

### Authors' correspondence

<sup>1</sup> Universidade de Marília  
Marília, SP - Brazil  
r.castanha@gmail.com

<sup>2</sup> Universidade Estadual Paulista  
Rio Claro, SP - Brazil  
franfranco.santos@gmail.com

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

Rafael Gutierrez Castanha<sup>1</sup> Francielle Franco dos Santos<sup>2</sup>

### 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 analyzes. 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, totally 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.

## BRAPCI Explorer: um novo ambiente web para análises bibliométricas a partir da base de dados referencial de artigos de periódicos em Ciência da Informação

### 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.

#### **CRedit**

- **Acknowledgments:** Not applicable.
- **Funding:** Not applicable.
- **Conflicts of interest:** The authors certify that they have no commercial or associative interest that represents a conflict of interest in relation to the manuscript.
- **Ethical approval:** Not applicable.
- **Availability of data and material:** The code is available at: <https://doi.org/10.5281/zenodo.8371105> Application running: <https://fctools.shinyapps.io/BRAPCIexplorer>
- **Authors' contributions:** Author contributions: Conceptualization, Research, Methodology, Software, Validation, Visualization, Resources Writing - original draft, Writing - review & editing, Formal analysis: CASTANHA, R. G.; FRANCO, F. Data curation, Project management, Supervision: CASTANHA, R. G.

| 2

**JITA:** BB. Bibliometrics methods



Article submitted to the similarity system

Submitted: 25/09/2023 – Accepted: 30/10/2023 – Published: 23/11/2023

Editor: Gildenir Carolino Santos

## 1 INTRODUÇÃO

Scientific databases can be understood as information retrieval systems that aim to store, index, represent, and make available scientific information based on the needs of users in a given community, using information retrieval mechanisms to contribute to the dissemination of science from scientific information sources (Arruda; Felipe; Santos, 2020).

A major historical milestone in the development of scientific databases is the creation of the Institute of Scientific Information (ISI) and the Science Citation Index (SCI) from the efforts of Eugene Garfield in the 1950s and 1960s, which culminated in the creation of the Web of Science database, launched in 1997, still in CD-ROM format after Thomson Reuters acquired ISI in 1992 (Clarivate, 2023). In 2001, the database incorporated the Essential Science Indicators, a set of bibliometric indicators related to the production and impact of productions at the level of authors, publication sources, articles and countries, which is widely disseminated and used in a variety of research.

Another database of enormous importance on the world stage is Scopus, which was launched in 2004 (Schotten et al., 2017; Thelwall; Sud, 2022) and belongs to the Elsevier group. Like Web of Science, Scopus is a database and a citation database, i.e., both are able to capture the connections made by the documents indexed in their respective databases. In addition, like Web Of Science, Scopus is a multidisciplinary database with broad coverage that provides bibliometric indicators to its users. As a result, they have become important data sources for research at the heart metric information studies and Information Science (IS).

In addition, it is possible to list databases specialized in certain fields of knowledge, such as PubMed/Medline (health sciences), MathSciNet (mathematics), IEEE Xplore (engineering and computer science), EconLit (economics), ERIC - Education Resources Information Center (education) and LISA - Library & Information Science Abstracts (information science). Specifically in the field of Information Science, BRAPCI (Reference Database of Journal Articles in Information Science) stands out at the national level. BRAPCI has been developed on the Web and is recognized as the most comprehensive repository of periodical scientific production in the field of Information Science in Brazil (Freitas; Bufrem; Gabriel Junior, 2010). The database is the result of the research project "Methodological options in research: the contribution of the field of information to the production of knowledge in higher education", by Leilah Santiago Bufrem, whose purpose was to support studies and proposals in the field of Information Science, based on institutionally planned activities (BRAPCI, 2023).

However, compared to other databases such as Web of Science or Scopus, BRAPCI provides few bibliometric metrics, limiting itself to production indicators (total number of publications) by author, source and year of publication. Therefore, given the relevance of BRAPCI, it is plausible to say that adding new bibliometric features to the database would be of great value and could significantly enrich the analysis and evaluation of scientific production in the field of Information Science. This would allow researchers to gain a more comprehensive understanding of the published work, as well as facilitate the identification of emerging research trends within the field.

For this to be possible, it is necessary to access all possible search results from the database, so this research relies on data extraction via API (Application Programming Interface) to propose a new web environment<sup>1</sup>. An Application Programming Interface (API) is a mechanism that enables interaction between different software, allowing developers to access and use specific functionalities of a system, application or service. Using an API, developers

---

<sup>1</sup> It should be clarified that this research initially proposed data extraction via web scraping. However, after the peer review process, the reviewers suggested contacting the BRAPCI administrators to request access to the database's API, given that BRAPCI is expected to be updated in the coming months after this publication, which could make the new environment obsolete shortly after its release. The tool, which uses web scraping as its data extraction method, will be available for as long as the current version of BRAPCI is still accessible, at: [https://fctools.shinyapps.io/brapciexplorer\\_ws/](https://fctools.shinyapps.io/brapciexplorer_ws/)

can create programs that connect directly to the database, selecting and manipulating information in an automated way, providing an effective way of obtaining data without relying on traditional navigation (through the graphical interface), being a valuable tool for the integration and automation of data-related processes, thus APIs can be used to enable direct access to the database, eliminating the need for manual interactions with user interfaces (BRASIL, 2023).

