



24º ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

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

COBERTURA DAS BASES SCIELO, WEB OF SCIENCE E SCOPUS: ANÁLISE COMPARATIVA DA PRODUÇÃO CIENTÍFICA BRASILEIRA

COVERAGE OF SCIELO, WEB OF SCIENCE AND SCOPUS DATABASES: A COMPARATIVE ANALYSIS OF BRAZILIAN SCIENTIFIC OUTPUT

Rogério Mugnaini – Universidade de São Paulo (USP)
Rafael Jeferson Pezzuto Damaceno – Universidade de São Paulo (USP)

Modalidade: Trabalho Completo

Resumo: A análise da cobertura de bases de dados é importante para compreender e avaliar a produção científica de um país. No Brasil, a base SciELO apresenta-se como uma importante fonte complementar às bases internacionais. O objetivo do estudo consistiu da análise da sobreposição e singularidade relativas das bases SciELO, Web of Science e Scopus na produção científica brasileira entre 2010 e 2020, ao longo do tempo e nas diversas áreas. A base SciELO foi tomada como complementar à Web of Science, permitindo que a comparação com a Scopus (com 624.535 artigos) tenha se dado em relação à Web of Science (589.139 artigos) e a Web of Science mais SciELO (672.702 artigos). A distribuição anual dos artigos mostrou que a SciELO propicia um aumento de cerca de 14% de artigos em relação à Web of Science, e que a expansão da Web of Science em 2015 altera significativamente a cobertura no período, com a criação do Emerging Sources Citation Index. A maior porcentagem de sobreposição se deu entre Web of Science e Scopus (90,1%), enquanto a menor foi entre Scopus e Web of Science (81,8%), revelando que a singularidade maior é da Scopus, mas se aproximando consideravelmente após 2015. A sobreposição dessas bases é mais alta nas ciências duras, enquanto a singularidade se revela mais claramente nas ciências brandas, o que sugere que estudos nestas últimas mereçam uma avaliação sobre o uso das bases combinadas, o que fica mais acentuado após a expansão da WoS. Por fim, sugere-se que pesquisas futuras aprofundem o estudo da cobertura das bases considerando os periódicos exclusivos de cada base, assim como é importante compreender se o aumento de singularidade se dá por um aumento de produção científica em periódicos do Brasil ou do exterior que passaram a ser indexados em determinada base.

Palavras-chave: produção científica do Brasil; cobertura; bases de dados.

Abstract: Analyzing the coverage of databases is important for understanding and evaluating a country's scientific output. In Brazil, the SciELO database is an important complementary source to international databases. The aim of the study was to analyze the relative overlap and singularity of SciELO, Web of Science and Scopus databases in Brazilian scientific output between 2010 and 2020, over time and in different areas. SciELO database was taken as complementary to Web of Science, allowing the comparison with Scopus (with 624,535 articles) to be made in relation to Web of Science (589,139 articles) and Web of Science plus SciELO (672,702 articles). The annual distribution of articles

showed that SciELO provides an increase of around 14% more articles than Web of Science, and that the expansion of Web of Science in 2015 significantly changes the coverage in the period, with the creation of the Emerging Sources Citation Index. The highest percentage of overlap was between Web of Science and Scopus (90.1 %), while the lowest was between Scopus and Web of Science (81.8 %), revealing that Scopus has the greatest singularity, but has come considerably closer since 2015. The overlap between these databases is higher in the hard sciences, while singularity is more evident in the soft sciences, which suggests that studies in the latter deserve an evaluation of the use of combined databases, which has become more pronounced since the expansion of Web of Science. Finally, it is suggested that future research study the coverage of the databases in greater depth, taking into account the journals exclusive to each database, and it is also important to understand whether the increase in singularity is due to an increase in scientific production in journals from Brazil or abroad that have been indexed in a particular database.

Keywords: Brazilian scientific output; coverage; databases.

1 INTRODUÇÃO

A escolha da base de dados para realização de análises bibliométricas é uma questão essencial, pois pode influenciar significativamente os resultados obtidos. Isto se dá principalmente devido à cobertura das mesmas, relacionada a sua abrangência (geográfica, temporal, temática, linguística, tipologia documental, dentre outras características). Adicionalmente, deve-se considerar o tipo de informação disponível: referencial ou de texto completo; ou a exaustividade de campos bibliográficos específicos, como o de “autor”, no que diz respeito à disponibilidade dos nomes de todos os autores, além da completude e normalização de seus nomes e instituições de afiliação (Costas; Moreno; Bordons, 2008).

Outro aspecto relaciona-se ao custo ou modalidade de acesso aos documentos, que nem sempre são declarados, e quando é o caso, a informação deste campo pode ser alterada ao longo do tempo – como é o caso da Web of Science (WoS), que obtém tais informações de uma parceria com a OurResearch¹.

Convém ainda considerar que, em se tratando de análise de citação, as exigências vão além da cobertura, sendo importante a qualidade dos dados e as ferramentas oferecidas pela base para busca, recuperação e análise de informação (Neuhaus; Daniel, 2008).

No passado, os estudos de cobertura se atinham aos periódicos científicos considerados nas fontes secundárias, que consistiam dos produtos dos serviços de resumo ou indexação. Da década de 1930, podem ser mencionados os estudos de Samuel C. Bradford, mas é na década de 1960 que John Martyn evidencia as primeiras contribuições para a proposição de uma teoria e análise de sobreposição de periódicos. Era o início dos sistemas e

¹ <https://webofscience.help.clarivate.com/en-us/Content/open-access.html>

algoritmos de busca, sem contar a ausência de padrão dos formatos de registros bibliográficos daquele momento, e o crescimento vertiginoso da literatura primária (Gluck, 1990).

