



Collection and Integration of Patent Data for Analysis and Validation of Brazilian Technical Production

Raulivan Rodrigo da Silva¹ · Thiago Magela Rodrigues Dias¹ · Washington Luís Ribeiro de Carvalho Segundo²

Accepted: 26 July 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

The main objective of this work is to present an overview of Brazilian technical production from the analysis of patents registered in national repositories such as the National Institute of Industrial Property (INPI), as well as in international repositories such as Espacenet. Initially, all the INPI patent identifiers are collected so that later, all the data of these patents are collected in the Espacenet repository. In this way, a large local data repository is characterized, as standing out for having consistent and structured data. This strategy makes it possible to analyze the entire dataset with the adoption of several metrics, making it possible to present a broad and unprecedented portrait of Brazilian technical production. With the dataset retrieved, it is still possible to validate other repositories or datasets, such as the curriculum base of the Lattes Platform, in which their curricula in general do not have all the information about the patents duly registered. However, it is concluded that the collection and integration of data extracted from patent documents and stored in local repositories, significantly enable patentometric analyses that could not be performed in distributed repositories.

Keywords Patents · INPI · Lattes platform · Espacenet · Patentometry

1 Introduction

The 21st century has been fertile ground for the creation of technological structures, and more than ever, the rapid evolution of these technologies has been visible. Daily, new devices, applications, and digital media permeate the market, bringing better versions of resources and/or functionalities that we knew until then or presenting new solutions. As a result, organizations in this market are committed to constantly monitoring their activities and the viability of their products and services offered, being necessary to implement innovations that retain or increase the customer

base [1]. Various inventions and innovations generated originate from research initiated in public and private educational institutions, which has been increasing over the years, making them a great pole of national innovation. This factor stimulated the creation of laws and norms that aim to protect the intellectual property generated in universities and also centers that aim to assist in the process of creating patents and enabling the relationship between the academic and marketing sectors, encouraging educational institutions and researchers to increasingly lean towards the filing of patents [2].

It is possible to find several studies that are based on the number of patents deposited and the number of patents granted, using them as parameters to determine the volume of technological innovation in a given country or institution, such as a work entitled “The state of the art of patentometry in journals of information science” by the authors [3] and the work “Mapping of intellectual property deposits at the INPI by students of the postgraduate program in intellectual property science at the federal university of Sergipe” written by [4]. However, this is not enough to understand the whole scenario. As [5] states, it is not possible to say that all patents contribute to the growth of science and technology, or even if they bear fruit for their holders and inventors.

✉ Thiago Magela Rodrigues Dias
thiagomagela@cefetmg.br

Raulivan Rodrigo da Silva
raulivan@cefetmg.br

Washington Luis Ribeiro de Carvalho Segundo
washingtonsegundo@ibict.br

¹ Federal Center for Technological Education of Minas Gerais (CEFET-MG), Minas Gerais, Brazil

² Brazilian Institute of Information in Science and Technology (IBICT), Brasília, Brazil

Therefore, as with scientific productions, in the context of technical production, there are also repositories of patent records, such as the pePI (Industrial Property Research) maintained by the Brazilian patent management body INPI (National Institute of Intellectual Property). As in Brazil, each country has its own body responsible for managing the filing and granting of patents, as well as making them available for consultation. In addition, there are international repositories for patent registration, some of which, such as Espacenet, are of recognized relevance. Espacenet, which makes it possible to consult patents from approximately 70 countries in a single repository, stands out because of the amount of data available.

In view of this, this work aims to broaden the understanding of nationally constructed patenting activities, seeking to evaluate the main actors, the hidden collaboration networks, and the result that these inflict on the evolution of science. Based on what is presented, a technological tool is proposed for the collection and treatment of national patent data available in the Espacenet repository, aiming at the construction of a local repository that allows all objective analyzes.

2 Theoretical Basis

The patent is one of the branches of intellectual protection, a patent is considered a formal document, which is issued by a public office, by which property rights and exclusive use are conferred and recognized for an invention widely described [6]. This is a privilege granted to creators, whether individuals or legal entities, holders of the right to invent products, manufacturing processes, or even the improvement of an existing product or process. The Inventor provides society with a new good, that is, knowledge, disseminating the technical information that will guide the company to reach its goal. Society, in turn, offers the inventor exclusivity of exploration, that is, production and commercialization of the object of his invention, for a certain interval of time.

Therefore, innovations protected by patents have potential commercial value, for this reason, there is specific legislation to guarantee their protection. The value of a patent document is territorial, that is, each country has its legislation and its own office for examining and granting patents. In Brazil, the Intellectual Property Code, law 9279/96, came into effect on May 14, 1996, which regulates rights and obligations relating to industrial property [7]. The body in charge of its application is the National Institute of Industrial Property (INPI), created in 1970, today a federal autarchy linked to the Ministry of Development, Industry and Foreign Trade (MDIC).