Therefore, this study aims to answer the following problem: How can a new web environment, obtained from the BRAPCI API, contribute to a better visualization, understanding and analysis of different bibliometric data? The objective is to propose a new web environment, called BRAPCI Explorer, aimed at bibliometric analysis, based on data requests via the BRAPCI database API, extending and structuring the visualization of indexed data. Specifically, to understand how this environment can contribute to the enrichment of the database, considering the importance of integrating bibliometric analysis into the databases through a web environment with an intuitive interface.

In this context, the proposal for BRAPCI Explorer is in line with BRAPCI's own needs, as described on its portal, when it states that BRAPCI is conducting an online survey with the aim of evaluating, from the user's point of view, the possibility of implementing future improvements to the interface, content and level of satisfaction (BRAPCI, 2023). In addition, there are other initiatives such as BRAPCI Livros by Silva (2023) and ScraperCI by Graciano (2020) and Graciano and Ramalho (2023). The latter is based on the web scraping method and both differ from BRAPCI Explorer, since ScraperCI focuses on the information retrieval process and the environment proposed here focuses on bibliometric elements such as data organization and visualization and the generation of indicators.

Just as OpenAlex, Semantic Scholar or Unpaywall complement large databases such as Web of Science and Scopus (Velez-Estevez et al., 2023), it is understood that complementary applications could be developed from BRAPCI. Other recent initiatives have also used the database to develop resources, such as BRAPCI Journals, Benancib and BRAPCI Events<sup>2</sup>, but it has not been possible to find any application that provides the functionalities offered by BRAPCI Explorer. The environment proposed here allows users to easily access bibliometric analyses without the need to export data, process them or use complementary tools.

## 2 METHODOLOGY

In order to propose a new web environment for bibliometric analysis, a code was implemented in the programming language R, capable of: accessing BRAPCI data via API; performing searches; organizing and structuring them in an interactive website; improving the visualization of the data indexed in the database. This new environment was called BRAPCI Explorer, available at <https://fctools.shinyapps.io/BRAPCIexplorer/>.

This code was able to generate the co-authorship network between the authors identified in a given search, to generate the frequency table of the list of the most productive authors and publication sources, to generate the production graph by year of publication, as well as to calculate bibliometric indicators related to the number of different authors found, the average number of documents per author, the co-authorship index, the total number of different sources, the average number of documents published per source, and the average number of documents per year. The code is open and available in Castanha and Franco (2023).

However, a priori access to the database API was requested from the BRAPCI administrators, which was promptly granted. At this point, it should be noted that the database is currently being updated and not all functionalities were available at the time of publication. Even though access to the API of the database was granted, before extracting the data we

---

<sup>2</sup> Resources available at: <https://cip.brapci.inf.br/>

checked the existence of BRAPCI's terms of use to see if there were any restrictions on the use of its API. This check was done by accessing the terms of use of the database (<https://cip.BRAPCI.inf.br/about> and <https://BRAPCI.inf.br/index.php/res/about>).

No copyright or usage license information was found. After these checks, the code was implemented. Thus, the design of BRAPCI Explorer took place on three fronts: i) outlining and programming the entire data extraction process via API request; ii) structuring and organizing the extracted data; iii) building the interface of the new environment. The entire process is illustrated in Figure 1.

The first stage of the data extraction process was based on understanding the file generated by the BRAPCI API. The API provides JavaScript Object Notation (JSON) files, which are very common for APIs. In the case of BRAPCI, it was necessary to identify all the objects and to understand what elements these objects referred to, identifying the total number of publications and the fields of authorship, source of publication and years of publication, since these elements would be used to build the co-authorship network, the tables and graphs and to calculate the bibliometric indicators.

After this identification, it was necessary to understand the composition of the parameters present in the BRAPCI API *url*. This was done in order to be able to manipulate the elements that make up the *url* when searching the database. It is important to mention that the understanding of these parameters is of the utmost importance, since the database offers different parameters in its searches and only displays 10 results per page. Therefore, in order to be able to extract all the data resulting from the searches, it was necessary to manipulate the *url* so that the programmed algorithm would be able to go through all the pages correctly. The base *url* is as follows

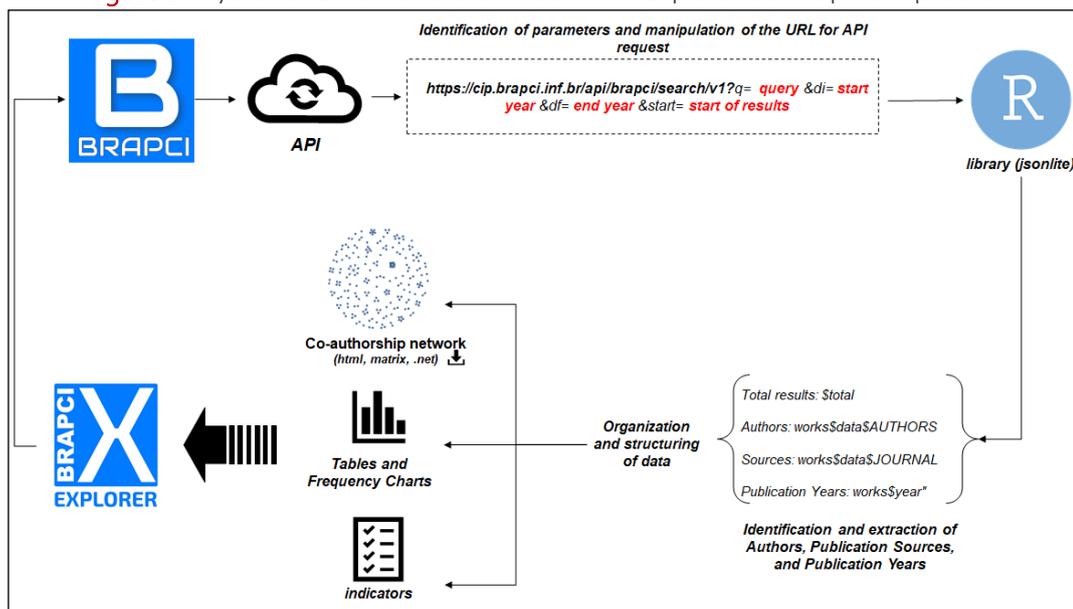
*https://cip.BRAPCI.inf.br/api/BRAPCI/search/v1?q=query&di=anoinitial&df=anofinal&start=iniciodosresultados*