Outra limitação que merece destaque diz respeito ao acesso às bases de dados, que dependia de pagamento das horas despendidas para acesso e recuperação de informação – desconsiderando aqueles estudos que dependiam de levantamento em fontes impressas –, justificando assim se recorrer a estudos amostrais. Além do que, se recorria a métodos variados para avaliação de sobreposição entre bases de dados, com vistas à avaliação de sua cobertura, sempre buscando garantir custo e benefício (Tenopir, 1982). Segundo Gluck (1990), é na década de 1970 que começam a surgir estudos exaustivos, eliminando as críticas feitas anteriormente a problemas estatísticos relacionados ao cálculo de erros amostrais.

Com o advento da Internet e a disponibilidade das bases de dados on-line o acesso foi significativamente facilitado para pesquisadores, no caso de bases comerciais, que contavam com o benefício de uma assinatura institucional. Contudo as bases comerciais internacionais e multidisciplinares existentes na virada do último milênio continham vieses, apresentando uma cobertura mais representativa para países do Norte Global e de língua inglesa. E como se podia notar, não apenas países do Sul Global ficavam condenados à invisibilidade (Gaillard, 1989; Gibbs, 1995). O preocupante é que a situação persiste no presente, conforme estudo de Asubiaro, Onaolapo e Mills (2024), evidenciando a centralidade da América do Norte e Europa e do idioma inglês.

Assim, com o surgimento de bases de dados temáticas e de abrangência regional ou local e suas variadas coberturas e políticas, estudos em áreas específicas precisam ser empreendidos a fim de compreender como diferentes bases se complementam ou sobrepõe, na representatividade da produção científica de uma região, como do estudo de Costas, Moreno e Bordons (2008) em Ciências da Saúde.

A seletividade da base WoS foi um dos fatores preponderantes para a criação do SciELO (Packer, 1998), além do esforço para que o Brasil pudesse contar com um índice de citação. No ano de 2014 o SciELO Citation Index é criado na plataforma de bases da WoS, mas seu uso para complementar a WoS já era verificado anteriormente (Mugnaini; Digiampietri; Mena-Chalco, 2014), e segue ocorrendo (Mugnaini; Igami; Krzyzanowski, 2022).

A Scopus chega em 2004 com uma proposta de incluir os periódicos já indexados na base SciELO, contudo esta promessa nunca foi escrita. Além disso, sua vantagem em relação

à WoS para representatividade da produção brasileira sempre foi aventada, aparentemente pelo fato de reunir um número maior de periódicos da região.

No mesmo ano o Google Acadêmico foi lançado, e sempre atrai atenção dos estudos bibliométricos e cientométricos. Não demorou muito para que Neuhaus e Hans-Dieter (2008) apresentassem uma detalhada análise comparativa entre índices de citação, enumerando diversas limitações do Google Acadêmico em relação a seus concorrentes, ainda que os autores reconheceram que o mesmo tinha potencial para se tornar uma poderosa ferramenta. O estudo de Martín-Martín *et al.* (2018) explorou a singularidade do Google Acadêmico em relação à WoS e Scopus, que evidenciava a presença de outros idiomas nas diversas áreas do conhecimento, além de uma vantagem de número de citações – especialmente em relação à WoS. Gusenbauer (2019) envidou esforços para estimar o tamanho do Google Acadêmico, evidenciando ser o maior entre diversas bases, além de seu crescimento mensal ser superior ao Microsoft Academic. E mais recentemente, Pereira e Mugnaini (2023) analisaram o interesse da comunidade especialista em Bibliometria e Cientometria no Google Acadêmico para análises avaliativas baseadas em publicações e citações, reconhecendo o crescente interesse nas ferramentas lançadas em 2011 e 2012, respectivamente, de citações e métricas.

Atualmente, o surgimento de outras fontes tem agregado o benefício do acesso via Interface de Programação de Aplicativos (API, sigla em inglês) – como Dimensions e OpenAlex – permitindo recuperar grande volume de dados, além de diversidade de tipologia documental, linguística e geográfica.) empreenderam uma análise dos metadados oferecidos por estas e outras seis bases, com base numa amostra de dados extraída do Crossref. Eles concluíram que bases baseadas em fontes externas – como é o caso de Dimensions, OpenAlex, The Lens e Scilit – podem oferecer mais e melhores metadados que do que motores de busca acadêmicos que extraem informação da rede – como o Google Acadêmico, Microsoft Academic e Semantic Scholar. Segundo os autores a Dimensions apresentou o maior número de campos bibliográficos e o maior grau de completude, enquanto que dentre os motores de busca o destaque foi da Microsoft Academic.

A disponibilidade de bases e dados mais amplas e suas limitações associadas tem sido foco das pesquisas atuais da comunidade científica, principalmente com a OpenAlex, contudo, a comparação com as bases WoS e Scopus segue sendo relevante (Culbert *et al.*, 2024), pois tais bases ganham proeminência entre a comunidade científica pelo fato de serem utilizada em processos avaliativos.

Por isso é importante observar a cobertura oferecida pelas mesmas, considerando a produção científica de áreas e países. Singh *et al.* (2021) verificam que as bases Scopus e Dimensions ofereciam alguma vantagem em relação à WoS – considerando a produção científica brasileira –, para as ciências sociais, e em menor magnitude para Artes e Humanidades, mas seus achados evidenciam a necessidade uma investigação mais aprofundada. Por essa razão, é importante comparar as coberturas proporcionada pelas bases SciELO, WoS e Scopus.

Em âmbito nacional, é difícil ter um levantamento exaustivo, que permita avaliar a cobertura oferecida pelas diversas bases de dados nas diferentes áreas. No Brasil a Plataforma Lattes é uma fonte compreensiva da produção científica nacional, permitindo avaliar mais exaustivamente a representatividade das bases, mensurando o volume e percentual de periódicos e artigos fora das bases SciELO, WoS e Scopus. Mugnaini *et al.* (2019) verificaram que 60% dos periódicos utilizados para publicação não estavam indexados nas referidas bases (mais de 23 mil periódicos) e publicaram 41% dos artigos, enquanto que os periódicos indexados na SciELO foram responsáveis por 21% da dos artigos, apesar de representarem apenas 3% dos periódicos utilizados.

