

Correspondência dos autores

¹ Universidade de Marília
Marília, SP - Brasil
r.castanha@gmail.com

² Universidade Estadual Paulista
Rio Claro, SP - Brasil
franfranco.santos@gmail.com

BRAPCI Explorer: um novo ambiente web para análises bibliométricas a partir da Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação

Rafael Gutierrez Castanha¹ Francielle Franco dos Santos²

RESUMO

Introdução: No campo da Ciência da Informação, a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) é uma referência nacional. No entanto, em comparação com outras bases, a BRAPCI oferece poucas funcionalidades bibliométricas. **Objetivo:** propõe um novo ambiente web, denominado BRAPCI Explorer, voltado a análises bibliométricas, a partir da requisição de dados via API da base de dados da BRAPCI, ampliando e estruturando a visualização dos dados indexados. **Metodologia:** Um código em linguagem R foi desenvolvido para realizar extração de dados da BRAPCI via API, organizando e estruturando os dados em um site interativo. Esse código gera rede de coautorias, tabelas de frequência de autores e fontes de publicação, gráfico de produção por ano e calcula indicadores bibliométricos, incluindo a quantidade de autores distintos, média de documentos por autor, índice de coautoria, total de fontes diferentes, média de documentos por fonte e por ano. **Resultados:** Ao testar o novo ambiente foi realizada uma busca com o termo "produção científica" encontrando 3099 documentos distintos produzidos por 3494 autores distintos em 51 fontes de publicação distintas. O BRAPCI Explorer foi capaz de gerar a rede de coautorias bem como elaborar todas tabelas e gráfico de frequência, além dos cálculos dos indicadores. **Conclusão:** Conclui-se que o BRAPCI Explorer contribui para uma melhor visualização dos dados disponibilizados pela BRAPCI, isso porque apresenta visualizações aprimoradas, em formato de rede, gráficos e tabelas, além de indicadores bibliométricos, em um mesmo ambiente.

PALAVRAS-CHAVE

Bibliometria. BRAPCI. Application Programming Interface (API), Visualização de dados. Ambiente web.

BRAPCI Explorer: a new web environment for bibliometric analyzes based on the BRAPCI

ABSTRACT

Introduction: In the field of Information Science, the Referential Database of the Database of Journal Articles in Information Science (BRAPCI) is a national reference. However, compared to other databases, BRAPCI provides limited bibliometric metrics. **Objective:** This search proposes a new web environment, named BRAPCI Explorer, aimed at bibliometric analyses. This would be achieved by requesting data through the BRAPCI database API, expanding and structuring the data visualization. **Methodology:** An R programming language code was

developed to extract data from BRAPCI via API, organizing and structuring the data into an interactive website. This code generates a co-authorship network, tables of author and publication source frequency, a production graph per year, and calculates bibliometric indicators, including the quantity of distinct authors, average documents per author, co-authorship index, total different sources, average documents per source, and per year. **Results:** When testing the new environment, a search with the term "scientific production" yielded 3099 distinct documents produced by 3494 distinct authors in 51 distinct publication sources. BRAPCI Explorer was capable of generating the co-authorship network as well as preparing all tables and frequency graphs, in addition to calculating the indicators. **Conclusion:** It is concluded that BRAPCI Explorer contributes to a better visualization of the data provided by BRAPCI because it presents enhanced visualizations in the form of networks, graphs, and tables, along with bibliometric indicators, all in a single environment.

KEYWORDS

Bibliometrics. BRAPCI. Application Programming Interface (API), Data visualization. Web environment.

CRedit

- **Reconhecimentos:** Não aplicável.
- **Financiamento:** Não aplicável.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não aplicável.
- **Disponibilidade de dados e material:** O código está disponível em: <https://doi.org/10.5281/zenodo.8371105> Aplicativo em funcionamento: <https://fctools.shinyapps.io/BRAPCIexplorer>
- **Contribuições dos autores:** Conceitualização, Investigação, Metodologia, Software, Validação, Visualização, Recursos Escrita – rascunho original, Escrita – revisão & edição, Análise formal: CASTANHA, R. G.; FRANCO, F. Curadoria de dados, Administração do projeto, Supervisão: CASTANHA, R. G.

| 2

JITA: BB. Bibliometric methods.



Artigo submetido ao sistema de similaridade

Submetido em: 25/09/2023 – Aceito em: 30/10/2023 – Publicado em: 23/11/2023

Editor: Gilденir Carolina Santos

1 INTRODUÇÃO

Bases de dados científicas podem ser entendidas como sistemas de recuperação da informação que visam armazenar, indexar, representar e disponibilizar informações científicas a partir de demandas de usuários de determinada comunidade, utilizando-se de mecanismos de recuperação da informação a fim de contribuir para a difusão da ciência a partir de fontes de informações científicas (Arruda; Felipe; Santos, 2020).

Um grande marco histórico do desenvolvimento das bases de dados científicas é a criação do *Institute of Scientific Information* (ISI) e do índice de citações (*Science Citation Index - SCI*) a partir dos esforços de Eugene Garfield durante as décadas de 1950 e 1960, que culminaria na criação da base de dados *Web Of Science*, lançada em 1997, ainda em formato de cd-rom após a Thomson Reuters adquirir o ISI em 1992 (Clarivate, 2023). Em 2001, a base incorporou o *Essential Science Indicators*, um conjunto de indicadores bibliométricos relacionados à produção e impacto de produções em nível de autores, fontes de publicação, artigos e países, sendo amplamente difundido e utilizado em pesquisas diversas.

Outra base de dados de enorme destaque no cenário mundial é a *Scopus*, lançada em 2004 (Schotten *et al.*, 2017; Thelwall; Sud, 2022) e pertencente ao grupo Elsevier. Assim como a *Web Of Science*, a *Scopus* é uma base de dados e de citação, ou seja, ambas são capazes de captar as conexões estabelecidas pelos documentos indexados em seus respectivos bancos de dados. Além disso, a *Scopus*, tal qual a *Web Of Science*, são bases multidisciplinares, de larga cobertura e que fornecem indicadores bibliométricos aos seus usuários. Com isso, tornaram-se importantes fontes de dados para pesquisas situadas no cerne dos estudos métricos da informação e da ciência da informação (CI).

Além destas, é possível elencar bases de dados especializadas em determinadas áreas do conhecimento, como a PubMed/Medline (Ciências da Saúde), a MathSciNet (Matemática), IEEE Xplore (Engenharias e Ciência da Computação), EconLit (Economia), ERIC - Education Resources Information Center (Educação) e LISA - Library & Information Science Abstracts (Ciência da Informação). Especificamente com relação à área de Ciência da Informação, destaca-se, em âmbito nacional, a BRAPCI (Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação). A BRAPCI foi desenvolvida na Web e é reconhecida como o repositório mais abrangente da produção científica periódica na área de Ciência da Informação no Brasil (Freitas; Bufrem; Gabriel Junior, 2010). A base de dados é resultado do projeto de pesquisa intitulado “Opções metodológicas em pesquisa: a contribuição da área da informação para a produção de saberes no ensino superior”, de Leilah Santiago Bufrem, que tinha como propósito fornecer apoio a estudos e propostas na área de Ciência da Informação, com base em atividades planejadas institucionalmente (BRAPCI, 2023).