Thus, *query* refers to the search term (parameter *q* of the *url*); start year (*di*) and end year (*df*) define the time window in which the search returns results; start of results determines the index of the element of the results page that the user consults (parameter *start*). With regard to this last parameter, each request to be made via the API will generate 10 results per page, with the first 10 results being displayed if this parameter is equal to zero, i.e. *start=0*, the next 10 results being displayed on the second page, starting from the parameter *start=10*, on the third, *start=20*, and so on.

For example, an API request via the *url*: *https://cip.BRAPCI.inf.br/api/BRAPCI/search/v1?q=ciencia&di=1972&df=2026&start=20* refers to a request to search for the term "science" (*q=science*) between the years 1972 and 2023 (*&di=1972&df=2023*) and will consult the third page of results for this search (*start=20*).

After understanding the composition of the parameters present in the *url*, the data extraction process was implemented using the *jsonlite* library from the R programming language. The functions *read\_json* and *fromJSON* were used to extract data on the total number of documents, authors, sources and year of publication of each document found. The former was used only to identify the total number of documents corresponding to the search, while the latter was used to convert the other data in JSON format available from the request via the BRAPCI API into a data frame. The total number of documents is stored in *\$total* and the data frame containing the publication data is stored in *\$works* with 14 columns, where the columns related to authors, sources and year of publication are identified as *works\$data\$AUTHORS*, *works\$data\$JOURNAL* and *works\$year*, respectively. Next, it was necessary to construct loops so that, given the total number of documents, it would be possible to iterate through all pages containing results. In this way, a code was programmed that was able to perform all the extractions and correctly collate the data obtained from the queries. The entire coding process for building the BRAPCI Explorer is shown in Figure 1.

Figure 1 – Systematization of the entire BRAPCI Explorer development process.



Source: By the authors.

The programmed functions correctly extracted all the data and from this extraction began the process of organizing and structuring the data collected to build the co-authorship network, tables, frequency graphs and bibliometric indicators mentioned above. To develop the co-authorship network, the code identified the list of authors by publication and converted this list into a co-authorship matrix which served as the basis for designing the network.

It is important to mention that the authorships, present in `works$data$AUTHORS`, identifies the list of all the authors of each publication separated by semicolons and when identifying all the authors per publication, the code returns the list of authors separated by semicolons, for example: *Author A; Author B; Author C*. In this case, it was necessary to carry out the split process, so that all the authors were considered for the frequency table. In programming, the "split" process is an operation that allows you to separate a string (a set of characters) into smaller elements based on a delimiter, in this case the semicolon. This procedure is useful when a string contains information separated by a character, such as a semicolon, and you want to split it into individual parts, in this case separating the authors of a publication, thus allowing you to count the frequency of each one. Based on this process, a matrix was built of the incidence of publications in relation to their respective authors. It was then converted into a co-authorship matrix by multiplying the incidence matrix with its transpose, thus obtaining a square co-authorship matrix (`mtx_coaut` element in the code).

Two R libraries were used to build and visualize the co-authorship network: *igraph* and *visNetwork*. The first is responsible for converting the matrix into a network format using the `graph_from_adjacency_matrix` function, which organizes the entire matrix into a list of weighted relationships between the nodes. The second library, on the other hand, uses this organization to provide a visualization of the interactive network, in which the user can select the nodes of interest via a drop-down box or by clicking on each one, as well as the possibility of zooming in and dragging the nodes, organizing them in any way the user wishes, as illustrated in Figure 4.

Once the co-authorship network had been built, tables were drawn up showing the frequency of production of authors, sources and years of publication, i.e. the number of documents produced per author, per source of publication and per year. Extracting the data on authors, sources and year of publication returned a list for each of these variables and to obtain the frequency tables, the `table` function (native to R) was applied, converting the data from each

list into a simple frequency table. Therefore, all the authors of each publication were considered when constructing the frequency table, not just the first author.

In addition, it was decided to visualize the temporal evolution of publications, i.e. the frequency of publications per year, using vertical bar graphs instead of tables. The ggplot2 library was used to create the graph. Using the information on publication frequency by author, source and year of publication, it was possible to calculate the proposed bibliometric indicators: number of different authors, average number of documents per author, co-authorship index, total number of different sources, average number of documents published per source and average number of documents per year. The first indicator refers to the identification of the total number of different authors found, while the average number of documents per author refers to the arithmetic average of the production, taking the number of documents produced by each author. The co-authorship index refers to the collaboration index (CI), defined in Grácio (2018) as the arithmetic average of authors per published article<sup>3</sup>.

