

CONJUNTO DE DADOS: ÍNDICES CIENTOMÉTRICOS DE PERIÓDICOS CIENTÍFICOS COMPARADOS AO ESTRATO QUALIS

Dataset: Scientometric Indexes of Scientific Journals compared to the QUALIS stratum

Josir Cardoso Gomes

Instituto Brasileiro de Informação em Ciência e Tecnologia / CGTI, Brasília, DF
Fundação Getulio Vargas / ECMI, Rio de Janeiro, RJ
josircg@gmail.com

<https://orcid.org/0000-0001-7629-1404> 

Fábio Castro Gouveia

Instituto Brasileiro de Informação em Ciência e Tecnologia / COEPI, Brasília, DF
Fundação Oswaldo Cruz
fgouveia@gmail.com

<https://orcid.org/0000-0002-0082-2392> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: Descreve a confecção e publicação do conjunto de dados cientométricos de periódicos nacionais e internacionais consolidados a partir do Qualis, Scopus, SJC, Diamond Journals e Google Scholar. Tem como objetivo dar suporte a pesquisadores brasileiros que queiram analisar e comparar diversos atributos cientométricos e, também, o modelo econômico utilizado pelos periódicos, mesmo que estes não estejam ainda no Qualis. Além do conjunto de dados, o artigo também descreve a implementação de uma consulta na web para que qualquer pesquisador possa realizar consultas rápidas sem a necessidade de baixar o conjunto de dados.

Método: Esta pesquisa caracteriza-se como descritiva e aplicada por envolver um problema prático que é a coleta dos dados de periódicos em diversas fontes. Para tanto, os dados foram agregados em um banco de dados relacional e exportados em um arquivo aberto no formato CSV.

Potencial de reutilização: A disponibilização dos conjuntos de dados dos periódicos científicos agregados permite que outros pesquisadores possam realizar consultas e análises bibliométricas a partir de uma única fonte.

PALAVRAS-CHAVE: Bibliometria, Periódicos Científicos, Citescore, API

ABSTRACT

Objective: It describes the creation and publication of the scientometric dataset of national and international journals consolidated from Qualis, Scopus, SJC, Diamond Journals and Google Scholar. The creation of the dataset aims to support Brazilian researchers who want to analyze and compare various scientometric attributes and also the economic model used by journals, even if they are not yet in Qualis. In addition to the dataset, the article also describes the implementation of a web query so that any researcher can carry out quick queries without the need to download the dataset.

Method: This research is characterized as descriptive and applied because it involves a practical problem: the collection of journal data from various sources. To this end, the data was aggregated in a relational database and exported in an open file in CSV format.

Reuse potential: The availability of aggregated scientific journal datasets allows other researchers to carry out bibliometric queries and analyses from a single source.

KEYWORDS: Bibliometrics, Scientific Journals, Citescore, API

1. APRESENTAÇÃO

Uma das tarefas corriqueiras de qualquer pesquisador é o planejamento da estratégia de publicação em periódicos científicos. A seleção de qual periódico publicar não é uma tarefa fácil pois o número de variáveis no processo decisório é muito grande. Tal escolha depende se o assunto de sua pesquisa tem aderência aos diversos “Call for Papers” (CFP) que surgem durante o ano, se existe custo para publicar e, especificamente para a realidade brasileira, se o periódico já faz parte do Qualis. Afinal, quem nunca ouviu a pergunta “mas qual é o Qualis da revista?” ao receber uma sugestão para publicar em um determinado periódico?

O chamado Qualis é um jargão utilizado no meio acadêmico brasileiro que resume a tabela Qualis Periódicos fornecida pela CAPES que, a cada quadriênio, determina uma classificação (ou índice de “qualidade”) para periódicos científicos. Na última avaliação quadrienal, a CAPES definiu 9 classificações: A1, A2, A3, A4, B1, B2, B3, B4 e C, com A1 sendo a classificação mais elevada que um periódico pode receber, e C, sendo a pior pontuação (CAPES, 2023a).

Apesar da própria instituição ressaltar em relatório técnico que “não recomenda e não se responsabiliza pela utilização do Qualis como uma fonte de classificação da qualidade dos periódicos científicos para outros fins que não a avaliação dos programas de pós-graduação stricto sensu” (CAPES, 2023b, p.1) efetivamente esse é um dos principais parâmetros de decisão por parte dos pesquisadores vinculados à pós-graduações no Brasil.