Nesse sentido vale a menção aos *Current Research Information Systems* (CRIS), que vem sendo integrados nacional e internacionalmente, podendo servir para este propósito. CRISTIN (o exemplo da Noruega) apresenta vários anos de experiência lidando com qualidade e completude dos dados. Uma parte do CRISTIN é denominada *Norwegian Science Index*, que busca assegurar completude e qualidade de dados, possibilitando análises bibliométricas (Sivertsen, 2019). Iniciativa como esta vai além de um repositório compreensivo da produção nacional, pois permite identificar a parte da pesquisa publicada em fontes mais seletivas – do ponto de vista avaliativo isto é importante.

A cobertura de produção científica de um país nas bases WoS e Scopus é objeto de pesquisas de muitos pesquisadores, em virtude da adoção dessas fontes em processos avaliativos. Sendo a WoS mais antiga, muita atenção era dada a ela na literatura, até que a Scopus começa a ocupar seu espaço. A WoS então reage, aumentando a cobertura global de periódicos entre 2005 e 2010 (Testa, 2011), evidenciando dois efeitos: a produção brasileira naquela base aumentou significativamente após indexação de periódicos do país (Leta, 2012); o português se tornou o segundo idioma mais frequente (após o inglês) dos artigos publicados

em periódicos latino-americanos (posição até então ocupada pelo espanhol), entre 2005 e 2011, graças ao aumento de periódicos da região (Collazo-Reyes, 2014).

Posteriormente, a partir de 2015 outro aumento da produção brasileira foi marcante na WoS, quando o Emerging Sources Citation Index – que reúne periódicos que ainda não são aceitos nos índices tradicionais – passou a compor a WoS Core Collection, incrementando a produção brasileira em 13,3% em 2015 e em 19,6% em 2016 (Mugnaini; Igami; Krzyzanowski, 2022).

Do cenário delineado, pode-se depreender que a análise de cobertura de bases avançou sobremaneira ao longo de quase um século, e que o contexto atual evidencia a necessidade de estudos que deem conta de delinear o panorama oferecido pelas bases em sua totalidade, ou sua representatividade em relação à produção científica de um país. Outro aspecto que merece atenção é a diferenciação entre as diversas áreas do conhecimento, assim como análises temporais, já que os conteúdos das bases estão continuamente expandindo (ou pelo menos espera-se que isto esteja ocorrendo) com vistas a uma cobertura mais adequada.

Ainda que outros estudos já tenham abordado este problema, pelo fato de a política científica brasileira seguir exigindo prioritariamente a publicação em periódicos da WoS e Scopus – ao basear seus critérios de excelência em indicadores de tais bases –, é importante explorar a complementaridade das mesmas, oferecendo aos pesquisadores uma visão específica e atualizada que possa guiá-los na busca de periódicos para publicação.

O presente estudo buscou analisar a representatividade da cobertura das bases SciELO, WoS e Scopus em relação à produção científica brasileira no período de 2010 a 2020, considerando sua sobreposição e singularidade ao longo do período e nas diferentes áreas.

2 METODOLOGIA

Do ponto de vista de seus objetivos a pesquisa é exploratória e sua abordagem é quantitativa. A pesquisa é de natureza básica, que conforme Glänzel (2003), é produzida pelos pesquisadores do grupo 1 (próprios "bibliometristas"), realizando pesquisa metodológica, sendo de potencial interesse dos demais grupos: grupo 2 (pesquisadores das áreas), para quem os resultados servirão para pesquisas aplicadas; grupo 3 (gestão e política científica), por facilitar a visão panorâmica da estrutura da ciência nacional e sua análise comparativa.

Outro aspecto relevante é que a presente proposta se enquadra no nível mais amplo de especialização dos estudos bibliométricos, por oferecer insumo relevante para a avaliação

de pesquisa em âmbito nacional, ao determinar a parcela de artigos indexados nas principais bases consideradas pela política científica nacional (Cox *et al.*, 2019).

O período do estudo abrange 11 anos (2010-2020), que corresponde à última trienal (2010-2012) de avaliação CAPES, e as duas quadrienais (2013-2016 e 2017-2020). A vinculação aos períodos dos ciclos avaliativos da CAPES tem sido observada na pesquisa, com o intuito de considerar os critérios avaliativos propostos pelas áreas de avaliação como fatores potenciais para interpretação das tendências observadas, quando pertinente.

A produção científica brasileira do estudo consiste dos artigos originais ou de revisão com pelo menos um autor com afiliação institucional do Brasil. Seguindo tais critérios, no referido período, foram identificados: 589.139 artigos na WoS; 672.702 artigos na WoS+SciELO (desconsideradas as sobreposições), e; 624.535 artigos na Scopus.

Os dados da WoS vêm sendo obtidos on-line ao longo dos anos – e usados em projetos de pesquisa sucessivos –, mediante recuperações sucessivas de acordo com o limite de registros permitido por seus mantenedores. Os dados da SciELO foram obtidos mediante solicitação à respectiva instituição, que permitiu a obtenção dos dados via servidor. Os dados da Scopus foram obtidos do ICSR Lab, mediante aprovação de projeto de pesquisa em uma chamada internacional.

O indicador utilizado foi o de sobreposição relativa (Gluck, 1990) para comparação dos pares de bases:

$$\% \text{ sobreposição em } A = 100 \times \frac{|A \cap B|}{|A|}$$

$$\% \text{ sobreposição em } B = 100 \times \frac{|A \cap B|}{|B|}$$

onde A representa, ora WoS e ora WoS+SciELO, e B representa Scopus.

Pelo fato da comparação ter se dado entre pares de bases, o complemento do % de sobreposição relativa de A, é a parte exclusiva, ou seja, a singularidade de A. A análise explorou este aspecto, para mensurar quanto da produção brasileira cada base aporta, em oposição ao que as mesmas têm em comum.