In Brazil, patents can be divided into three types [8], namely:

- Patent of Invention (IP): Products or processes that meet the requirements of inventive activity, novelty, and industrial application. Its validity is 20 years from the filing date.
- Utility Model Patent (MU): Object of practical use, or part thereof, susceptible of industrial application, which presents a new form or arrangement, involving an inventive act, which results in functional improvement in its use or manufacture. Its validity is 15 years from the filing date.
- Invention Addition Certificate: Not quite a type, but an addendum to an existing patent. Improvement or development introduced in the object of the invention, even if devoid of inventive activity, but still within the same inventive concept. The certificate will be an accessory to the patent and with the same final date of validity as this.

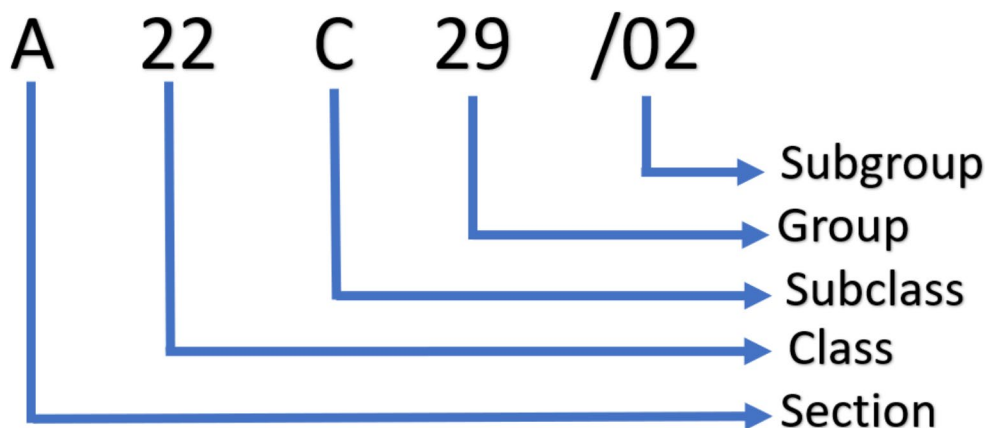
Granted patents are classified according to their nature following the International Patent Classification (IPC) standard. IPC is the patent classification system used in Brazil and over 90 other countries. Several countries mobilized to develop the classification system based on the international treaty signed in 1954, the European Convention for the International Classification of Patents for Inventions. Sponsored by WIPO, the Strasbourg Agreement on the International Patent Classification was signed in 1971, and entered into force in 1975 [9].

IPC aims to allow the search and retrieval of documents referring to patents; organize them to facilitate access to technological information; help in the investigation of the state of the art in certain fields of technology; base for the elaboration of statistics on industrial property. It is worth mentioning that attention can receive more than one classification, as many as necessary, if there is no specific section that best classifies the invention, it will be inserted in the most appropriate area.

IPC is a hierarchical classification which divides technological knowledge into eight large areas called sections. At the highest level, there are sections, represented by the letters A to H; each section is divided into classes, each class into subclasses, and each subclass into groups, which may contain subgroups. The eight sections, the first hierarchical level, are: A – Human needs; B – Processing operations, transport; C – Chemistry; Metallurgy; D – Textiles; Paper; E – Fixed constructions; F – Mechanical engineering; Lighting; Heating; Weapons; Explosions; G – Physics; H – Electricity.

The IPC classification received by a patent comprises a combination of letters and numbers, such as the classification A22C29/02 [10]. Figure 1 presents the details of each element that make up the classification:

Fig. 1 Details of a patent classification. Details: Section A: Human Needs; Class 22: Killing of animals; meat processing; processing poultry, or fish; Subclass C: Processing of meat, poultry or fish; Group 29: Processing of crustaceans, e.g. ex. oysters, and lobsters; e Subgroup 02: Shrimp, lobster, or similar processing



A patent deposited in Brazil is only valid in Brazilian territory; if you want to protect the invention in another country, it is necessary to deposit the patent in the desired country to defend it, thus constituting a family of patents. In general, a family of patents consists of a set of patents filed in different countries but with the same purpose, that is, to protect the same invention; the priority right of the first filing is extended to the others in the different countries in which they were filed deposited. According to the INPI patent family, it is a set of patents and applications, whether national or international, directly related to the claim of ownership or sharing of the same international deposit.