Isso ocorre porque o resultado desta estratificação é utilizado no sistema Sucupira e influi em grande medida na avaliação dos programas de pós-graduação (PPGs) e, por consequência, determinam o status dos PPGs e dos recursos financeiros que cada programa terá acesso.

Assim, este formato de classificação tem gerado polêmicas e conflitos no meio acadêmico, sendo que tais conflitos chegaram a ser tão graves que em 2021, no auge do desmonte de políticas públicas educacionais, durante o governo Bolsonaro, o Ministério Público do Rio de Janeiro foi acionado para mediar o caso. Tal ação resultou em um termo de adequação de conduta que deu uma maior transparência ao processo e atendeu a diversas reivindicações de pesquisadores e programas que se sentiram lesados (Palhares, 2021).

Um dos pontos centrais de polêmica do Qualis é a temporalidade da classificação, pois o estrato pode ser alterado no final do quadriênio e uma queda na classificação pode prejudicar a estratégia que um PPG utilizou para escolher em quais periódicos publicar. Em contrapartida, a metodologia necessita utilizar critérios referentes ao quadriênio que se tenta avaliar e, a princípio, estes critérios só poderão ser calculados após o fim dele.

Um outro ponto a considerar é o critério para a escolha de quais periódicos devem entrar na seleção na definição do Qualis. O critério atual contempla os periódicos onde os pesquisadores vinculados aos PPGs publicaram no quadriênio em análise. Este levantamento se dá a partir da base Sucupira e como o modelo atual faz uso de métricas cujos estratos são definidos pelo corpus total de periódicos analisados, ele reflete as escolhas feitas pelos pesquisadores dos PPGs dentre os periódicos existentes. Assim, apesar de mais de 60 mil periódicos estarem indexados nas grandes bases de dados (Web of Science (WoS) e Scopus), apenas 28.417 fazem parte do Qualis Periódicos de 2017-2020.

Em função desta temporalidade, uma estratégia de publicação deveria tentar prever quais serão as publicações que deverão se manter com uma boa classificação. E uma das formas de prever qual será o estrato de um determinado periódico é o acompanhamento dos índices cientométricos estrangeiros que são utilizados no cálculo dos estratos. Mesmo que estes sejam influenciados pelo corpus do quadriênio anterior, ainda sim é uma medida que pode auxiliar nesta previsão.

A partir destas constatações, buscou-se construir um banco de dados que agregasse diversos índices bibliométricos de um conjunto de periódicos maior que o disponibilizado pelo Qualis e pela Scopus para que um pesquisador ou PPG pudesse, de forma rápida, comparar os principais índices existentes e, a partir desta base, fosse possível cruzar tais informações através de análises estatísticas e cientométricas.

Buscou-se também incluir o índice-H5 do Google Scholar ao banco de dados, pois o grupo de trabalho do Qualis passou a incluir também tal índice como critério na montagem dos estratos. Tal escolha parece ter sido acertada, pois os índices da Scopus e WoS não têm a mesma amplitude que o Google Scholar para periódicos do Sul Global (Canto *et al.*, 2022). No entanto, é importante ressaltar que o índice-H5 do Google em seu cálculo não leva em consideração o número de documentos indexados como divisor, algo que faz parte do *Journal Impact Factor* da WoS e o *CiteScore* da Scopus, configurando assim um indicador com uma dinâmica distinta desses dois indicadores.

É fato que já existem algumas iniciativas que buscam agregar índices bibliométricos. Um dos mais utilizados é o *Publish or Perish* da empresa Harzing, que apesar de ser gratuito não tem o software aberto, ou seja, não é possível a auditoria de como a coleta é realizada e quais os critérios que o software utiliza para fornecer os índices apresentados.