A identificação do volume de artigos nas bases se deu a partir do código ISSN do periódico, em cada ano do estudo. A consideração do ano visou garantir a checagem de sobreposição ao longo do período do estudo, uma vez que a indexação pode ter ocorrido nesse ínterim. Para garantir que o código ISSN de um mesmo periódico coincidissem entre as bases, recorreu-se ao ISSN-link, comumente utilizado em situações deste tipo.

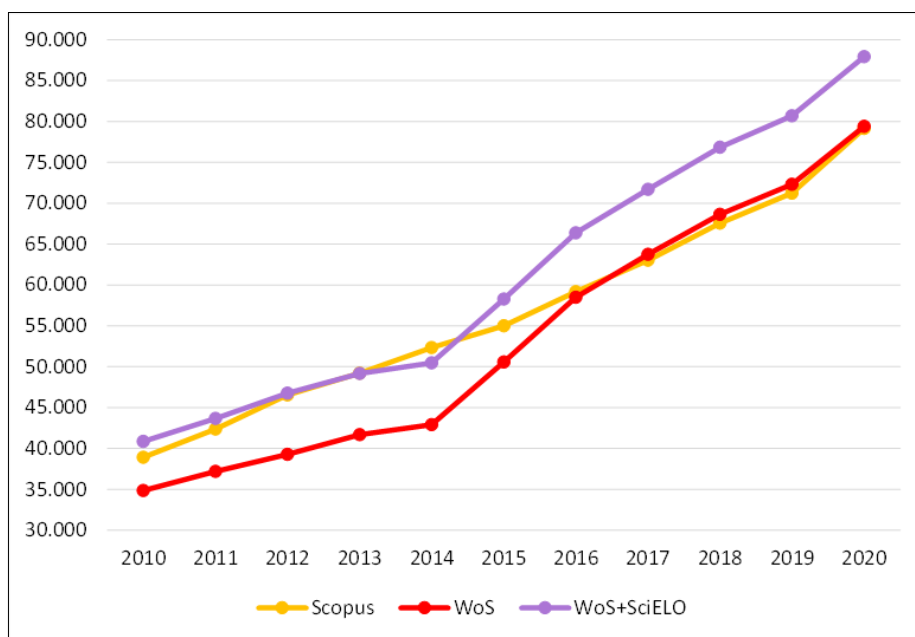
Para a análise nas diversas áreas, recorreu-se à classificação Essential Science Indicators (ESI) para análise da produção na WoS e WoS+SciELO, e à All Science Journal Classifications (ASJC) para a produção na Scopus. Por esta razão, e pelo fato de centrar-se na singularidade, não houve necessidade de compatibilização das categorias de cada classificação de periódicos que compõem a sobreposição (ou seja, indexados em mais de uma base).

A análise buscou detalhar a singularidade entre anos e áreas, especialmente o efeito da expansão da cobertura da WoS após inclusão do Emerging Sources Citation Index na WoS Core Collection.

4 RESULTADOS

A distribuição anual dos artigos nas bases é apresentada na Figura 1, podendo-se notar que a inclinação na WoS muda a partir de 2015, como consequência do já referido aumento de cobertura da base (Mugnaini; Igami; Krzyzanowski, 2022). O incremento da SciELO pode ser observado na distância da linha roxa em relação à amarela, representando um adicional de 6 mil artigos no início do período e chegando a 8,5 mil no final. Equivalendo a 14% em relação à WoS, o incremento diz respeito à produção em periódicos indexados exclusivamente na SciELO, pois aqueles indexados em ambas as bases já estão considerados na categoria WoS.

Figura 1 – Números absolutos da produção científica brasileira nas bases



Fonte: Elaborado pelos autores

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

A base Scopus apresenta uma taxa de crescimento constante em todo o período, evidenciando uma quantidade muito próxima da WoS após a expansão de sua cobertura. Antes de 2015, porém, o montante na Scopus era similar ao da WoS+SciELO.

Apesar das magnitudes observadas, importa avaliar se os conteúdos são similares, ou seja, se o percentual de sobreposição relativa a cada base. A Tabela 1 permite observar que o percentual de sobreposição do número de artigos da WoS+SciELO em relação à Scopus é inferior ao da WoS em relação à Scopus, evidenciando que o SciELO agrega periódicos exclusivos também quando comparada com a Scopus. A parte da produção exclusiva da SciELO representa a singularidade desta base para a produção brasileira.

Considerando a base Scopus, observa-se uma maior sobreposição com WoS+SciELO do que quando comparada somente com WoS, e como era de se esperar, sua singularidade é mais pronunciada quando SciELO não é adicionada.

Tabela 1 – Porcentagem de sobreposição da produção científica brasileira entre as bases