Contudo, a BRAPCI fornece poucas métricas de cunho bibliométrico se comparada com outras bases como *Web of Science* ou *Scopus*, limitando-se aos indicadores de produção (quantidade total de publicações) por autor, fonte e ano de publicação. Assim sendo, dada a relevância da BRAPCI, é plausível afirmar que a inserção de novas funcionalidades bibliométricas na base seria de grande valia e poderia enriquecer significativamente a análise e avaliação da produção científica na área de Ciência da Informação. Isso permitiria aos pesquisadores uma compreensão mais abrangente dos trabalhos publicados, além de facilitar a identificação de tendências de pesquisa em ascensão dentro da área.

Para que isso seja possível, é necessário acessar todos resultados de buscas possíveis da base, assim, esta pesquisa apoia-se na extração de dados via API¹ (*Application Programming*

¹ É necessário esclarecer que, inicialmente, esta pesquisa propôs extração de dados via *web scraping* (raspagem de dados). Entretanto, após o processo de revisão por pares, os revisores sugeriram contato junto aos administradores da BRAPCI para requisitar acesso à API da base, dada a sinalização de atualização da BRAPCI nos próximos meses após esta publicação, podendo tornar o novo ambiente obsoleto em pouco tempo após sua divulgação. A ferramenta que utiliza *web scraping* enquanto método de extração de dados estará disponível enquanto a versão atual da BRAPCI ainda esteja acessível, em: https://fctools.shinyapps.io/BRAPCIexplorer_ws/

Interface - Interface de Programação de Aplicação) para propor um novo ambiente web. Uma Interface de Programa de Aplicação (API) é um mecanismo que possibilita a interação entre diferentes softwares, permitindo que desenvolvedores acessem e utilizem funcionalidades específicas de um sistema, aplicação ou serviço. Utilizando uma API, desenvolvedores podem criar programas que se conectam diretamente à base de dados, selecionando e manipulando informações de forma automatizada proporcionando uma maneira eficaz de obter dados sem depender da navegação convencional (pela interface gráfica), sendo uma ferramenta valiosa para a integração e a automação de processos relacionados a dados, assim, APIs podem ser empregadas para viabilizar o acesso direto ao banco de dados, eliminando a necessidade de interações manuais com interfaces de usuário (Brasil, 2023)

Portanto, este estudo pretende responder ao seguinte problema: de que modo um novo ambiente web, obtido a partir da API da BRAPCI, pode contribuir para melhor visualização, compreensão e análise de diferentes dados bibliométricos? Tendo como objetivo propor um novo ambiente web, denominado BRAPCI Explorer, voltado a análises bibliométricas, a partir da requisição de dados via API da base de dados da BRAPCI, ampliando e estruturando a visualização dos dados indexados. De maneira específica, compreender de que forma, este ambiente pode contribuir para o enriquecimento da base de dados, considerando a importância da integração das análises bibliométricas às bases por meio de um ambiente web de interface intuitiva.

Nesse contexto, a proposição do BRAPCI Explorer se alinha às próprias necessidades da BRAPCI descritas em seu portal ao afirmarem que a BRAPCI está fazendo uma pesquisa online com a finalidade de avaliar, sob o ponto de vista do usuário, a possibilidade de implementar futuras melhorias de interface, conteúdo e o nível de satisfação (BRAPCI, 2023). Além disso, existem outras iniciativas como a BRAPCI Livros de Silva (2023) e o ScraperCI de Graciano (2020) e Graciano e Ramalho (2023). Este último baseado no método de *web scraping*, e ambos, diferente do BRAPCI Explorer, uma vez que o ScraperCI foca no processo de recuperação da informação e o ambiente aqui proposto, em elementos bibliométricos tais como organização e visualização de dados e geração de indicadores.

Assim como a *OpenAlex*, a *Semantic Scholar* ou *Unpaywall* complementam grandes bases de dados, como a *Web of Science* e a *Scopus* (Velez-Estevez *et al.*, 2023), entende-se que aplicações complementares também poderiam ser desenvolvidas a partir da BRAPCI. Outras iniciativas recentes também fazem uso da base para o desenvolvimento de recursos como a BRAPCI Revistas, a Benancib e a BRAPCI Eventos², contudo não foi possível localizar nenhuma aplicação que apresente as funcionalidades disponibilizadas pelo BRAPCI Explorer. O ambiente aqui proposto permite que usuário acesse análises bibliométricas de maneira facilitada, sem a necessidade de exportação, processamento dos dados e utilização de ferramentas complementares.

2 METODOLOGIA

Para propor um novo ambiente web voltado às análises bibliométricas, implementou-se, em linguagem de programação R, um código capaz de: acessar os dados da BRAPCI via API; realizar buscas; organizá-los e estruturá-los em um web site interativo, aprimorando a visualização dos dados indexados na base. A este novo ambiente deu-se o nome de BRAPCI Explorer³. Este código foi capaz de gerar a rede de coautoria entre os autores identificados sobre determinada busca, construir a tabela de frequência da lista de autores e fontes de publicação mais produtivas, de elaborar o gráfico de produção por ano de publicação, além de calcular indicadores bibliométricos referentes a quantidade de autores distintos encontrados, média de

² Recursos disponíveis em: <https://cip.BRAPCI.inf.br/>

³ <https://fctools.shinyapps.io/BRAPCIexplorer/>

documentos por autor, índice de coautoria, total de fontes distintas, média de documentos publicados por fontes e média de documentos por ano. O código é aberto e está disponível em Castanha e Franco (2023).

Contudo, *a priori*, foi solicitado, junto aos administradores da BRAPCI, o acesso à API da base, o qual prontamente foi concedido. Neste momento é imprescindível elucidar que a base de dados está em processo de atualização e nem todas as funcionalidades estavam disponíveis no momento desta publicação. Mesmo com acesso à API da base concedida, antes da extração dos dados foi verificada a existência dos termos de uso da BRAPCI para averiguar possíveis restrições com relação ao uso de sua API. Esta verificação foi realizada acessando os termos de uso da base (<https://cip.BRAPCI.inf.br/about> e <https://BRAPCI.inf.br/index.php/res/about>). Nenhuma informação relacionada aos direitos autorais ou licença de uso foi encontrada. Após tais verificações, prosseguiu-se com a implementação do código. Assim sendo, a concepção do BRAPCI Explorer se desenvolveu em três frentes: i) delineamento e programação de todo processo de extração de dados via requisição por API; ii) estruturação e organização dos dados extraídos; iii) construção da interface do novo ambiente. Todo o processo é apresentado na Figura 1.

O processo de extração de dados se estabeleceu, em sua primeira etapa, a partir do entendimento do arquivo gerado pela API da BRAPCI. A API disponibilizada fornece arquivos do tipo JSON (*JavaScript Object Notation*), extremamente comum à APIs. No caso da BRAPCI, foi necessário identificar todos os objetos e compreender a quais elementos estes objetos fazia referência identificando o total de publicações e os campos de autoria, fonte de publicação e anos de publicação, visto que estes itens seriam utilizados para construção da rede de coautoria, das tabelas e gráficos e para os cálculos dos indicadores bibliométricos.