2.1 Repositories for Querying Patent Documents

To file a patent application, one of the requirements is to know the state of the art, which, by definition, is everything that is accessible to the public before the filing date of the patent application, which can be done using a textual description, making use of any type of device, whether national or international, including being disclosed by the inventor himself, that is, everything that already exists about the claimed application in the world. Therefore, all technical information contained in patent documents is available in freely accessible databases via the Internet. Among the various existing databases in the world, [11] presents seven databases that provide free consultation of patent documents on the internet. To know:

- PePI (Research in Industrial Property)¹ also known as Bus-ca Web, is maintained by the INPI agency, to search for patents deposited in Brazil, as well as access to patent applications that are not in secrecy, patents already granted and patents archived. Through anonymous access in which it is possible to view basic patent information, such as title, abstract, and authors, among others.

¹ <https://busca.inpi.gov.br/pePI/>.

To view digitized documents, it is necessary to register, free of charge, to grant access to such information;

- Latipat² for searches in public patent documents in Latin America, such as Brazil, Argentina, Mexico, Cuba, Chile, Uruguay, and Peru, among others, and even the only European country, Spain;
- Espacenet³, maintained by the European Patent Office (EPO), is the best-known and most widely used for searching patent documents. It is a database containing data on more than 120 million patent documents from around the world.
- Patentscope⁴, which is WIPO's search tool, offers free access to documents from over 40 countries, such as China, the United States, Australia, Japan, Mexico, and Brazil. Including deposits made via PCT (Patent Cooperation Treaty);
- USPTO (United States Patent and Trademark Office)⁵ for searches on patents deposited in the United States of America (USA), is the page of the North American patent office;
- SIPO (State Intellectual Property Office of China)⁶ for searches on patents filed in the People's Republic of China;
- JPO (Japan Patent Office)⁷ for searches on patents deposited in Japanese territory and even on Google patents.

However, it is possible to make use of paid tools, which allow searches in several repositories in a single place in a transparent way for the user, such as, for example, the Derwent Innovation Index – DII46 base, by Thomson Reuters Scientific, accessible by institutions participating in the

² <https://lp.espacenet.com/>.

³ <https://worldwide.espacenet.com/>.

⁴ <https://patentscope.wipo.int/search/en/search.jsf>.

⁵ <https://www.uspto.gov/>.

⁶ <https://english.cnipa.gov.cn/>.

⁷ <https://www.jpo.go.jp/>.

Journal Portal of the Coordination for the Improvement of Higher Education Personnel (Capes/MEC).

2.2 Patentometry

Several works have used the term “Patentometry” to describe the metric studies of patents [12, 13]. Patentometrics is part of the group that constitutes bibliometrics because, for the analysis of patent documents, adaptations of indicators applied to the production of other types of documents are used [14]. The authors [15] define patentometrics as the science of analyzing information contained in patent documents seeking to identify possible relationships and trends, through strategies such as:

- Patent intelligence: “Use of patent information to identify the technical resources of an ICT (Science and Technology Institutions) or NIT (Technological Innovation Center) and the use of this intelligence to develop technological strategic planning” [15];
- Patent mapping: “Uses patent data to create graphical representations that show a technological overview of an area of knowledge or an invention” [15];
- Analysis of patent citations: “Assists in valuing a technology and identifying partners and licensing new technological discoveries” [15].

The concept of patentometry was used in 1995 by Narin [16], however, the term was established by Sánchez [17] who highlights that Patentometry is an approach that combines certified knowledge and participation in industrial innovation. Already [18] define patentometrics as the science of analyzing a large volume of data on intellectual property, aiming to discover relationships, trends, and patterns. Making use of mathematics, statistics, computer programming, and operational research, to obtain information that can assist in decision making.

3 Related Works

Studies based on information obtained through patent data have different approaches and have quantitative, technological, innovation management, and social, among others. Because it is a topic with a wide variety of works published, a small group was selected, in which the main theme is patentometry. The criteria defined for the selection of these studies were: publications that are available on the web, in Portuguese or English, published in the last 10 years, and whose objective is to analyze patent documents of any nature. The research was carried out in the year 2021 between April and July, using the following search terms:

in Portuguese “patentometria” and English “patentometry”. Using the following repositories: Arca (Fiocruz), Digital Library of Theses and Dissertations, Google Academic, Microsoft Academic, CAPES/MEC Journal Portal, and SciELO.

The authors address the dimensions of intellectual property in their work, conceptualizing and highlighting patents as a rich source of data for research in the field of Human and Social Sciences, particularly in Information and Communication Science, as well as the importance of using them in the transfer of technology between countries and organizations, since the information contained in patent documents, in rare cases, is reapplied in other sources of information. In their work, the authors do not carry out a specific analysis of patent data, however, they emphasize the importance of using the information contained in patent documents and present several strategies for data analysis, presenting directions for researchers interested in the subject [19].