	WoS+SciELO	WoS	Scopus
WoS+SciELO			86,0%
WoS			90,1%
Scopus	89,0%	81,8%	

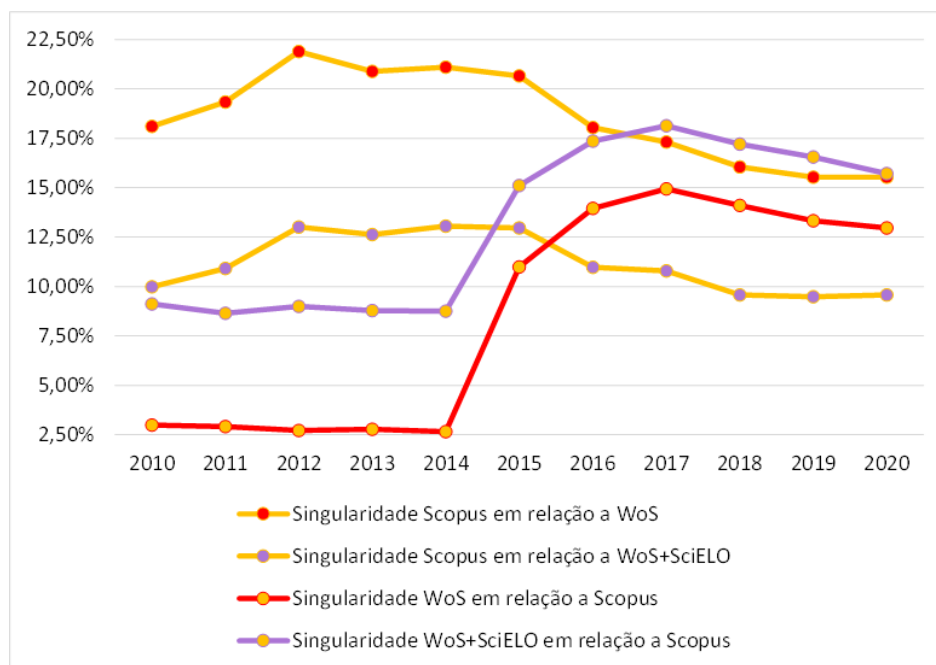
Fonte: Elaborado pelos autores

Analisando a sobreposição pode-se concluir que as bases têm alto percentual do número de artigos em comum, contudo, importa avaliar se esses percentuais variam ao longo dos anos do estudo. Outra questão importante a se considerar é se a produção em áreas específicas apresenta percentuais homogêneos de sobreposição entre as bases. Se isto não for confirmado, restará ver em quais áreas deve-se priorizar uma base específica, ou se mais bases em conjunto.

No entanto, ao prosseguir as análises, o foco será dado na singularidade das bases, ou no percentual de seu conteúdo exclusivo, sempre considerando o número de artigos. A Figura 2 denota que antes de 2015 as bases apresentavam percentuais díspares de singularidade. A Scopus em relação à WoS apresentava o maior percentual (ao redor de 20%), revelando queda após 2015 e chegando a cerca de 15% em 2020. A Scopus em relação à WoS+SciELO apresentou perfil similar, contudo diminuiu de cerca de 12,5% para 10% após 2015. Na direção oposta, WoS e WoS+SciELO em relação a Scopus apresentaram tendência praticamente constante antes de 2015, aumentando significativamente sua singularidade após 2015.

A partir de 2016 o percentual de singularidade das bases torna-se similar, mas da mesma forma que na comparação de Singh *et al.* (2021), os percentuais não especificam as diferenças. Nesse sentido, as Figuras 3 e 4 apresentam a singularidade nas áreas de cada base.

Figura 2 – Singularidade das bases na cobertura da produção científica brasileira



Fonte: Elaborado pelos autores

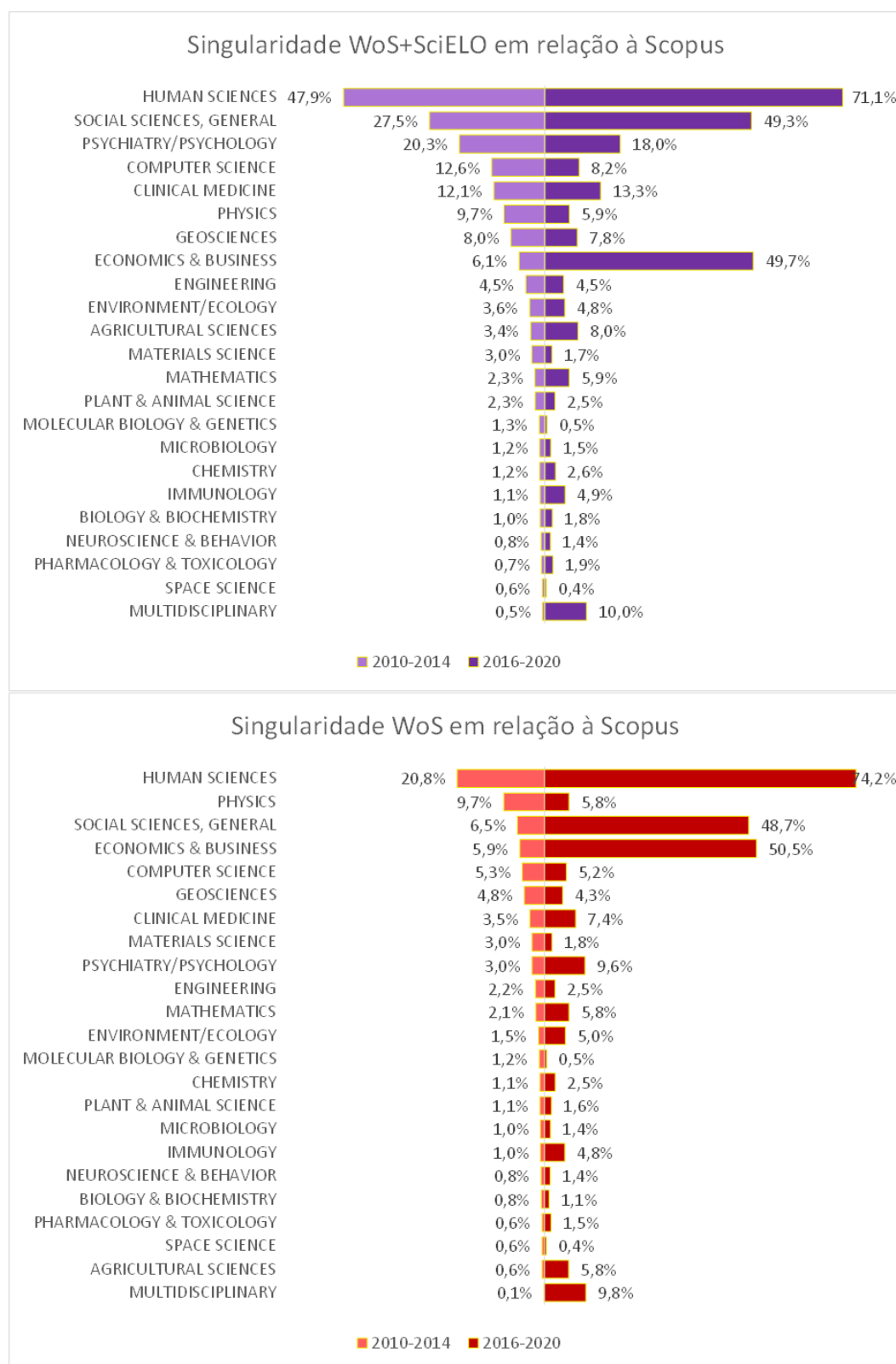
Na Figura 3 observa-se que o período posterior à expansão da WoS Core Collection evidencia um aumento significativo do percentual de singularidade para algumas áreas, não importando a consideração de WoS+SciELO ou somente a WoS: Ciências Humanas, Ciências Sociais e Economia & Negócios. Contudo, no período anterior à expansão, os percentuais de singularidade eram mais evidentes quando considerada a base SciELO, o que sugere que parte dos periódicos do SciELO não estavam indexados na Scopus, podendo-se destacar as seguintes áreas: Psiquiatria/Psicologia, Ciências da Computação e Medicina Clínica.

