases de dados Scopus e WoS podem estar relacionadas a fatores como qualidade das revistas e grande potencial de divulgação proporcionada por estas duas bases. No entanto, as medianas (bastante próximas entre as três bases) sugerem que alguns poucos artigos são altamente citados, puxando a média para cima.

Das publicações analisadas, 22% possuem pelo menos uma citação captada por alguma das fontes de dados consideradas, mas, 77,9% dos artigos não receberam nenhuma citação. A base de dados Crossref conta com 19,9% das publicações com ao menos uma citação, sendo a que apresentou maior cobertura de citação identificada entre as três bases analisadas, seguida pela Scopus, com 7,3%, e pela WoS, com 1,4% (Tabela 3).

Tabela 3 – Cobertura de citação dos artigos do portal de periódicos da UFG por base de dados

Base de Dados	Cobertura de Citação*	%
Crossref	2448	19,9
Scopus	905	7,3%
Web of Science	178	1,4%
Total**	2703	22,0%

* Quantidade de artigos, em relação à totalidade (12.248) com ao menos uma citação.

** O Total é calculado a partir da totalidade de artigos do portal de periódicos.

Fonte: dados de pesquisa.

As bases WoS e Scopus possuem uma cobertura muito inferior à da Crossref, resultado que concorda com estudos anteriores que questionam o uso de indicadores provindos destas duas bases de dados bibliométricas para avaliar a qualidade de publicações brasileiras. Deve-se utilizar preferencialmente fontes de dados que possuam uma maior cobertura da produção científica brasileira, ou seja, que representam com maior fidelidade o

cenário nacional (MUGNAINI *et al.*, 2019). Portanto, a base de dados Crossref fornece indícios, a partir dos artigos do portal da UFG, de ser mais adequada para o monitoramento e avaliação das revistas, artigos e da própria ciência nacional, quando comparada com a Scopus e WoS.

Para analisarmos as relações entre as citações nas diferentes fontes de dados bibliométricas, foi utilizada a correlação de Spearman. A correlação entre citações da Crossref e da Scopus demonstrou uma força moderada de 0,425, maior do que as demais, conforme a Tabela 4.

Tabela 4 – Correlações de Spearman entre indicadores de citação da totalidade de artigos do portal de periódicos da UFG

	Crossref	Scopus	WoS
Crossref	1	0,425	0,157
Scopus	0,425	1	0,247
WoS	0,157	0,247	1

Fonte: dados de pesquisa.

Este resultado indica que há uma semelhança moderada entre as citações destas duas bases para artigos do Portal de Periódicos da UFG, evidenciando que as fontes podem captar as mesmas citações, mas como a correlação não é alta, cada uma delas também captam citações diferentes. Por outro lado, as correlações da WoS com as demais bases são muito baixas, indicando que há grande diferença entre os dados de citação entre ela e as demais bases. Um dos fatores que podem explicar isso é a diferença de indexação de artigos e revistas pelas três bases.

4 CONSIDERAÇÕES FINAIS

A cobertura de revistas, artigos e, por consequência citações, do portal de periódicos da UFG, nas bases de dados Scopus e WoS, é substancialmente mais baixa quando comparada à Crossref. Os dados confirmam a sub-representação da indexação das revistas do portal de periódicos da UFG na Scopus e WoS, o que, muito provavelmente, se estende a uma quantidade expressiva das revistas nacionais e da produção científica desenvolvida no país. Além do número de revistas indexadas pela Scopus e WoS serem baixas (16% em ambas), essas bases de dados não indexam a totalidade de artigos dessas revistas (40% e 17,2%, respectivamente). Comparando os dados da Scopus e a WoS, é possível afirmar que a cobertura da primeira é menos pior. A Crossref, naturalmente, cobre a totalidade de artigos e revistas com DOI do Portal da UFG. Do ponto de vista de cobertura de citações, foi possível verificar que 19,9% dos artigos foram citados ao menos uma vez na Crossref, porcentagem