Após esta identificação foi preciso compreender a composição dos parâmetros presentes na *url* da API da BRAPCI. Este procedimento foi realizado para que fosse possível manipular os elementos que compõem a *url* durante as buscas realizadas na base. É importante mencionar que a compreensão destes parâmetros é de suma importância na medida em que a base de dados fornece diferentes parâmetros em suas buscas e exibe somente 10 resultados por página. Desta forma, afim de viabilizar a extração de todos os dados resultantes das buscas, foi essencial manipular a *url* para que o algoritmo programado fosse capaz de percorrer todas as páginas de maneira correta. A *url* da base é composta da seguinte maneira:

https://cip.BRAPCI.inf.br/api/BRAPCI/search/v1?q=query&di=ano inicial&df=ano final&start=inicio dos resultados

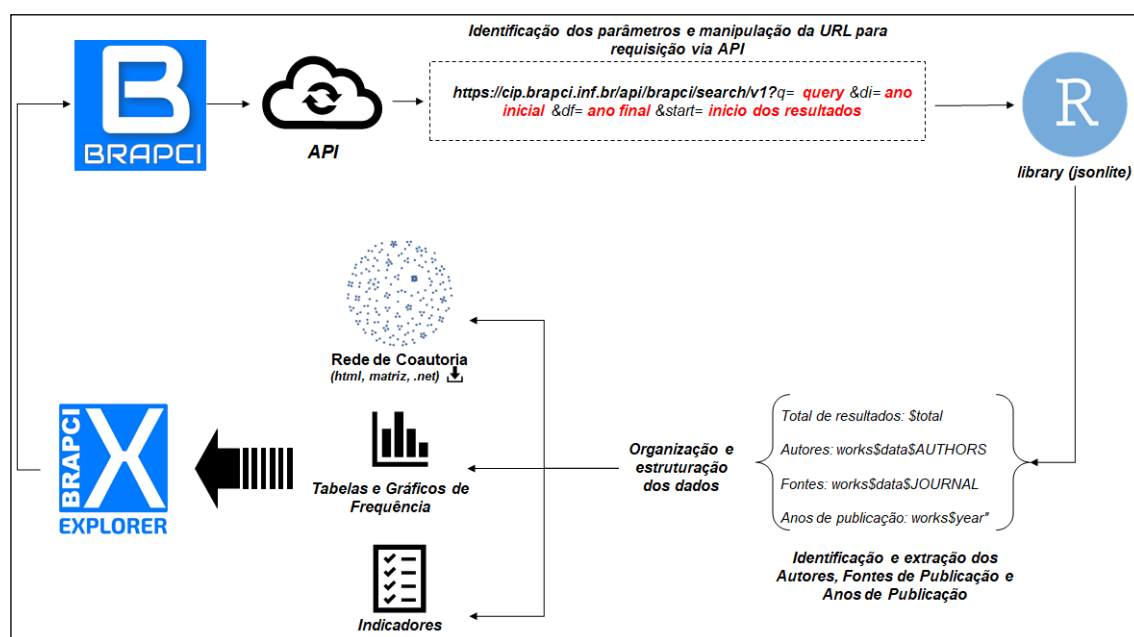
Sendo assim, *query* refere-se ao termo de busca (parâmetro *q* da *url*); *ano inicial* (*di*) e *ano final* (*df*) estabelecem a janela temporal em que a busca retornará resultados; *inicio dos resultados* determina o índice do elemento da página de resultados que o usuário está consultando (parâmetro *start*). Referente a este último parâmetro, cada requisição a ser feita via API gera 10 resultados por páginas em que os 10 primeiros resultados serão apresentados quando este parâmetro é igual a zero, isto é *start=0*, os próximos 10 resultados, estarão na segunda página iniciando a partir do parâmetro *start=10*, na terceira, *start=20*, e assim por diante.

Por exemplo, uma requisição de API pela *url*: *https://cip.BRAPCI.inf.br/api/BRAPCI/search/v1?q=ciencia&di=1972&df=2026&start=20* refere-se a uma requisição de uma busca utilizando o termo “ciência” (*q=ciencia*) no entre os anos de 1972 e 2023 (*&di=1972&df=2023*) e está consultando a terceira páginas de resultados desta busca (*start=20*).

Após a compreensão da composição dos parâmetros presentes na *url*, foi implementado o processo de extração de dados e para isso, foram utilizada a biblioteca *jsonlite* da linguagem de programação R. Para extrair os dados referentes ao total de documentos, às autorias, fontes e ano de publicação de cada documento encontrado utilizou-se as funções *read_json* e *fromJSON*. Assim, a primeira foi utilizada somente para identificar o total de documentos

correspondentes à busca, e a segunda para permitir a conversão dos demais dados em formato JSON disponibilizados a partir da requisição via API da BRAPCI em quadro de dados (*dataframe*). O total de documentos está presente em *\$total* e o quadro de dados com os dados das publicações encontra-se em *\$works* contendo 14 colunas em que as colunas relativas aos dados de autoria, fontes e ano de publicação são identificadas em *works\$data\$AUTHORS*, *works\$data\$JOURNAL* e *works\$year*, respectivamente. A seguir, foi necessário construir *loopings* para que, a partir do total de documentos, fosse possível percorrer todas as páginas contendo resultados. Dessa forma, foi programado um código capaz de realizar todas as extrações, coletando corretamente dos os dados obtidos a partir das requisições. Todo o processo de codificação para construção do BRAPCI Explorer está presente na Figura 1.

Figura 1. Sistematização de todo o processo de desenvolvimento do BRAPCI Explorer.



Fonte: Elaboração dos autores.

As funções programadas extraíram corretamente todos os dados e a partir desta extração, iniciou-se o processo de organização e estruturação dos dados coletados para a construção da rede de coautoria, as tabelas, os gráficos de frequência e os indicadores bibliométricos, mencionados anteriormente. Para o desenvolvimento da rede de coautoria, o código identificou a lista de autores por publicação e converteu esta lista em uma matriz de coautorias que serviu de base para a concepção da rede.

É importante mencionar que as autorias, presentes em *works\$data\$AUTHORS*, identifica a lista de todos os autores de cada publicação separados por ponto e vírgula e ao identificar todos os autores por publicação o código retorna a lista dos autores separados por ponto e vírgula, por exemplo: *Autor A; Autor B; Autor C*. Neste caso, foi necessário realizar o processo de *split*, para que todos os autores fossem considerados para tabela de frequência. Em programação, o processo de “*split*” é uma operação que permite separar uma *string* (um conjunto de caracteres) em elementos menores com base em um delimitador, neste caso o ponto e vírgula. Este procedimento é útil quando uma *string* contém informações separadas por um caractere, como ponto e vírgula, e deseja-se dividi-la em partes individuais, neste caso, separar os autores de uma publicação, permitindo assim contar a frequência de cada um. A partir deste processo, construiu-se uma matriz de incidência das publicações com relação aos seus respectivos autores. Depois, foi realizada a conversão para uma matriz de coautorias a partir da

multiplicação da matriz de incidência com sua transposta, obtendo assim uma matriz quadrada de coautorias (elemento *mtx_coaut* no código).