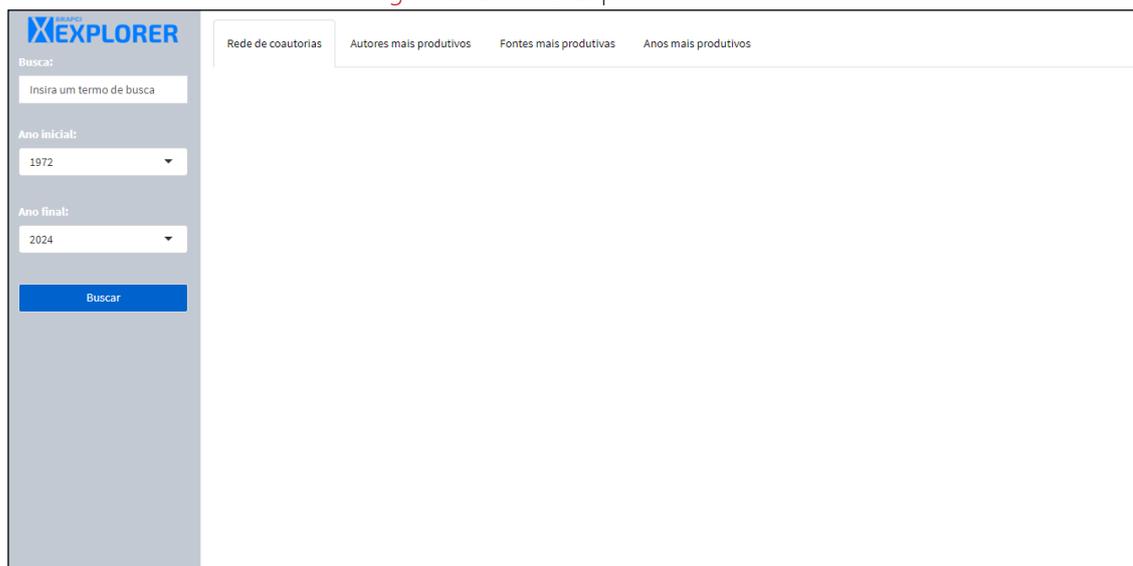
Similarly to the indicators based on authorship data, the indicators related to publication sources (total of different sources) can be understood as the number of different sources found in the search, while the average number of documents per source corresponds to the arithmetic average number of documents produced by each different source. Similarly, the last indicator (average number of documents per year) establishes the ratio between the number of documents and the years of publication.

After defining the calculation of the indicators, the process of building the BRAPCI Explorer web environment began. The whole process described above (extraction, organization, structuring of the data and calculation of the indicators) was incorporated into a web application developed from the R programming language itself using the Shiny libraries. The library makes it possible to create interactive web environments that can perform the same processing and display various data visualizations, similar to what can be done directly in R (or RStudio), and to customize these environments by manipulating css and html elements. The bibliometrics package is one of the most popular applications of the shiny library because it contains an extensive set of implemented techniques as well as an intuitive interface (Moral-Muñoz et al., 2020). Since BRAPCI Explorer processes BRAPCI data from the database's API, the proposed environment should be able to replicate searches from BRAPCI itself, where this replication procedure is done by manipulating the *url*. In this sense, Figure 2 shows the BRAPCI Explorer interface, where the sidebar replicates the search to be scraped from BRAPCI.

---

<sup>3</sup>  $IC = \frac{\sum_{i=1}^n X_i}{n}$

Figure 2. BRAPCI Explorer interface.



Source: By the authors.

The sidebar is made up of text input fields (search field), two selectors (start year and end year) and a button (Search). All the fields have default values in order to assist the user in the search process. This setting allows BRAPCI Explorer to perform the search without interruptions since all the fields will always be filled in. The data is presented in a set of four tabs (tabset): "Co-authorship network" (*Rede de coautorias*), "Most productive authors" (*Autores mais produtivos*), "Most productive sources" (*Fontes mais produtivas*) and "Most productive years" (*Anos mais produtivos*). In addition to the main results, shown in the center of the screen, there is also additional information (total documents found, total authors and different sources, co-authorship index, average number of documents per author and per source and average number of documents per year). There are also interactive features that allow the user to intervene in the co-authorship network, using the author selector or manipulating the network nodes (Figure 4).

Once the BRAPCI Explorer interface had been presented, several searches were carried out to validate the tool, manipulating all the search parameters in the new environment, and compared with the results provided by the database itself (<https://cip.BRAPCI.inf.br/>)<sup>4</sup> to check the accuracy of the new tool. In all searches, BRAPCI Explorer found the same results as those provided by BRAPCI.