Outro aspecto que merece atenção é a ordem das áreas em cada gráfico da Figura 3, a partir da singularidade no período anterior à expansão da WoS, como é o caso de áreas que se destacam em singularidade ao se considerar a SciELO: Ciências Agrárias sobe 11 posições pelo fato de aumentar sua singularidade de 0,6% para 3,4%, enquanto Psiquiatria/Psicologia aumenta 6 posições e a singularidade de 3,0% para 20,3%. E no sentido contrário, algumas áreas aumentam 4 posições quando WoS é considerada exclusivamente: Economia &

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Negócios, Ciências dos Materiais e Física. E no período posterior à expansão a diferença do ranqueamento da singularidade das áreas é bem menos pronunciada.

Figura 3 – Singularidade das bases WoS+SciELO e WoS em relação à Scopus, nas diversas áreas



Fonte: Elaborado pelos autores

A Figura 4 destaca que a singularidade da base Scopus é mais pronunciada – principalmente no período anterior à expansão da WoS, como já se pôde observar na Figura 2. Outro aspecto que chama a atenção é que a SciELO faz real diferença complementando a WoS, pois a singularidade da Scopus em relação à WoS é significativamente maior no gráfico inferior da Figura 4.

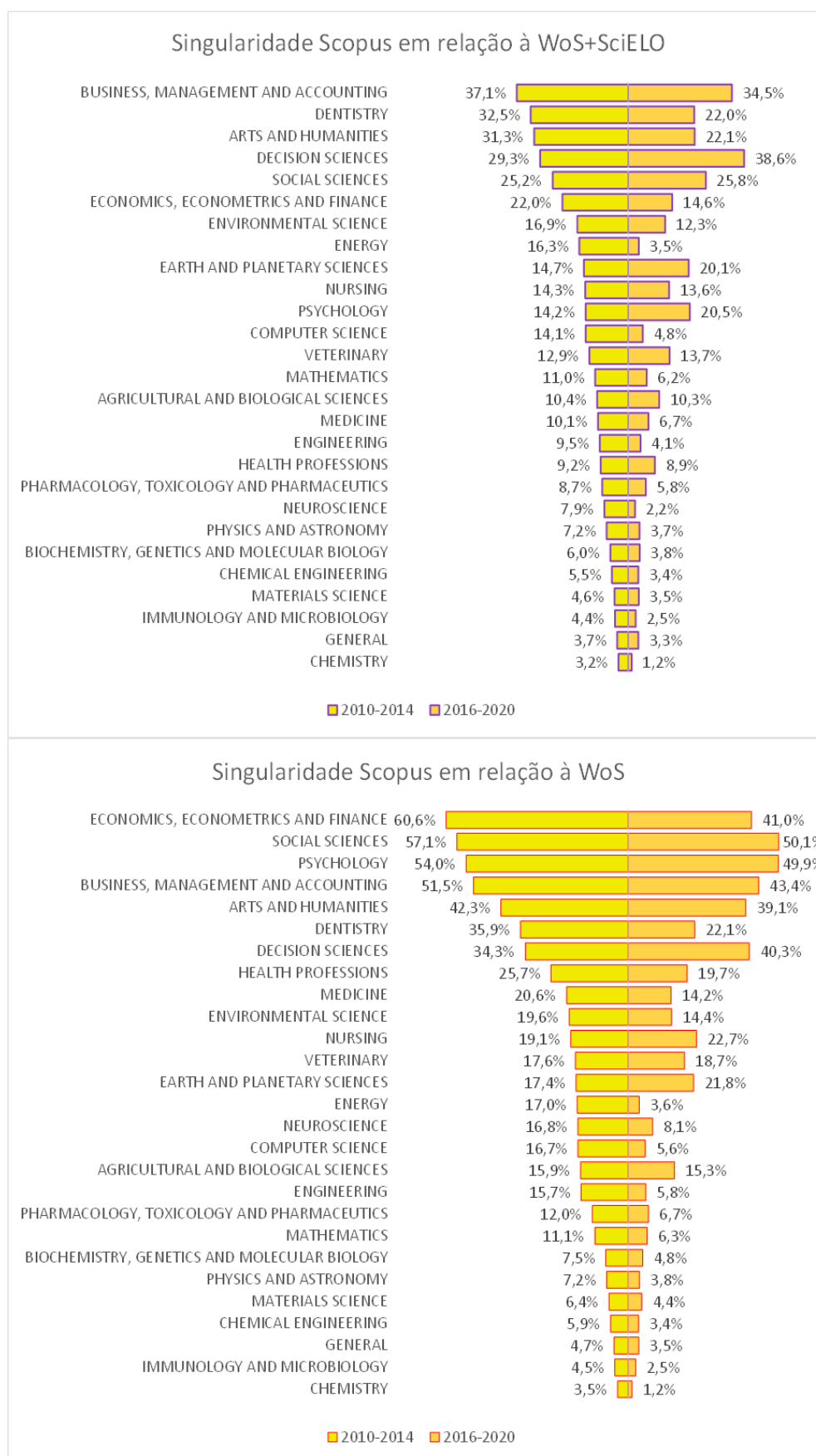
É interessante observar na Figura 4 a área de Economia, Econometria e Finança, que evidencia a maior perda de singularidade da Scopus após expansão da WoS, quando considerada a WoS exclusivamente. Este efeito resulta de um aumento significativo na área de Economia e Negócios, conforme observado na Figura 3. Apesar de diminuir, conclui-se que os altos percentuais de singularidade desta área em cada uma das bases sugerem que estudos deveriam considera-las em conjunto, haja vista sua clara complementaridade (principalmente após 2015). O mesmo pode-se dizer das áreas de Ciências Sociais e Humanas e subárea de Negócios, que na base Scopus, correspondem respectivamente às categorias: Ciências Sociais, Artes e Humanidades e Negócios, Administração e Contabilidade.