Outras iniciativas também muito conhecidas são o *SCImago Journal Rank* (SJR) editado pela Universidade de Granada na Espanha e o MIAR da Universidade de Barcelona. A deficiência do SJR é que ele se baseia exclusivamente na listagem da Scopus e desta forma deixa de incluir vários periódicos. Já a MIAR procura ser bem mais inclusiva, pois se utiliza de várias outras fontes de dados mas, infelizmente, não disponibiliza a base completa tal com a proposta aqui detalhada. Vale notar que outros trabalhos também relatam a construção de bases semelhantes que buscam comparar os diversos índices bibliométricos (Martín-Martín *et al.*, 2021; Rautenberg; Souza; Hild, 2021).

Em resumo, o objetivo da pesquisa foi construir um conjunto de dados que dê subsídios para que coordenadores de PPGs e pesquisadores no campo da Cientometria possam avaliar, filtrar ou agregar o conjunto integral de periódicos científicos **sem depender de softwares proprietários**. A partir dos dados aqui disponibilizados, pode-se realizar cruzamentos entre os diversos índices cientométricos e responder perguntas do tipo: quais periódicos que contém H5 elevado e não estão presentes na base Scopus? Ou, quais os periódicos que existem em uma determinada área que tem H5 elevado e que não estão presentes no Qualis?

Por exemplo, utilizando o conjunto de dados descrito aqui, foi possível filtrar apenas os periódicos específicos da área de Comunicação Social e descobriu-se que o Brasil é o terceiro país que mais tem periódicos nesta área. Se a mesma consulta fosse realizada apenas com periódicos presentes na Scopus ou na base Qualis Periódicos, este resultado não seria obtido, pois ambas contêm menos registros que o conjunto apresentado aqui.

2. MÉTODOS E INSTRUMENTOS

O passo inicial foi a construção de um banco de dados em MySQL que pudesse armazenar os atributos necessários de cada periódico bem como o detalhamento da

origem do índice e, também, as áreas de conhecimento para que fosse possível realizar filtros para as áreas de interesse dos usuários que venham a realizar consultas nesta base. Após a criação da base, iniciou-se a importação de cada uma das fontes de dados através de scripts em Python.

A importação de cada um dos arquivos em *Comma-separated values* (CSV) contou também com a validação do International Standard Serial Number (ISSN) no portal do ISSN. Caso o ISSN não existisse, o registro não seria gravado e um arquivo de ocorrências de erro era registrado. Esta verificação é necessária, pois na planilha fornecida pela CAPES, existem diversos ISSNs inválidos.

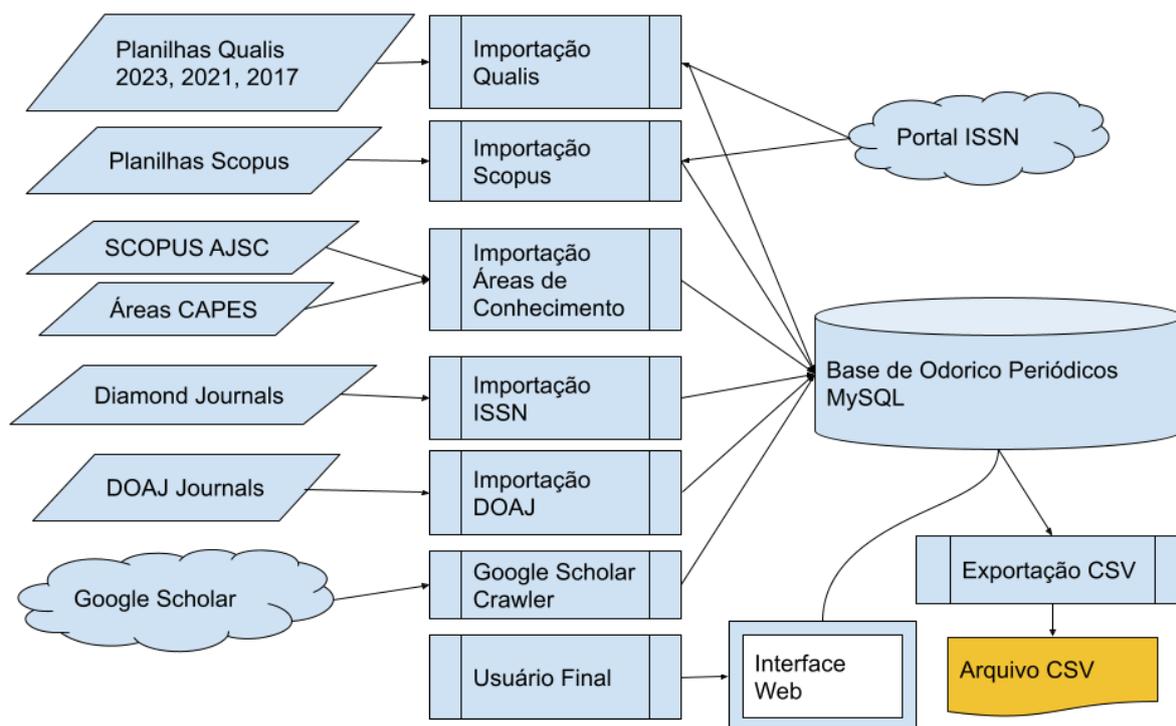
Uma vez que todas as bases em CSV foram importadas, iniciou-se a importação do H5 do Google através do script *Google Scholar Crawler* (Gomes; Gouveia; Pimenta, 2018). Como o Google Scholar não armazena o ISSN das publicações, a estratégia para encontrar o H5 do periódico foi buscar através do nome completo do periódico ou partes do nome, caso o nome completo não fosse encontrado e, por conta disso, são admitidas algumas falhas no processamento pelo *crawler*.