Para a construção e a visualização da rede de coautorias foram utilizadas duas bibliotecas do R: *igraph* e *visNetwork*. A primeira, responsável por converter a matriz em formato de rede por meio da função *graph_from_adjacency_matrix* que organiza toda a matriz em uma lista de relações ponderadas entre os nós. Já a segunda biblioteca, utiliza-se desta organização para fornecer uma visualização da rede interativa, em que o usuário pode selecionar os nós de interesse por meio de uma caixa suspensa ou clicando em cada um, além da possibilidade de aplicar *zoom* e arrastar os nós, organizando-os da forma em que o usuário desejar, como ilustrado na Figura 4.

Após a construção da rede de coautorias, elaborou-se as tabelas de frequência da produção dos autores, das fontes e dos anos de publicação, isto é, foi identificada a quantidade de produção de documentos por autor, por fonte de publicação e por ano. A extração dos dados referentes aos autores, fontes e ano de publicação retornou uma lista para cada uma destas variáveis e para obter as tabelas de frequência, aplicou-se a função *table* (nativa do R) que converteu os dados de cada lista em tabela de frequência simples. Assim sendo, considerou-se todos os autores de cada publicação para construção da tabela de frequência e não só o primeiro autor.

Ademais, optou-se por visualizar a evolução temporal das publicações, isto é, a frequência de publicações por ano, por meio de gráficos de barras verticais e não por tabela. Para construção do gráfico foi utilizado a biblioteca *ggplot2*. A partir das informações referentes às frequências de publicação por autor, fonte e ano de publicação foi possível calcular os indicadores bibliométricos propostos: quantidade de autores distintos, média de documentos por autor, índice de coautoria, total de fontes distintas, média de documentos publicados por fontes e média de documentos por ano. O primeiro indicador refere-se à identificação do total de autores distintos encontrados, a média de documentos por autor refere-se à média aritmética da produção tomando a quantidade de documentos produzidos por cada autor. Já o índice de coautoria refere-se ao Índice de Colaboração (IC) definido em Grácio (2018) como a média aritmética de autores por artigo publicado⁴.

Analogamente aos indicadores baseados nos dados de autoria, os indicadores respectivos às fontes de publicação (total de fontes distintas) podem ser entendidos como a quantidade de diferentes fontes encontradas na busca, enquanto a média de documentos por fontes corresponde à média aritmética de documentos tomando a produção de cada fonte distinta. Da mesma forma, o último indicador (média de documentos por ano) estabelece a razão entre a quantidade de documentos e os anos de publicação.

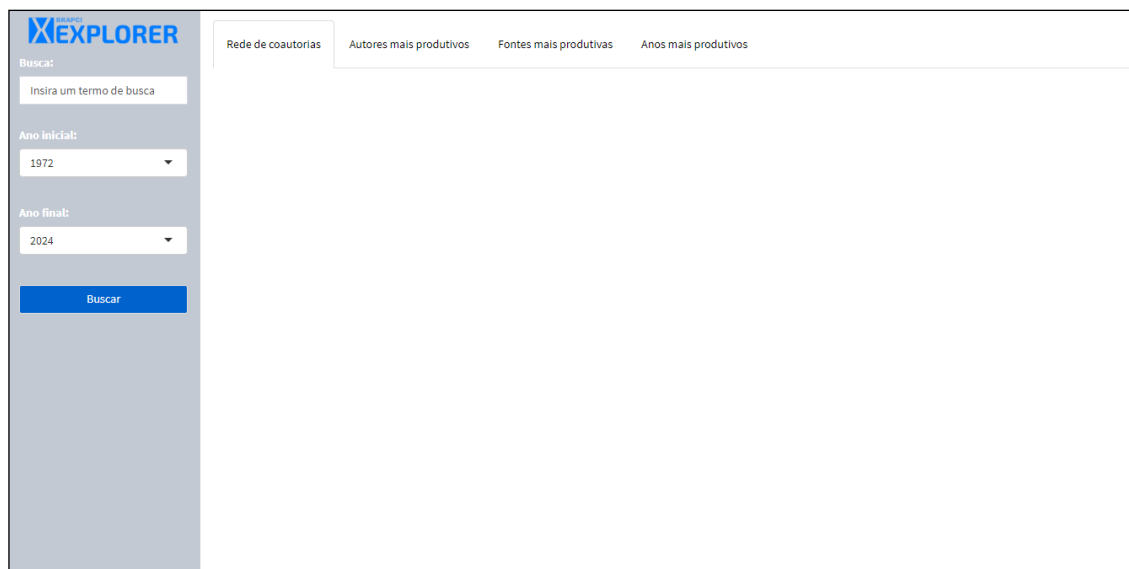
Tendo definido o cálculo dos indicadores, iniciou-se o processo de construção do ambiente web do BRAPCI Explorer. Assim, todo o processo descrito anteriormente (extração, organização, estruturação dos dados e cálculo dos indicadores) foi incorporado em um aplicativo web desenvolvido a partir da própria linguagem de programação R por meio da biblioteca *shiny*. A biblioteca possibilita a criação de ambientes web interativos que podem executar os mesmos processamentos e exibir diversas visualizações de dados, de forma semelhante ao que é possível fazer diretamente no R (ou no RStudio), além de possibilitar a personalização destes ambientes por meio manipulação de elementos *css* e *html*. O pacote *Bibliometrix* é uma das aplicações mais populares da biblioteca *shiny* porque contém um extenso conjunto de técnicas implementadas, além de uma interface intuitiva (Moral-Muñoz *et al.*, 2020).

Visto que o BRAPCI Explorer processa os dados da BRAPCI a partir da API da base, o ambiente proposto deveria ser capaz de replicar as buscas da própria BRAPCI, em que este

⁴ $IC = \frac{\sum_{i=1}^n X_i}{n}$

procedimento de replicação se dá a partir da manipulação da *url*. Nesse sentido, conforme apresentado na Figura 2, tem-se a interface do BRAPCI Explorer, em que na coluna lateral (*sidebar*), ocorre a replicação da busca a ser raspada da BRAPCI.

Figura 2. Interface do BRAPCI Explorer.



Fonte: Elaboração dos autores.

A *sidebar* é formada por campos de *input* de texto (campo de busca), dois seletores (ano inicial e ano final) e um botão (Buscar). Todos os campos possuem valores padrão com a finalidade de auxiliar o usuário no processo de busca, essa definição permite que o BRAPCI Explorer execute a busca sem interrupções visto que todos os campos estarão sempre preenchidos. Os dados são apresentados em um conjunto de quatro abas (*tabset*): Rede de coautorias, Autores mais produtivos, Fontes mais produtivas, Anos mais produtivos. Além dos resultados principais, mostrados no centro da tela, ainda são apresentadas informações complementares (total de documentos encontrados, total de autores e fontes distintas, índice de coautoria, média de documentos por autor e por fonte e média de documentos por ano). Ainda existem recursos interativos que permitem que o usuário faça intervenções na rede de coautoria, por meio do seletor de autor ou da manipulação dos nós da rede (Figura 4).

Apresentada a interface do BRAPCI Explorer, para validação da ferramenta foram realizadas diversas buscas, manipulando todos os parâmetros de busca no novo ambiente, e comparadas com os resultados fornecidos pela própria base de dados (<https://cip.BRAPCI.inf.br/>)⁵ para verificar a acurácia da nova ferramenta. Em todas as buscas o BRAPCI Explorer encontrou os mesmos resultados fornecidos pela BRAPCI.