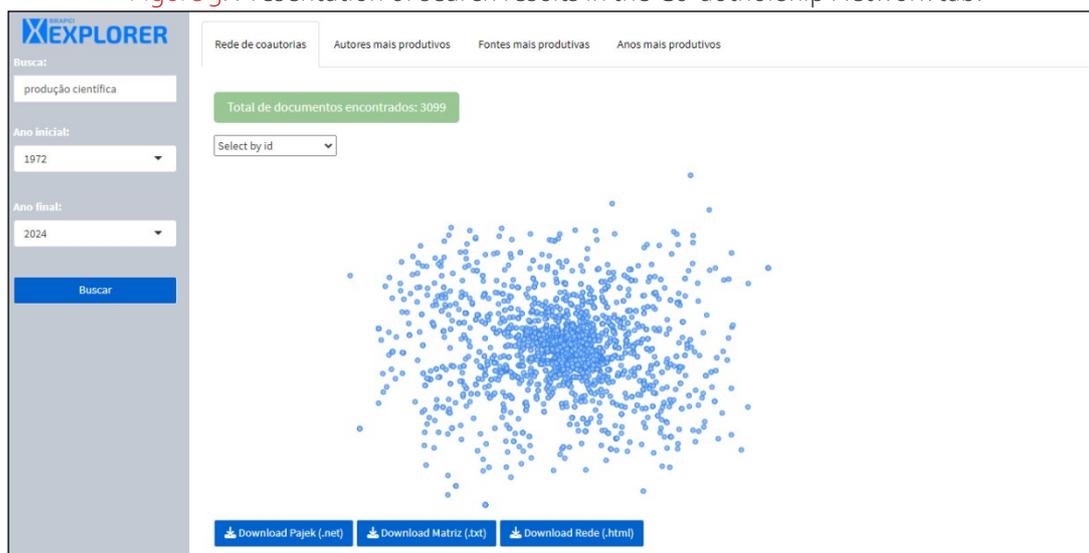
The results of this research include a presentation and description of the functionalities of BRAPCI Explorer, using only one search strategy. The search term "scientific production" between 1972 and 2024 was used (Figures 3, 4, 5, 6). In addition, it is proposed that BRAPCI Explorer be added to the BRAPCI portal (Figures 8 and 9), to demonstrate the feasibility of linking the two environments by means of buttons on the home page and on the search results page. The entire code implementation and testing process was carried out using RStudio software version 4.1.2, on a computer with Windows 10 operating system, Intel Core(TM) i7-8550UCPU @ 1.80GHz-1.99 GHz processor, 200Gb SSD and 8Gb of RAM. Processing was carried out on November 13, 2023.

<sup>4</sup> This address corresponds to the new version of BRAPCI, which uses the same data from the API used to develop BRAPCI Explorer. This address differs in terms of results from the old one: <https://brapci.inf.br>.

### 3 RESULTS

Once the search term "scientific production" has been entered, all the features of BRAPCI Explorer are presented, organized into four tabs. The first tab, shown in Figures 3 and 4, called "Co-authorship network" (*Rede de coautorias*), shows the weighted co-authorship network between all the authors of all the retrieved productions. Along with the network, the green text box shows the number of documents found by the search entered, in the example represented by 3099 documents.

Figure 3. Presentation of search results in the Co-authorship Network tab.



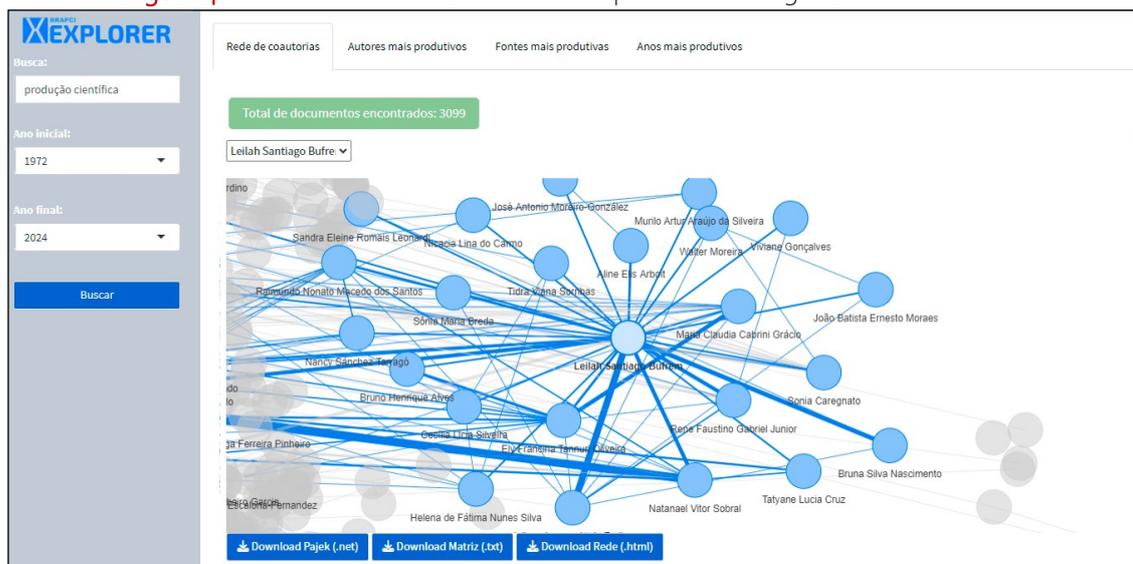
Source: By the authors.

| 9

In Figure 3, each node represents an author and the edges represent the co-authorship relationships established between the authors, where the variation in the thickness of the edges shows the number of documents produced jointly by the authors. The automatic generation of co-authorship networks in BRAPCI Explorer demonstrates an innovative and unprecedented feature in scientific databases, since traditionally co-authorship analyses are carried out after the user exports the search data and processes it in external software. Examples of these platforms are VosViewer or Bibliometrix, bibliometric software capable of reading and generating co-authorship networks from a document exported from databases such as Web of Science, Scopus, Dimensions, Pubmed, among others.