muito superior às demais bases (inclusive pelo fato destas contarem com baixo número de revistas e artigos indexados). Apesar disso, pudemos notar que as médias de citações por artigos na Scopus e WoS são maiores do que na Crossref, porém as medianas são próximas. Os dados demonstram que as citações das bases de dados analisadas possuem correlações entre moderada e baixa entre elas, dando a entender que as citações capturadas pelas três bases de dados são diferentes e, portanto, provavelmente complementares. Porém, é evidenciado que, dentre as três bases de dados, a mais adequada para avaliar as revistas do portal de periódicos da UFG e, provavelmente, da ciência brasileira como um todo, é a Crossref. É importante frisar que esta pesquisa é um estudo de caso dos periódicos da UFG, não sendo possível inferir que a dinâmica apresentada representa a totalidade de revistas, artigos e citações brasileiras. Portanto, sugerimos que outras pesquisas semelhantes a esta sejam realizadas em outros contextos, para que se tenha uma visão mais completa da problemática no cenário nacional.

REFERÊNCIAS

- BAR-ILAN, J. Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar. *Scientometrics*, Budapest, v. 82, n. 3, p. 495-506, 2010. Disponível em: <https://doi.org/10.1007/s11192-010-0185-9>. Acesso em: 16 jan. 2022.
- CROSSREF. **Documentation**. [S. l.]: Crossref, 2020. Disponível em: <https://www.crossref.org/documentation/retrieve-metadata/rest-api/>. Acesso em: 5 nov. 2021.
- MARTÍN-MARTÍN, A.; THELWALL, M.; ORDUNA-MALEA, E.; LÓPEZ-CÓZAR, E. D. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, Budapest, v. 126, n. 1, p. 871-906, 2021. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03690-4>. Acesso em: 12 dez. 2021.
- MELO, J. H. N.; TRINCA, T. P.; MARICATO, J. M. Limites dos indicadores bibliométricos de bases de dados internacionais para avaliação da Pós-Graduação brasileira: a cobertura da *Web of Science* nas diferentes áreas do conhecimento. *Transinformação*, Campinas, v. 33, 2021. Disponível em: <https://doi.org/10.1590/2318-0889202133e200071>. Acesso em: 3 jan. 2022.
- NORRIS, M.; OPPENHEIM, C. Comparing alternatives to the Web of Science for coverage of the social sciences’ literature. *Journal of informetrics*, Amsterdam, v. 1, n. 2, p. 161-169, 2007. Disponível em: <https://doi.org/10.1016/j.joi.2006.12.001>. Acesso em: 16 jan. 2022.
- PACKER, A. L. Os periódicos brasileiros e a comunicação da pesquisa nacional. *Revista USP*, São Paulo, n. 89, p. 26-61, 2011.
- RAMOS, T. B. C.; MARICATO, J. M. **Odisseia métrics**: ferramenta de extração de dados. Versão 1.0. [S. l.: s. n.], 2020. Disponível em: https://colab.research.google.com/drive/1L75mJbqteVE0kMoTMK1_kLVXWFZpQumi#scrollTo=ohRqrEb7RXlm. Acesso em: 16 jan. 2022.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. **Information Processing & Management**, London, v. 28, n. 1, p. 1-3, 1992. Disponível em: [https://doi.org/10.1016/0306-4573\(92\)90087-G](https://doi.org/10.1016/0306-4573(92)90087-G). Acesso em: 21 out. 2021.

TENNANT, J. Web of science and scopus are not global databases of knowledge. **European Science Editing**, [S. l.], v. 46, 2020. Disponível em: <https://osf.io/preprints/socarxiv/qhvgr>. Acesso em: 04 jan. 2022.

VERA-BACETA, M. A; THELWALL, M.; KOUSHA, K. Web of science and scopus language coverage. **Scientometrics**, Budapest, v. 121, n. 3, p. 1803-1813, 2019.