Enquanto resultados desta pesquisa, tem-se a apresentação e descrição das funcionalidades do BRAPCI Explorer, e para isso utilizou-se somente uma estratégia de busca. Foi utilizado o termo de busca “produção científica” entre os anos 1972 e 2024 (Figuras 3, 4, 5, 6). Ademais, é proposto a inserção do BRAPCI Explorer junto ao portal da BRAPCI (Figuras 8 e 9), para demonstrar a viabilidade da vinculação entre os dois ambientes por meio de botões na página inicial e na página dos resultados de busca. Todo o processo de implementação do código e testes foi realizado utilizando o software RStudio em sua versão 4.1.2, em um computador com sistema operacional Windows 10, processador *Intel Core(TM) i7-8550UCPU*

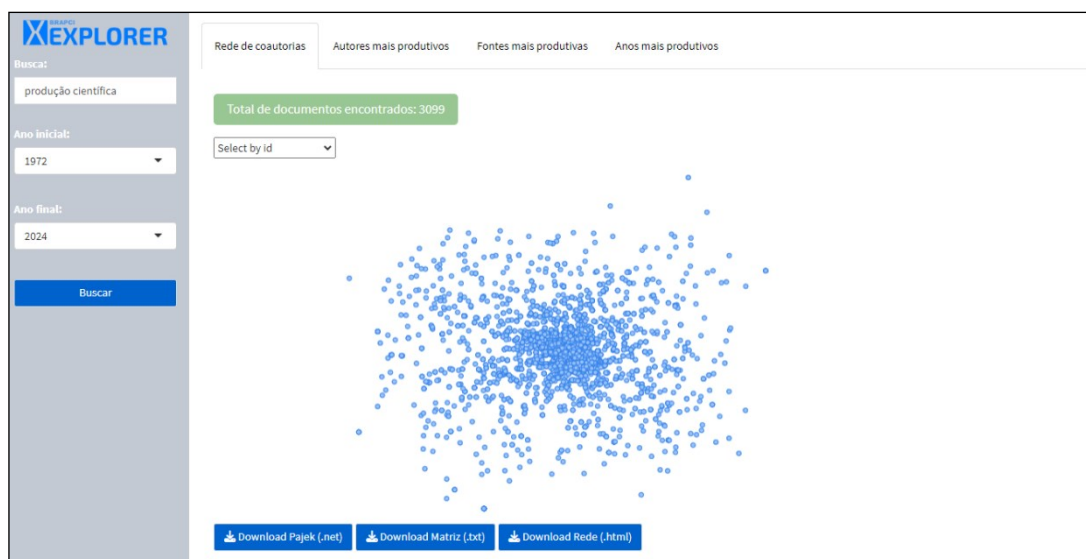
⁵ Este endereço corresponde da nova versão da BRAPCI que utiliza dos mesmos dados oriundos da API utilizada para o desenvolvimento do BRAPCI Explorer. Este endereço difere, em termos de resultados, do antigo: <https://BRAPCI.inf.br>.

@ 1.80GHz-1.99 GHz, SSD 200Gb e 8Gb de memória RAM. Os processamentos foram realizados em 13 de novembro de 2023.

3 RESULTADOS

Inserido o termo de busca “produção científica”, apresentam-se todas as funcionalidades, organizadas em quatro abas, do BRAPCI Explorer. A primeira aba, demonstrada nas Figuras 3 e 4, denominada “Rede de coautorias”, apresenta a rede ponderada de coautorias entre todos os autores de todas as produções recuperadas. Juntamente com a rede, na caixa de texto em verde, é apresentada a quantidade de documentos encontrados pela busca inserida, no exemplo representado por 3099 documentos.

Figura 3. Apresentação dos resultados de busca na aba de Rede de Coautorias.

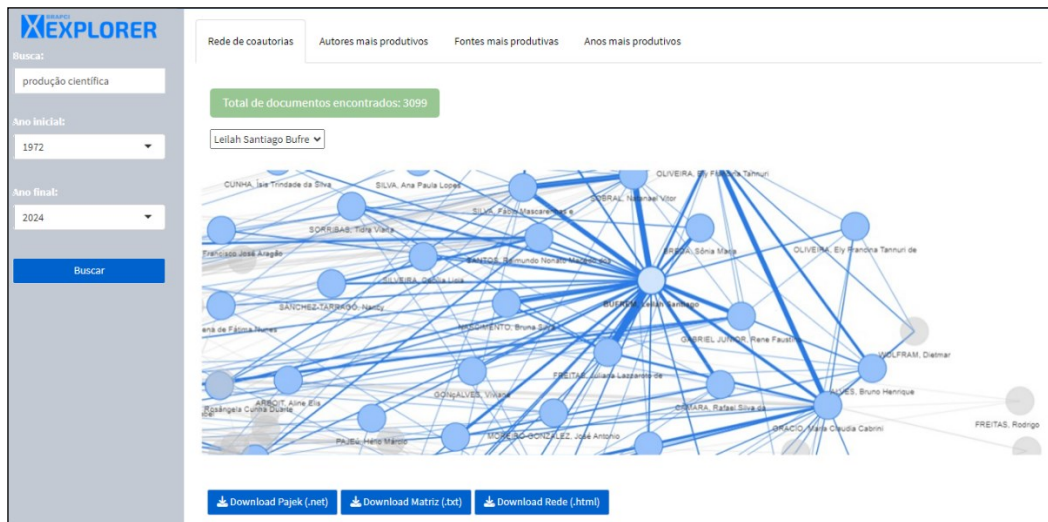


Fonte: Elaboração dos autores.

Com relação à rede, na Figura 3, cada nó representa um autor e as arestas as relações de coautoria estabelecida entre os autores em que a variação da espessura das arestas demonstram a quantidade de documentos produzidos em conjunto entre os autores. A geração automática das redes de coautoria do BRAPCI Explorer demonstra uma funcionalidade inovadora e inédita às bases de dados científicas, pois, tradicionalmente, análises de coautorias são realizadas após o usuário exportar os dados da busca e processá-los algum em software externo. Exemplos destas plataformas são o *VosViewer* ou *Bibliometrix*, softwares de cunho bibliométricos capazes de ler e gerar redes de coautoria a partir de um documento exportado de bases como *Web of Science*, *Scopus*, *Dimensions*, *Pubmed*, entre outras.

Ainda, caso o usuário prefira a visualização desta rede em softwares de visualização de redes, o BRAPCI Explorer oferece a opção de *download* da rede em formato *.net*. Esta extensão é certamente um dos arquivos de rede mais versáteis podendo ser executado (para visualização de redes) em softwares como *Pajek*, *Ucinet*, *Gephi*, *VosViewer*, entre outros. Além desta opção, é possível realizar o *download* da rede em configuração de matriz e formato *.txt* (tabulado). É sabido que, em redes muito grandes, o acesso à matriz de dados é de grande valia, visto que, é possível observar as principais relações estabelecidas pelos nós da rede. E mais, o acesso a matriz de coautorias é um resultado incomum aos softwares de análises bibliométricas. Outra funcionalidade é a opção do usuário interagir com a rede por meio das ferramentas de zoom e seleção de nó, como apresentado na Figura 4.