In addition, if the user prefers to view this network in network visualization software, BRAPCI Explorer offers the option of downloading the network in .net format. This extension is certainly one of the most versatile network files and can be run (for network visualization) in software such as Pajek, Ucinet, Gephi, VosViewer, among others. In addition to this option, it is also possible to download the network in matrix configuration and in .txt format (tabulated). It is well known that, in very large networks, access to the data matrix is of great value, since it is possible to observe the main relationships established by the nodes in the network. What's more, access to the co-authorship matrix is an unusual result for bibliometric analysis software. Another feature is the option for the user to interact with the network using the zoom and node selection tools, as shown in Figure 4.

Figure 4. Interaction with the co-authorship network using the author selector.



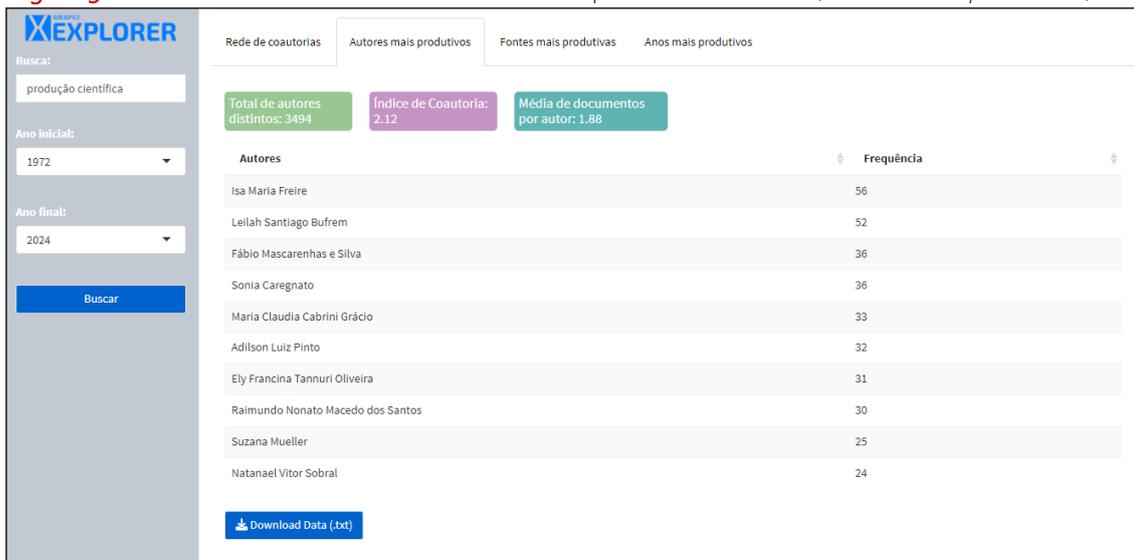
Source: By the authors.

Figure 4 shows the user's interaction when selecting, in the author selection box, and zooming in on the *Leilah Santiago Bufrem* node (name), Leilah Santiago Bufrem referring to the researcher Leilah Santiago Bufrem, creator of BRAPCI. If the user chooses to save the network the way BRAPCI Explorer displays it, it is possible to download it in html format, so that the saved file retains the interaction properties described above. The second tab displays a table of the most productive authors in relation to the search carried out, along with indicators describing the total number of different authors found, the co-authorship index and the average number of documents per author, as shown in Figure 5.

To show the most productive authors, BRAPCI Explorer displays a frequency table containing the 10 most productive authors. It is important to note that the count (frequency) described in this table corresponds to the count of all the documents for which each author has their name on the list of authors, i.e. BRAPCI Explorer does not split the authorship count. In Figure 5, for example, when identifying the researchers Isa Maria Freire and Leilah Santiago Bufrem, it can be seen that they were responsible for 56 and 52 documents respectively, the tool does not distinguish whether this production comes from individual production or in co-authorship with other researchers found by the search.

With regard to the indicators calculated, BRAPCI Explorer identifies the total number of different authors. In this case, 3494 different authors were responsible for producing the 3099 documents. The next indicator refers to the Coauthorship Index, described as the ratio between the number of authors per article and the total number of documents found. This is 2.12, meaning that each publication has an average of 2.12 authors. The other indicator calculates the average number of documents per author, i.e. the arithmetic mean of the number of documents produced per author. In this case, it is 1.88. In other words, on average, each author produced 1.88 documents according to the search carried out.