A tabela de *Diamond Journals* (Bosman; Frantsvåg; Kramer, 2021) foi utilizada especificamente para deduzir qual o modelo econômico de cada periódico, ou seja, caso o ISSN esteja presente nesta tabela, sabe-se que o modelo é aberto e sem cobranças de taxas para publicação (*Article Processing Charges* ou APCs).

Optou-se por não utilizar a WoS, pois até o fechamento da pesquisa não foi encontrado o detalhamento da política de dados que permite ou não a divulgação dos índices bibliométricos gerados pela empresa Clarivate. Além disso, o número de periódicos disponibilizados é bem menor do que o encontrado na Scopus. Na verdade, a partir do processamento realizado neste trabalho, pode-se verificar que apenas 6 periódicos indexados nas bases da Clarivate não estavam disponíveis no conjunto das outras bases já importadas.

Após o final de todos os processamentos de importação, um novo *script* foi executado para gerar a tabela CSV detalhada neste artigo, com os indicadores agregados para todos os periódicos coletados durante as cargas anteriores. O arquivo resultante, contém 62.855 periódicos ativos. O processamento completo pode ser visualizado na Figura 1.

Figura 1: Processamento Completo



Fonte: próprios autores

Também foram disponibilizados dois serviços na web para facilitar o acesso aos dados, caso o pesquisador não queira fazer o download do arquivo. O primeiro permite que o usuário faça consultas pelo nome do periódico, pelo ISSN, pela área de conhecimento ou pelo modelo econômico do periódico.

Ao clicar em nome do periódico, o sistema já encaminha o usuário para o site do periódico e se clicar sobre o índice H5 ou Scopus, o usuário é direcionado para o Google Scholar ou para a página da Scopus equivalente. Um exemplo da tela pode ser visualizado na Figura 2.

Já o segundo serviço é uma API¹ que permite que programaticamente se obtenha os dados de um determinado através da chamada https://odorico.irdx.com.br/get_issn/0961-0006. O resultado da chamada da API é um arquivo JSON no formato UTF-8 como o indicado no exemplo abaixo:

```
{"nome": "JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE",  
"issn": "0961-0006", "eissn": null, "url": "https://journals.sagepub.com/home/lis",
```

¹ API é a sigla para *Application Programming Interface* utilizada para representar um serviço que disponibiliza uma comunicação entre programas de computador.

```
"status": "A", "qualis": "A1", "pais": "England", "sistema": "-", "referencia": null,
"dt_validacao": "2021-04-17", "assuntos": [{"cod_externo": "3309",
"descricao": "Library and Information Sciences"},
{ "cod_externo": null,
"descricao": "COMUNICA\u00c7\u00c3O E INFORMA\u00c7\u00c3O"}]}
```

Figura 2: Tela de Consulta de Peri\u00f3dicos

Listagem de Peri\u00f3dicos

ISSN: Q \u00c1rea CNPQ: CI\u00cancia da Informa\u00e7\u00e3o \u00c1rea ASJC / CrossRef: Todos Nome do Peri\u00f3dico: Q