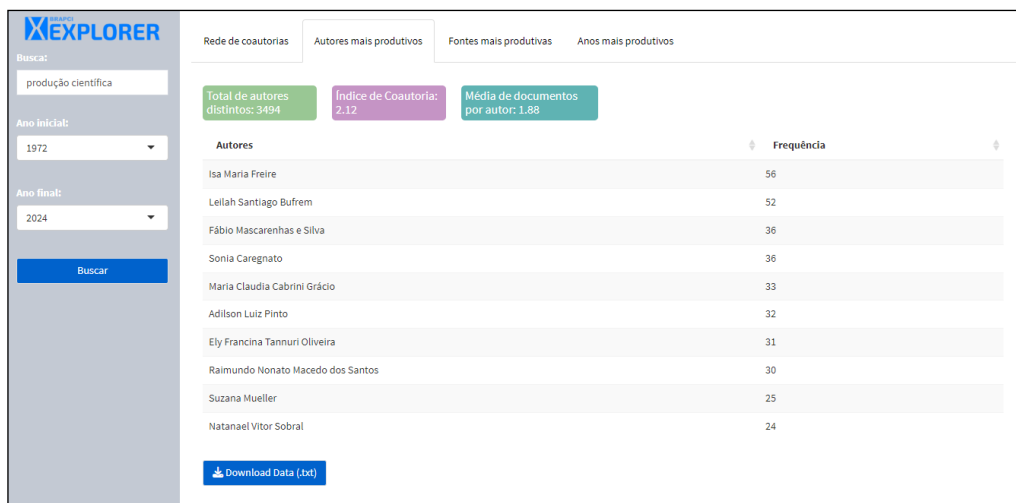
Figura 4. Interação com a Rede de coautorias utilizando seletor de autores.



Fonte: Elaboração dos autores.

A Figura 4 exemplifica a interação do usuário ao selecionar, na caixa de seleção de autores, e do zoom sobre o nó (nome) *BUFREM, Leilah Santiago* referente a pesquisadora Leilah Santiago Bufrem, idealizadora da BRAPCI. Caso o usuário opte por salvar a rede da forma que o BRAPCI Explorer a exibe, é possível realizar o *download* em formato *html*, de modo que o arquivo salvo mantém as propriedades de interação anteriormente descritas. A seguir, na segunda aba, tem-se a exibição da tabela dos autores mais produtivos com relação a busca realizada, juntamente com os indicadores que descrevem o total de autores distintos encontrados, o índice de coautoria e a média de documentos por autor, como apresentado na Figura 5.

Figura 5. Apresentação dos resultados de busca na aba Autores mais produtivos.



Fonte: Elaboração dos autores.

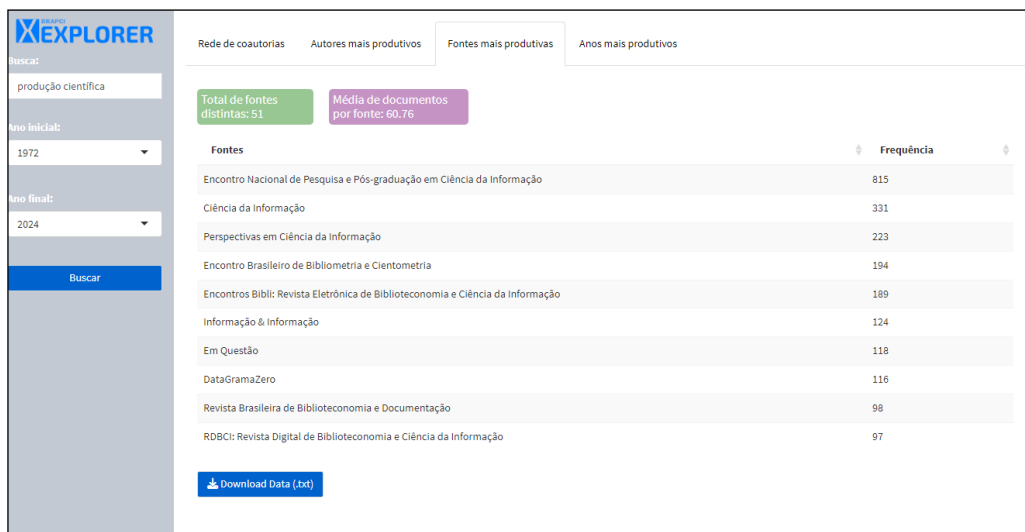
Para apresentação dos autores mais produtivos, o BRAPCI Explorer exibe uma tabela de frequência contendo os 10 autores mais produtivos. É importante ressaltar que a contagem (frequência) descrita nesta tabela corresponde a contagem de todos os documentos aos quais cada autor possui seu nome na lista de autores, ou seja, o BRAPCI Explorer não fraciona a contagem de autorias. Na Figura 5, por exemplo, ao identificar que as pesquisadoras Isa Maria Freire e Leilah Santiago Bufrem, nota-se que elas foram responsáveis por 56 e 52 documentos

respectivamente, a ferramenta não distingue se esta produção é oriunda de produção individual ou em coautoria com outros pesquisadores encontrados pela busca.

Com relação aos indicadores calculados, o BRAPCI Explorer identifica a quantidade total de autores distintos. Neste caso, 3494 autores distintos foram responsáveis pela produção dos 3099 documentos. O próximo indicador refere-se ao Índice de Coautoria, descrito como a razão entre a quantidade de autores por artigo e a quantidade total de documentos encontrados. Tem-se 2,12, ou seja, cada publicação possui, em média, 2,12 autores. Já o outro indicador, calcula a média de documentos por autor, isto é, estabelece a média aritmética da quantidade de documentos produzidos por autor. Neste caso, tem-se 1,88. Isto é, em média, cada autor produziu 1,88 documentos segundo a busca realizada.

É importante enfatizar que mesmo que a aba exiba somente os 10 resultados com o maior número de Frequência, o usuário tem a possibilidade de salvar a tabela contendo todos os autores distintos e suas respectivas produções ao clicar em “Download Data (.txt)” gerando um arquivo intitulado *Autores.txt* contendo uma tabela com os dados completos. Todos os indicadores são calculados levando em consideração todos resultados e não somente os 10 a serem exibidos nesta aba. De forma semelhante aos indicadores em nível de autor, tem-se, na Figura 6, a terceira aba contendo a tabela de frequência das 10 fontes de publicação mais produtivas e os indicadores que denotam o total de fontes distintas e a média de documentos por fonte.

Figura 6. Apresentação dos resultados de busca na aba Fontes mais produtivas.



Fonte: Elaboração dos autores.

Analogamente a tabela de autores mais produtivos, exibida na Figura 5, o BRAPCI Explorer realiza a contagem (Frequência) de publicações por fonte. Ao realizar a análise de frequência, foi possível determinar o total de fontes distintas, ou seja, em quantas fontes de publicação únicas os documentos foram publicados. Neste caso, tem-se que os 3099 documentos foram publicados por 3494 autores em 51 fontes distintas.