Algumas áreas se destacaram por um aumento mais pronunciado de singularidade na base Scopus após 2015, sugerindo que a expansão da WoS não deve ter beneficiado essas áreas. Ciências da Decisão aumenta 9,3% e 6,0%, respectivamente, quando considerada a SciELO ou não; enquanto para Psicologia esses valores são, respectivamente, de 6,3% e 4,0%; para Ciências da Terra e Planetárias de 5,4% e 4,4%.

A alteração de ranqueamento de singularidade entre as áreas nos gráficos da Figura 4 antes da expansão da WoS também merecem atenção. Profissões da Saúde aumenta 10 posições no ranqueamento quando WoS é considerada exclusivamente, graças a um aumento de 9,2% para 25,7. Contrariamente, Energia e Matemática sobem 6 posições quando considerada a base SciELO. E depois da expansão Neurociências sobe 11 posições quando considerada a WoS exclusivamente, enquanto Ciências da Decisão sobe 4 posições ao se considerar a SciELO.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Figura 4 – Singularidade da base Scopus em relação às bases WoS+SciELO e WoS, nas diversas áreas



Fonte: Elaborado pelos autores

5 CONSIDERAÇÕES FINAIS

O estudo buscou oferecer um panorama comparativo da cobertura das bases SciELO, WoS e Scopus em relação à produção científica brasileira no intuito de facilitar a escolha das bases em análises nas diversas áreas e períodos. A observação dos efeitos da expansão da literatura na WoS Core Collection em 2015, após criação o Emerging Sources Citation Index, sinaliza a importância de monitoramentos que permitam decidir as fontes de informação para delineamento de análises bibliométricas e revisão de literatura. Considera-se que os objetivos da pesquisa foram atingidos, oferecendo um maior detalhamento na análise de cobertura, considerando as diferenças nas áreas e variações temporais.

Dentre os achados da pesquisa, pode-se destacar que a consideração das bases WoS e SciELO combinadas segue sendo importante para garantir maior singularidade em relação à Scopus. Além disso, ficou evidente que a expansão da WoS em 2015 logrou um nível de singularidade comparável à Scopus, sendo que esta tinha vantagem no período anterior.

Por outro lado, a especificidade das áreas mostrou que a sobreposição dessas bases é mais alta nas ciências duras, e que ambas apresentam maior singularidade nas ciências brandas, o que sugere que estudos nestas últimas mereçam uma avaliação sobre o uso das bases combinadas, o que fica mais acentuado após a expansão da WoS.

Pesquisas futuras devem aprofundar o estudo da cobertura das bases considerando os periódicos exclusivos de cada base, assim como é importante compreender se o aumento de singularidade se dá por um aumento de produção científica em periódicos do Brasil ou do exterior que passaram a ser indexados em determinada base. A nacionalidade de edição do periódico é importante num contexto em que a política científica nacional estimula com grande ênfase a publicação internacional.

Uma limitação da pesquisa diz respeito ao uso exclusivo da produção brasileira nas bases de dados com alta seletividade. Seus vieses de seleção resultam na exclusão de parte significativa da ciência brasileira. Outra limitação reside no fato da comparação ter sido feita utilizando ISSN e ano, o que se sabe que utilizando cada artigo conferiria maior precisão, o que pode ser considerado para continuidade da pesquisa.

AGRADECIMENTOS

Este trabalho utilizou dados da base Scopus fornecidos pela Elsevier através do ICSR Lab.

Agradecemos ao CNPq, pelo financiamento da pesquisa (Bolsa Produtividade #308800/2022-2 de Rogério Mugnaini).

REFERÊNCIAS

ASUBIARO, Toluwase; ONAOLAPO, Sodiq; MILLS, David. Regional disparities in Web of Science and Scopus journal coverage. **Scientometrics**, [S. l.], p. 1-23, 2024. DOI: <https://doi.org/10.1007/s11192-024-04948-x>

COLLAZO-REYES, Francisco. Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. **Scientometrics**, [S. l.], v. 89, n. 1, p. 197-209, 2014. DOI: <https://doi.org/10.1007/s11192-013-1036-2>

COSTAS, Rodrigo; MORENO, Luz.; BORDONS, María. Solapamiento y singularidad de MEDLINE, WoS e IME para el análisis de la actividad científica de una región en Ciencias de la Salud. **Revista Española De Documentación Científica**, Madrid, v. 31, n. 3, p. 327-343, 2008. DOI: <https://doi.org/10.3989/redc.2008.v31.i3.432>

COX, Andrew; GADD, Elizabeth; PETERSOHN, Sabrina; SBAFFI, Laura et al. Competencies for bibliometrics. **Journal of Librarianship and Information Science**, [S. l.], v. 51, n. 3, p. 746-762, 2019. DOI: <https://doi.org/10.1177/0961000617728111>

CULBERT, Jack; HOBERT, Anne; JAHN, Najko; HAUPKA, Nick; SCHMIDT, Marion; DONNER, Paul; MAYR, Philipp. Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus. **arXiv preprint**, [S. l.], v. 1, n. 1, 2024. <https://arxiv.org/abs/2401.16359>

DELGADO-QUIRÓS, Lorena; ORTEGA, José Luis. Completeness degree of publication metadata in eight free-access scholarly databases. **Quantitative Science Studies**, [S. l.], v. 5, n. 1, p. 31-49, 2024. DOI: https://doi.org/10.1162/qss_a_00286

GAILLARD, Jacques. ¿Es visible la ciencia del tercer mundo? **Mundo Científico**, London, v. 9, n. 93, p. 764-768, 1989. Disponível em: https://www.researchgate.net/publication/32982226_Es_visible_la_ciencia_del_tercer_mundo. Acesso em: 10 jul. 2024.