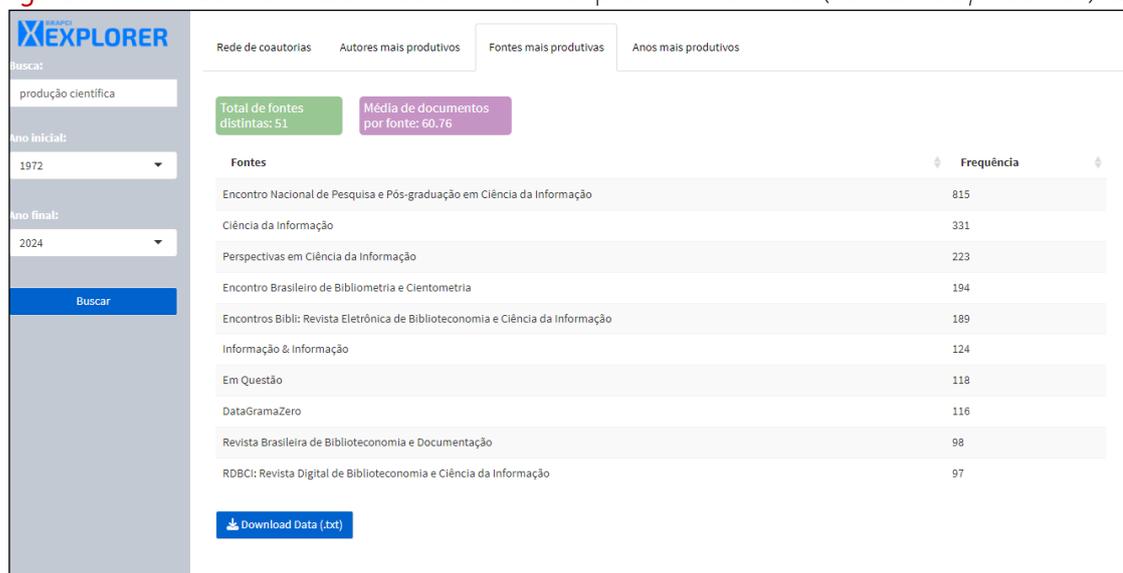
Figure 5. Presentation of search results in the Most productive authors (*Autores mais produtivos*) tab.



Source: By the authors.

It is important to emphasize that even though the tab only displays the 10 results with the highest number of Frequency, the user has the possibility of saving the table containing all the different authors and their respective productions by clicking on "Download Data (.txt)", generating a file entitled Authors.txt (*Autores.txt*) containing a table with the complete data. All indicators are calculated taking into account all results and not just the 10 to be displayed in this tab. Similarly to the author-level indicators, Figure 6 shows the third tab containing the frequency table of the 10 most productive publication sources and the indicators denoting the total number of different sources and the average number of documents per source.

Figure 6. Presentation of search results in the Most productive sources (*Fontes mais produtivas*) tab.



Source: By the authors.

Similarly to the table of most productive authors shown in Figure 5, BRAPCI Explorer counts the number of publications per source (frequency). By carrying out the frequency analysis, it was possible to determine the total number of different sources, i.e. how many unique publication sources the documents were published in. In this case, the 3099 documents were published by 3494 authors in 51 different sources.

The other indicator on the tab is the average number of documents per source. This indicator measures the arithmetic mean of production according to publication source, and for this search we have 60.76. In other words, each source published an average of 60.76 documents relating to the expression "scientific production". To access all the information relating to both indicators, the user can save the complete table containing all the sources and their respective outputs by clicking on "Download Data (.txt)" which will generate a file named Publication Sources.txt (*Fontes de Publicação.txt*).

The "Most productive years" (*Anos mais produtivos*) tab, shown in Figure 7, refers to the temporal analysis of publications by year. Unlike the other tabs, we opted to visualize the data using a bar chart showing production over the last 15 years.

Figure 7. Presentation of search results in the Most Productive Years (*Anos mais produtivos*) tab.



Source: By the authors.

It was decided to export the complete data in .txt format instead of a graph (in image format) because, depending on the size of the search, visualization could be impaired. In addition, this tab has an indicator that calculates the average number of documents per year. In this case, the value 61.98 was obtained, i.e. an average production of 61.98 articles per year was found.

Now that BRAPCI Explorer has been presented, we propose inserting the new tool alongside the rest of BRAPCI's resources. The database has some resources which are presented at the bottom of <https://cip.BRAPCI.inf.br>. At the time of publication of this research, some of these resources are in the development and updating phase (<https://cip.BRAPCI.inf.br/books/about>), as mentioned by Silva (2023). Therefore, two proposals for linking the environments were suggested, the first of which is shown in Figure 8.

Figure 8. Proposed insertion of the BRAPCI Explorer access link on the BRAPCI homepage.



Source: By the authors.

In this proposal, the BRAPCI Explorer is presented together with the new functionalities, or environments, still under development by BRAPCI. In Figure 8, the card is clickable and is in the focus state, i.e. selected by the user, unlike the other cards. The suggestion to insert a link on the base's home page is justified by the fact that it would help popularize the base's complementary tools, so that users can get to know the existing search and analysis options. It is also suggested that a link be added to the BRAPCI search results page, as shown in Figure 9.

| 13

Figure 9. Proposed insertion of the BRAPCI Explorer access link on the BRAPCI search results page.



Source: By the authors.

In this scenario, the user performed the search on the BRAPCI home screen and clicked on the "Search" button, being directed to the search results screen. The suggestion here is to present a link in the form of a button, similar to the "Analyze Results" buttons on Web of Science and Scopus. As the user has already carried out the search in the BRAPCI environment, ideally they should be redirected to BRAPCI Explorer with the search results already displayed, however, due to the fact that BRAPCI Explorer is still a database-independent environment, the redirect after the user clicks on the "BRAPCI Explorer" button on the BRAPCI results page would take them to the BRAPCI Explorer interface (Figure 2). This way, they only have to enter the search term again to access the results in the BRAPCI Explorer environment. It is thought that this proposal (ideal model) would be the most appropriate, given that it is the practice adopted by other databases.

## 4 CONCLUSION

Scientific databases are recognized as important for the preservation and dissemination of scholarship by storing, indexing, representing, and making available scientific information. The emergence of the Web of Science and Scopus databases in the late 1990s and early 2000s were important milestones in the development of these environments. In the context of Brazilian information science, BRAPCI, launched in 2010, can be mentioned as the most comprehensive repository of periodical scientific production in the country. However, compared to Web of Science and Scopus, BRAPCI lacks bibliometric resources beyond production indicators.