Com relação ao outro indicador presente na aba, tem-se a média de documentos por fonte. Este indicador afere a média aritmética da produção segundo fonte de publicação, e para esta busca, tem-se 60,76. Em outras palavras, cada fonte publicou, em média, 60,76 documentos relativos à expressão “produção científica”. Para acessar todas as informações referentes a ambos indicadores, o usuário poderá salvar a tabela completa contendo todas as fontes e suas respectivas produções clicando em “Download Data (.txt)” que irá gerar um arquivo nomeado como *Fontes de Publicação.txt*.

A aba “Anos mais produtivos”, exibida na Figura 7, refere-se à análise temporal das publicações por ano. Diferente das demais, optou-se pela visualização dos dados por meio de um gráfico de barras que exibe a produção referente aos últimos 15 anos.

Figura 7. Apresentação dos resultados de busca na aba Anos mais produtivos.



Fonte: Elaboração dos autores.

Optou-se pela exportação dos dados completos em formato .txt ao invés de um gráfico (em formato de imagem), pois, a depender do tamanho da busca, a visualização poderia ser prejudicada. Além disso, está aba conta com o indicador que calcula a média de documentos por ano. Neste caso, obteve-se o valor 61,98, ou seja, foi encontrada uma produção média de 61,98 artigos por ano.

Apresentado o BRAPCI Explorer, propõe-se a inserção da nova ferramenta junto aos demais recursos da BRAPCI. A base conta com alguns recursos que são apresentados no rodapé do endereço <https://cip.BRAPCI.inf.br>, até o momento da publicação desta pesquisa alguns destes recursos estão em fase de desenvolvimento e atualização (<https://cip.BRAPCI.inf.br/books/about>), conforme mencionado por Silva (2023). Portanto, foram sugeridas duas propostas de vinculação entre os ambientes, a primeira é apresentada na Figura 8.

Figura 8. Proposta de inserção do link de acesso ao BRAPCI Explorer na página inicial da BRAPCI.



Fonte: Elaboração dos autores.

Nesta proposição o BRAPCI Explorer é apresentado juntamente com as novas funcionalidades, ou ambientes, ainda em desenvolvimento pela BRAPCI. Na Figura 8, o *card* é clicável e encontra-se no estado *focus*, ou seja, selecionado pelo usuário ao contrário dos demais *cards*. A sugestão de inserir um link na página inicial da base justifica-se pelo fato de ajudar na popularização das ferramentas complementares da base, dessa forma o usuário pode conhecer as opções de busca e análise existentes. Além disso, sugere-se ainda a inserção de um link na página de resultados de busca da BRAPCI, apresentado na Figura 9.

Figura 9. Proposta de inserção do link de acesso ao BRAPCI Explorer na página de resultados de busca da BRAPCI.



Fonte: Elaboração dos autores.

| 13

Neste cenário o usuário realizou a busca na tela inicial da BRAPCI e clicou no botão “Pesquisar” sendo direcionado para a tela de resultados de busca. A sugestão aqui é de apresentar uma vinculação em formato de botão, similar aos botões “*Analyze Results*” na *Web of Science* e na *Scopus*. Como o usuário já realizou a busca no ambiente da BRAPCI, idealmente ele deveria ser redirecionado para o BRAPCI Explorer com os resultados da busca já apresentados, entretanto, devido ao fato do BRAPCI Explorer ser um ambiente, ainda, independente da base, o redirecionamento após o usuário clicar no botão “BRAPCI Explorer” na página de resultados da BRAPCI levaria para a interface do BRAPCI Explorer (Figura 2). Assim, basta que ele insira novamente o termo de busca para ter acesso aos resultados no ambiente do BRAPCI Explorer. Entende-se que esta proposição (modelo ideal) seria a mais adequada tendo em vista que é a prática adotada por outras bases.

4 CONCLUSÃO

As bases de dados científicas têm sua importância reconhecida para a manutenção e disseminação da ciência, por meio do armazenamento, indexação, representação e disponibilização de informações científicas. O surgimento das bases *Web of Science* e *Scopus*, no fim da década de 90 e início dos anos 2000, foram grandes marcos para o desenvolvimento destes ambientes. No contexto da Ciência da Informação brasileira, pode-se mencionar a BRAPCI, lançada em 2010, como o repositório mais abrangente da produção científica periódica no país. Contudo, quando comparada à *Web of Science* e *Scopus*, a BRAPCI carece de recursos bibliométricos para além dos indicadores de produção.

Assim, entende-se que a BRAPCI reúne um conjunto de informações científicas relevantes e passíveis de análises bibliométricas. Nesse sentido, esta pesquisa propôs um novo ambiente, inovador e interativo, denominado BRAPCI Explorer, voltado às análises bibliométricas, a partir da extração de dados via API da base, um método de grande valia no processo de coleta e estruturação de dados favorecendo análises diversas. A partir desta técnica de extração, foi possível coletar os dados disponibilizados pela base e apresentá-los no BRAPCI Explorer

Este novo ambiente foi construído a partir da linguagem de programação R utilizando a biblioteca *shiny* para elaboração da interface e a biblioteca *jsonlite* para o processo de coleta de dados. O desenvolvimento da interface partiu da organização dos campos de *input* de texto, dos seletores de campo e período e do botão de busca em uma estrutura lateral, *sidebar*. A apresentação dos dados, a partir da busca realizada na *sidebar*, ocorreu em uma estrutura de abas, *tabset*, onde o usuário pode navegar entre as opções de “Rede de coautorias”, “Autores mais produtivos”, “Fontes mais produtivas” e “Anos mais produtivos”.

Concebido o BRAPCI Explorer, foram realizados testes para avaliar o uso do novo ambiente web. De forma preliminar, os testes se alinharam aos resultados da própria BRAPCI, isto é, os resultados extraídos (via API) pelo BRAPCI Explorer foram condizentes com aqueles fornecidos pelo novo *web site* da BRAPCI. Ademais, os testes demonstraram as possibilidades de visualização e análise de dados via BRAPCI Explorer. Uma das funcionalidades inovadoras é a geração automatizada da rede de coautorias, presente na primeira aba da ferramenta, facilitando as análises bibliométricas porque o usuário não precisa realizar o *download* dos dados e utilizar outras ferramentas, já que tudo é disponibilizado na interface do BRAPCI Explorer. Além disso, oferece ainda a possibilidade de salvar os dados caso o usuário tenha necessidade em diferentes formatos: *.net*, *matriz* e *html*. Os recursos interativos incluem manuseio dos nós da rede de coautorias e identificação dos nós por meio do seletor de autores.

Na segunda, terceira e quarta abas tem-se visualização de autores, fontes e anos de publicação mais produtivos, juntamente com indicadores bibliométricos referentes a essas unidades de análise. Em que, nos dois primeiros casos, a interface apresenta apenas os 10 primeiros resultados, organizados em tabela de frequência e no caso dos anos de publicação mais produtivos, optou-se pela apresentação dos dados de frequência em formato de gráfico de barras. O usuário ainda tem acesso aos dados completos por meio dos botões de *download*.

No que diz respeito aos autores, tem-se os indicadores de total de autores distintos, índice de coautoria e média de documentos por autores. Já com relação às fontes de publicação, o novo ambiente fornece os indicadores referentes às fontes distintas e a média de documentos produzidos por fonte de publicação. E na última aba, é calculado a média de documentos por ano de publicação. Com isso, o BRAPCI inova ao disponibilizar os indicadores automaticamente a seus usuários.