GIBBS, Wayt W. Lost science in the third world. **Scientific American**, New York, v. 273, n. 2, p. 92-99, 1995. DOI: <https://doi.org/10.1038/scientificamerican0895-92>

GLÄNZEL, Wolfgang. **Bibliometrics as a research field**: a course on theory and application of bibliometric indicators. Belgica: Course Handouts, 2003. Disponível em: https://www.researchgate.net/publication/242406991_Bibliometrics_as_a_research_field_A_course_on_theory_and_application_of_bibliometric_indicators. Acesso em: 10 jul. 2024.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

GLUCK, Myke. A review of journal coverage overlap with an extension to the definition of overlap. **Journal of the American Society for Information Science**, [S. l.], v. 41, n. 1, p. 43-60, 1990. [https://doi.org/10.1002/\(SICI\)1097-4571\(199001\)41:1%3C43::AID-ASI4%3E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-4571(199001)41:1%3C43::AID-ASI4%3E3.0.CO;2-P)

GUSENBAUER, Michael. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. **Scientometrics**, [S. l.], v. 118, n. 1, p. 177-214, 2019. DOI: <https://doi.org/10.1007/s11192-018-2958-5>

LETA, Jacqueline. Brazilian growth in the mainstream science: The role of human resources and national journals. **Journal of Scientometric Research**, New Delhi, v. 1, n. 1, p. 44-52, 2012. DOI: <https://doi.org/10.5530/jscires.2012.1.9>

MARTÍN-MARTÍN, Alberto; ORDUNA-MALEA, Enrique; THELWALL, Mike; LÓPEZ-CÓZAR, Emilio Delgado. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. **Journal of informetrics**, [S. l.], v. 12, n. 4, p. 1160-1177, 2018. DOI: <https://doi.org/10.1016/j.joi.2018.09.002>

MUGNAINI, Rogério; DIGIAMPETRI, Luciano Antonio; MENA-CHALCO, Jesús Pascual. Comunicação científica no Brasil (1998-2012): indexação, crescimento, fluxo e dispersão. **TransInformação**, Campinas, v. 26, n. 3, p. 239-252, 2014. DOI: <http://dx.doi.org/10.1590/0103-37862014000300002>

MUGNAINI, Rogério; DAMACENO, Rafael Jeferson Pezzuto; DIGIAMPIETRI, Luciano Antonio; MENA-CHALCO, Jesús Pascual. *Panorama da produção científica do Brasil além da indexação: uma análise exploratória da comunicação em periódicos*. **TransInformação**, Campinas, v. 31, 2019. DOI: <https://doi.org/10.1590/2318-0889201931e190033>

MUGNAINI, Rogério; IGAMI, Mery Piedad Zamudio; KRZYZANOWSKI, Rosaly Favero. Acesso aberto e financiamento da pesquisa no Brasil: características e tendências da produção científica. **Encontros Bibli**, Florianópolis, v. 27, p. 1-26, 2022. DOI: <https://doi.org/10.5007/1518-2924.2022.e78818>

NEUHAUS, Christoph; DANIEL, Hans-Dieter. Data sources for performing citation analysis: an overview. **Journal of documentation**, [S. l.], v. 64, n. 2, p. 193-210, 2008. DOI: <https://doi.org/10.1108/00220410810858010>

PACKER, Abel Laerte. SciELO: uma metodologia para publicação eletrônica. **Ciência da Informação**, Brasília, DF, v. 27, n. 2, p. 109-121, 1998. DOI: <https://doi.org/10.1590/S0100-19651998000200001>

PEREIRA, Fabiana Andrade; MUGNAINI, Rogério. Mapping the use of Google Scholar in evaluative bibliometric or scientometric studies: A bibliometric review. **Quantitative Science Studies**, [S. l.], v. 4, n. 1, p. 233-245, 2023. DOI: https://doi.org/10.1162/qss_a_00231

SINGH, Vivek Kumar; SINGH, Prashasti; KARMAKAR, Mousumi; LETA, Jacqueline; MAYR, Philipp. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. **Scientometrics**, [S. l.], v. 126, p. 5113-5142, 2021. DOI: <https://doi.org/10.1007/s11192-021-03948-5>

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

SIVERTSEN, Gunnar. Developing current research information systems (CRIS) as data sources for studies of research. *In*: GLÄNZEL, Wolfgang; MOED, Henk F.; SCHMOCH, Ulrich; THELWALL, Mike. (ed.). **Springer handbook of science and technology indicators**. Dordrecht: Springer, 2019. p. 667-683. DOI: https://doi.org/10.1007/978-3-030-02511-3_25

TENOPIR, Carol. Evaluation of database coverage: A comparison of two methodologies. **Online Review**, [S. l.], v. 6, n. 5, p. 423-441, 1982. DOI: <https://doi.org/10.1108/eb024110>

TESTA, James. **The globalization of Web of Science**. São Paulo: Thomson Reuters Expert Essays. 2011. Disponível em: <https://wokinfo.com/media/pdf/globalwos-essay>. Acesso em: 10 jul. 2024.