Thus, it is assumed that BRAPCI brings together a set of relevant scientific information that can be analyzed bibliometrically. In this sense, this research proposed a new, innovative and interactive environment, called BRAPCI Explorer, aimed at bibliometric analysis, based on data extraction through the API of the database, a method of great value in the process of collecting and structuring data, favoring different analyses. Using this extraction technique, it was possible to collect the data provided by the database and present them in BRAPCI Explorer.

This new environment was developed in the R programming language, using the shiny library to design the interface and the jsonlite library for the data collection process. The interface was developed by organizing the text input fields, the field and period selectors, and the search button in a sidebar structure. The presentation of the data, based on the search performed in the sidebar, took place in a tab structure where the user could navigate between the options "Co-authorship network" (*Rede de coautorias*), "Most productive authors" (*Autores mais produtivos*), "Most productive sources" (*Fontes mais produtivas*) and "Most productive years" (*Anos mais produtivos*).

BRAPCI Explorer was designed and tested to evaluate the use of the new web environment. Preliminarily, the tests were in line with BRAPCI's own results, i.e. the results extracted (via API) by BRAPCI Explorer were consistent with those provided by the new BRAPCI website. In addition, the tests demonstrated the possibilities of data visualization and analysis via BRAPCI Explorer. One of the innovative features is the automatic generation of the co-authorship network, present in the first tab of the tool, which facilitates the bibliometric analysis, since the user does not need to download the data and use other tools, since everything is available in the BRAPCI Explorer interface. It also offers the possibility of saving the data if the user needs them in different formats: .net, matrix and html. The interactive features include the management of the nodes of the co-authorship network and the identification of the nodes using the author selector.

The second, third and fourth tabs show the most productive authors, sources and publication years, along with bibliometric indicators for these units of analysis. In the first two cases, the interface shows only the top 10 results, organized in a frequency table, and in the

case of the most productive publication years, it was decided to present the frequency data in bar graph format. The user can also access the full data using the download buttons.

For authors, there are indicators for the total number of different authors, the co-authorship index, and the average number of documents per author. For publication sources, the new environment provides indicators for different sources and the average number of documents produced per publication source. And on the last tab, the average number of documents per year of publication is calculated. In this way, BRAPCI innovates by making the indicators automatically available to its users.

When testing the new environment, a search using the term "scientific production" found 3099 different documents produced by 3494 different authors in 51 different publication sources. BRAPCI Explorer was able to generate the co-authorship network containing all these authors, as well as the co-authorship matrix and the .net network file, create all the tables and frequency graphs, and calculate the indicators. It's important to note that this process took approximately 1 minute and 40 seconds, which means that 6,102,271 interactions took place between all the authors identified when building the co-authorship network. Although this process may seem slow, BRAPCI Explorer provides all results ready for use and/or export via download. After the tests, to demonstrate the viability of BRAPCI Explorer, two ways of linking the environments were proposed: a button on the home page and another on the BRAPCI search results page, thus demonstrating that linking the environments would not require much programming effort.

One of the limitations is that for the web application version, the tool will have problems processing large datasets, as its current hosting on the shinyapps.io server has an instance size limit of 1 gigabyte. For scenarios where the search exceeds the processing limit, we suggest using the BRAPCI explorer directly through the R software, as well as adjusting the memory limit (`memory.size` and `memory.limit` functions). It should be noted that the tests were carried out with the configurations mentioned in the methodology section, so the computational overload may vary according to the scenario of each device. However, the possibility of using BRAPCI Explorer online makes it possible to use it in different Internet browsers such as Google Chrome, Microsoft Edge, Mozilla Firefox, Safari, among others, without the need for the user to install R.

It should also be noted that BRAPCI Explorer was initially designed to perform a web scraping process; however, after accessing the database API, the environment was recoded. Since the data connection via API is faster, it is understood that this change made the new environment more computationally viable. At the time of publication, it is not yet possible to add more filters to BRAPCI Explorer searches, as presented on the BRAPCI website: search field that limits where the query will be searched (title, abstract, keywords, authors, all fields) and selection of information sources (journals or events). It is hoped that this increment will be present in the next updates of BRAPCI Explorer, completely replicating the database search. However, it must be stressed that the database is currently being updated.

In view of the above, it can be concluded that BRAPCI Explorer contributes to a better visualization of the data made available by BRAPCI, because it presents improved visualizations, in network format, graphs and tables, as well as bibliometric indicators, without the user having to leave the web environment and without the need for programming knowledge. Finally, the interface is intuitive and has a visual identity close to that of BRAPCI in order to make the user feel confident in navigating between the environments. The next steps and for future research are to keep BRAPCI Explorer up and running and carry out updates to the interface, as well as usability tests.

## REFERENCES

- 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>.
- BRASIL. API de dados. **Portal da transparência – Controladoria-Geral da União**. 2023. Disponível em: <https://portaldatransparencia.gov.br/api-de-dados>. Access on: 10 nov. 2023.
- BRAPCI, Sobre - Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). Available at: <https://brapci.inf.br/index.php/res/about>. Access on: 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. Available at: <bit.ly/3EPufU0>. Access on: 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**. Available at: <https://clarivate.com/the-institute-for-scientific-information/history-of-isi/>. Access on: 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. Available at: <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](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. Available at: <https://lume.ufrgs.br/handle/10183/257995>. Access on: 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>.