Ao testar o novo ambiente foi realizada uma busca com o termo “produção científica” encontrando 3099 documentos distintos produzidos por 3494 autores distintos em 51 fontes de publicação distintas. O BRAPCI Explorer foi capaz de gerar a rede de coautorias contendo todos estes autores, bem como gerar a matriz de coautorias e arquivo de redes *.net*, elaborar todas tabelas e gráfico de frequência, além dos cálculos dos indicadores. É importante destacar que este processo durou aproximadamente 1 minuto e 40 segundos, ou seja, ao construir a rede de coautorias foram realizadas 6.102.271 interações entre todos os autores identificados. Mesmo que este processo possa aparentar certa lentidão, o BRAPCI Explorer fornecerá todos resultados prontos para o uso e/ou exportação via *download*. Após os testes, para demonstrar a viabilidade do BRAPCI Explorer foram sugeridas duas formas de vinculação entre os ambientes, um botão da página inicial e outro na página de resultados de busca da BRAPCI, demonstrando assim que a conexão entre os ambientes não demandaria grande esforço de programação.

Dentre as limitações destaca-se que, para a versão *web-app*, a ferramenta enfrentará problemas em processar grandes conjuntos de dados, pois, atualmente, sua hospedagem no servidor *shinyapps.io* possui limite de *instance size* de 1 *gigabyte*. Para cenários em que a busca ultrapasse o limite de processamento, sugere-se a utilização do BRAPCI Explorer direto pelo software R além do ajuste do limite de memória (funções *memory.size* e *memory.limit*). Destaca-se que os testes foram realizados com as configurações mencionadas na seção de metodologia, assim a sobrecarga computacional pode variar de acordo com o cenário de cada dispositivo. Entretanto, a possibilidade de utilizar o BRAPCI Explorer de modo online favorece o uso em diversos navegadores de internet como *Google Chrome*, *Microsoft Edge*, *Mozilla Firefox*, *Safari*, entre outros, sem a necessidade de o usuário instalar o R.

Além disso, reitera-se que inicialmente, o BRAPCI Explorer foi idealizado para realizar um processo de *web scraping*, contudo, após o acesso à API da base, o ambiente foi recodificado. Tendo em vista que a conexão de dados via API é mais rápida, compreende-se que essa alteração tornou o novo ambiente mais viável computacionalmente. Até o momento desta publicação, ainda não é possível adicionar mais filtros às buscas do BRAPCI Explorer, como apresentado no *web site* da BRAPCI: campo de busca, que delimita onde será pesquisado a *query* (título, resumo, palavras-chave, autores, todos campos) e seleção de fontes de informação (revistas ou eventos). Almeja-se que este incremento esteja presente nas próximas atualizações do BRAPCI Explorer, replicando completamente a busca da base. Contudo, é necessário reforçar que atualmente a base está em processo de atualização.

Em vista do exposto, pode-se concluir que o BRAPCI Explorer contribui para uma melhor visualização dos dados disponibilizados pela BRAPCI, isso porque apresenta visualizações aprimoradas, em formato de rede, gráficos e tabelas, além de indicadores bibliométricos, sem que o usuário precise sair do ambiente web e sem a necessidade de conhecimentos de programação. Por fim, destaca-se a interface intuitiva e como identidade visual próxima à da BRAPCI a fim de causar identificação, para que o usuário se sinta seguro em navegar entre os ambientes. As próximas etapas e para pesquisas futuras pretende-se manter o BRAPCI Explorer em pleno funcionamento e realizar atualizações referentes à interface, bem como realizar teste de usabilidade.

REFERÊNCIAS

ARRUDA, W. R.; FELIPE, C. B. M.; SANTOS, R. F. dos. Avaliação da qualidade das bases de dados BRAPCI e PERI da área de Ciência da Informação. **Ciência da Informação em Revista**, Maceió, v. 7, n. 1, p. 121–137, 2020. DOI:

<https://doi.org/10.28998/cirev.2020v7n1h>. Acesso em: 10 set. 2023

BRASIL. API de dados. **Portal da transparência – Controladoria-Geral da União**. 2023. Disponível em: <https://portaldatransparencia.gov.br/api-de-dados>. Acesso em: 10 nov. 2023.

BRAPCI, Sobre - Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). Disponível em: <https://brapci.inf.br/index.php/res/about>. Acesso em: 10 set. 2023.

BUFREM, L. S.; COSTA, F. D. O.; GABRIEL JUNIOR, R. F.; PINTO, J. S. P. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 2, p.22-41, 2010. Disponível em: bit.ly/3EPufU0. Acesso em: 10 set. 2023.

CASTANHA, R. G.; FRANCO; F. **BRAPCI Explorer**. DOI: <https://doi.org/10.5281/zenodo.10126529>.

CLARIVATE. **The History of ISI and the work of Eugene Garfield**. Disponível em: <https://clarivate.com/the-institute-for-scientific-information/history-of-isi/>. Acesso em: 09 set. 2023.

FREITAS, J. L.; BUFREM, L. S.; GABRIEL JUNIOR, R. F. Proposta de metodologia para a recuperação da produção científica em ciência da informação na base BRAPCI. **Ponto de Acesso**, Salvador, v. 4, n. 3, p. 45-67, 2010. DOI: <https://doi.org/10.9771/1981-6766rpa.v4i3.4629>.

GRÁCIO, M. C. C. Colaboração científica: indicadores relacionais de coautoria. **Brazilian Journal of Information Science: research trends**, Marília, v. 12, n. 2, 2018. DOI: <https://doi.org/10.36311/1981-1640>.

GRACIANO, H. L. S. **ScraperCI**: um protótipo de Web scraper para coleta de dados. 2022. Dissertação (Mestrado em Ciência da Informação). Universidade Federal de São Carlos, São Carlos, SP. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/17166>. Acesso em 25 set 2023.

MORAL-MUÑOZ, J *et al.* Software tools for conducting bibliometric analysis in science: An up-to-date review. **El profesional de la información**, Madrid, v. 29, n. 1, e290103. 2020. 15/09/2023. DOI: DOI: <https://doi.org/10.3145/epi.2020.ene.03>.

THELWALL, M.; SUD, P. Scopus 1900–2020: Growth in articles, abstracts, countries, fields, and journals. **Quantitative Science Studies**, Cambridge, MA, v. 3, n. 1, p. 37-50, 2022. DOI: https://doi.org/10.1162/qss_a_00177.

SCHOTTEN, M. *et al.* A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In: CANTU-ORTIZ, F.J. (ed.). **Research analytics: boosting University Productivity and competitiveness through scientometrics**. Auerbach Publications, 2017. p. 31-58. DOI: <https://doi.org/10.1201/9781315155890>.

SILVA, L. M. **BRAPCI livros**: uma proposta de organização e recuperação de livros digitais científicos abertos em Ciência da Informação. 2023. (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Grande do Sul, Rio Grande do Sul. Disponível em: <https://lume.ufrgs.br/handle/10183/257995>. Acesso em: 25 set 2023.

VELEZ-ESTEVEZ, A. *et al.* New trends in bibliometric APIs: A comparative analysis, **Information Processing & Management**, London, v. 60, ed. 4, 2023. DOI: <https://10.0.3.248/j.ipm.2023.103385>.