Diamond (Acesso Livre)

NOME COMPLETO	ISSN	QUALIS	E-ISSN	MODELO	GOOGLE HS	CITESCORE	SJR
INFORMATION SCIENCES	0020-0255	A1	-		102,0	11,300	1,7230
JOURNAL OF DOCUMENTATION	0022-0418	A1	-		27,0	2,500	0,8900
LIBRARY QUARTELY	0024-2519	A1	1549-652X		18,0	2,900	0,7270
ACERVO: REVISTA DO ARQUIVO NACIONAL	0102-700X	A1	2237-8723		7,0	-	-
TRANSFORMA\u00c7\u00c3O	0103-3786	A1	2318-0889		12,0	0,700	0,1960
JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE	0961-0006	A1	-		21,0	2,600	0,7110
Journal of Knowledge Management	1367-3270	A1	-		54,0	8,500	1,7520
INFORMATION RESEARCH	1368-1613	A1	-		21,0	1,700	0,4540
INFORMATION, COMMUNICATION & SOCIETY	1369-118X	A1	1468-4462		65,0	8,000	2,5790
EL PROFESIONAL DE LA INFORMACION	1386-6710	A1	1699-2407		36,0	2,100	0,4800
INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES (PRINT)	1432-5012	A1	-		-	5,400	0,5590
INFORMATION AND ORGANIZATION	1471-7727	A1	-		-	6,000	1,8320

Fonte: Pr\u00f3prios Autores

3. TABELA DE ESPECIFICA\u00c7\u00d5ES

\u00c1rea de Conhecimento	BB. M\u00e9todos bibliom\u00e9tricos.
\u00c1rea de assunto espec\u00edfica	Bibliometria
Idioma	Multil\u00edngue
Tipo de Arquivo	Tabela no formato CSV
Localiza\u00e7\u00e3o da fonte de dados	https://doi.org/10.5281/zenodo.10202106
Como os dados foram adquiridos	Os dados brutos utilizados na montagem do arquivo agregado podem ser baixados nos links abaixo: QUALIS (2023): https://sucupira.capes.gov.br/sucupira/public/consultas/col

	<p>eta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf</p> <p>Deve-se desmarcar todas as opções para que o arquivo completo fique disponível</p> <p>Scopus (2022): https://www.scopus.com/sources Deve-se clicar no botão <i>Download Source</i></p> <p>SJR (2022): https://www.scimagojr.com/journalrank.php Deve-se clicar no botão <i>Download Source</i></p> <p>DOAJ (2023): https://doaj.org/docs/public-data-dump/ Deve-se clicar no link “exportable version of the journal metadata”</p>
Estado dos dados	Agregados
Parâmetros para coleta de dados	Sem necessidade de parâmetros
Artigo de pesquisa relacionado	<p>GOMES, J. C.; GOUVEIA, F. C.; PIMENTA, R. Google scholar como fonte de medição da produção científica lusófona. Anais do I Congresso de Altimetria e Ciência Aberta na América Latina. Anais... Em: I CONGRESSO DE ALTMETRIA E CIÊNCIA ABERTA NA AMÉRICA LATINA. Niterói, RJ: 2018. Disponível em: https://app.uff.br/riuff/handle/1/7928</p>

3.1. Descrição do conjunto de dados

O conjunto de dados é disponibilizado em um único arquivo CSV no formato UTF-8 separado por ponto-e-vírgula e tendo os elementos textuais encapsulados por aspas. Cada linha representa um periódico contendo as seguintes colunas:

ISSN: Código ISSN Principal do periódico;

eISSN: Código ISSN específico da publicação online;

Título: Título completo sem abreviações ou extensões do tipo (ONLINE) ou (PRINT);

ASJC: Descrição da Área utilizado pela Scopus. Caso exista mais de um, a descrição será separada por vírgula (pode conter nulos);

Qualis: Estrato Qualis referente publicado em 2023 (Quadriênio 2017-2020) - Além dos 8 estratos, pode conter ND indicando que o Qualis não foi calculado;

Google H5: Índice H5 referente a 2020 calculado pelo Google (pode conter nulos)

Google H5M: Índice H5 Mediana referente a 2020 calculado pelo Google (pode conter nulos)

Citescore: Índice *Citescore* da Scopus (pode conter nulos)

SJR: Índice SJR (pode conter nulos)

URL: Endereço do periódico na web (pode conter nulos)

País: Nome do País (em inglês) da Editora do Periódico (pode conter nulos)

Modelo Econômico: pode assumir 3 valores: (D) Diamond, (A) APC e (T) Trancado. Onde “Diamond” indica periódicos de acesso aberto sem nenhum tipo de cobrança ou restrição de acesso. “APC” (*Article Processing Charges*) indica que o periódico cobra o autor e “Trancado” indica que o acesso ao conteúdo é restrito sendo disponibilizado apenas a quem paga por ele.

REFERÊNCIAS

BOSMAN, J.; FRANTSVÅG, J. E.; KRAMER, B.. **OA Diamond Journals Study**. Dataset. Zenodo, 9 mar. 2021. Disponível em: <https://zenodo.org/record/4553103>. Acesso em: 18 jun. 2023

CAPES. **CAPES publica metodologia do Qualis**. Brasília, 2023a. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/capes-publica-metodologia-do-qualis>. Acesso em: 18 jun. 2023.

CAPES. **Relatório Final do Grupo de Trabalho Qualis Periódicos**. Brasília, 2023b. Disponível em: <https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/avaliacao/relatorios-tecnicos-e-grupos-de-trabalho>. Acesso em: 18 jun. 2023

CANTO, F. L. *et al.* Latin American and Caribbean journals indexed in Google Scholar Metrics. **Scientometrics**, [S. l.], v. 127, n. 2, p. 763–783, fev. 2022. Disponível em: <https://link.springer.com/article/10.1007/s11192-021-04237-x>. Acesso em: 18 jun. 2023.

GOMES, J. C.; GOUVEIA, F. C.; PIMENTA, R. M. Google Scholar como fonte de medição da produção científica lusófona. In: CONGRESSO DE ALTMETRIA E CIÊNCIA ABERTA NA AMÉRICA LATINA (LATMETRICS), 1., 2018, Niterói. **Anais [...]**. Rio de Janeiro: UFF, 2018.

MARTÍN-MARTÍN, A. *et al.* Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. **Scientometrics**, [S. l.], v. 126, n. 1, p. 871–906, 1 jan. 2021. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03690-4>. Acesso em: 18 jun. 2023.

PALHARES, I. Justiça suspende avaliação de pós-graduação por suposto uso de critério ilícito. **Folha de São Paulo**, 23 set., 2021. Disponível em:

<https://www1.folha.uol.com.br/educacao/2021/09/justica-suspende-avaliacao-de-pos-graduacao-por-suposto-uso-de-criterio-ilicito.shtml>. Acesso em: 18 jun. 2023.

RAUTENBERG, S.; SOUZA, L.; HILD, T. A. Conjuntos de dados para pesquisas cientométricas: recuperando QUALIS, SJR, SNIP a partir da Web de Dados. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 26, 6 ago. 2021. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/80155>. Acesso em: 18 jun .2023.

NOTAS

AGRADECIMENTOS

Agradecimentos: ao Dr. Washington Segundo (IBICT) que incentivou o desenvolvimento da pesquisa com sugestões no âmbito do projeto BrCRIS. E principalmente ao IBICT e ao Laboratório em Rede de Humanidades Digitais (Larhud) por fornecer suporte tecnológico e financeiro para o desenvolvimento desta pesquisa. Este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processos 315521/2020-1 e 315689/2023-4.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: J. C. Gomes

Coleta de dados: J. C. Gomes

Análise de dados: J. C. Gomes, F. C. Gouveia

Discussão dos resultados: J. C. Gomes, F. C. Gouveia

Revisão e aprovação: J. C. Gomes, F. C. Gouveia

FINANCIAMENTO

Capes 001

Cnpq 315521/2020-1 e 315689/2023-4

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

não se aplica.

CONFLITO DE INTERESSES

Não se aplica.

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva.

HISTÓRICO

Recebido em:15-03-2023 – Aprovado em: 19-02-2024 – Publicado em: 15-04-2024