In the work of the authors, they present in their work an overview of patentometry in international journals of information science. They begin their works citing bibliometrics and scientometrics as the most relevant methodologies in the studies of Informetrics, as they cover a large area of study. Complementing, they emphasize that patentometry is a tool of competitive intelligence in science and innovation, providing indicators as a substantial basis for decision-making. Since patentometry is characterized by establishing indicators that seek to identify innovation activities and technologies, through information extracted from patent documents. To establish an overview, the authors carried out a study at the international level, verifying the state of the art of research in patentometry in the area of information science. Per research in journals with Qualis A1 to B2 concepts from 1998 to 2013, for terms referring to patentometry, such as patent, industrial property, innovation patent, and analysis patent. They conclude the work by stating that the analysis of patents is an important tool for competitiveness in research, which allows the identification and conversion of scientific knowledge into technological knowledge [20].

The author presents in his master’s dissertation, an analysis of patents granted by Brazilian Public Universities to identify factors and/or characteristics that encourage the filing, and consequently, the granting of patents. The collection of patent data for analysis was performed in the database of the INPI, in which it was possible to collect patent data from 29 universities, out of the total of 106 universities surveyed, 65.2% correspond to universities in the southeastern region of parents. Based on the IPC (International Patent Classification) of the collected patents, it was possible to observe that the area of knowledge with the highest number of patents granted is “Chemistry and Metallurgy”, in which the researchers who patent the most were professors with a

background in Chemistry or Mechanical Engineering. The author also emphasizes that there is a strong presence of the public sector, since for the most part, coownership is constituted by funding and research institutions. In conclusion, the authors highlight the importance of Technological Innovation Centers (NITs) in public institutions as actors that contribute to implanting the culture of patenting in the country, which consequently, contributes in an indispensable way to the technological evolution of the country [21].

In this work, the authors propose a methodology to assess the technology, through scientometric and patentometry techniques, thus serving as a decision-making support tool for organizations regarding the introduction of new technologies. Through a model composed of 9 steps, (1) Definition of terms, where the words or expressions that you want to obtain greater information; (2) Validation of terms by experts, where defined terms are reviewed and approved by experts in the related field; (3) Definition of the bases to be used, through previously established criteria; (4) search for terms in the defined bases; (5) Extract the found records to form a local database, to enable the analyses; (6) Perform data cleaning, correction of possible errors typing, duplication and standardization; (7) Generate lists, graphs, and tables; (8) Analyze the information generated; (9) Conclusions and results. The authors further argue that the proposal makes it possible to provide decision-makers with scientific databases that will serve as a basis for decision-making, as evidence of offers of knowledge for the development of technology, as well as its demand [22].

Patentometry is seen as an indispensable tool for proposing technologies for the chemical industry, since the prospection based on quantitative and qualitative methods, are important to enable the conceptualization of plausible proposals in the state of the art. Next, they explore some indicators presenting the process of construction of its relevance in the technological panorama. You authors conclude by mentioning that patentometry is a little-explored area in Brazil despite its competitive value [23].

Several other works have used data recorded in patent documents to analyze technical production [24, 25]. Based on the above, the works present analyzes of specific sets of patents, this work aims to contribute by presenting a collection strategy that enables analysis of broader data sets.

4 Methodology

This article is a case study, that is, an empirical study that investigates a particular phenomenon, within a context in which there are still gaps in the literature, as stated [26].

A priori, information was collected regarding patent documents deposited at the INPI in the period from 01/01/1900

to 12/31/2020. With the data in hand, the patent data was consulted in the Espacenet repository, using the patent filing number collected at the INPI. This set of data extracted at INPI as well as at Espacenet is the set of data analyzed in this study, a study of great relevance, mainly due to the techniques implemented, as well as the breadth and consistency of the data collected.

After collecting patent data, using the LattesDataXplorer framework developed by Dias [27], data were collected from the curricula registered on the CNPq Lattes Platform⁸ (The Lattes Platform is a curriculum base maintained by the National Council for Scientific and Technological Development (CNPq) with the objective of integrating academic curriculum bases from public and private institutions into a single platform), which have registration information and/or participation in the patent filing. With this data, algorithms were developed to equalize the entire database obtained from the curricula, and also to remove information from users without registered technical production or a valid number of patent registrations. The objective of this strategy is to verify the consistency of the patent data registered in the curricula registered on the Lattes Platform, as well as to enable analyses that can consider various information from the proponents that are not in the patent records, but informed in their CVs, such as for example, data on academic training, areas of activity and scientific production.

Several other studies used the framework LattesDataXplorer to extract curricular data from the Lattes Platform [28–31].

4.1 Data Acquisition

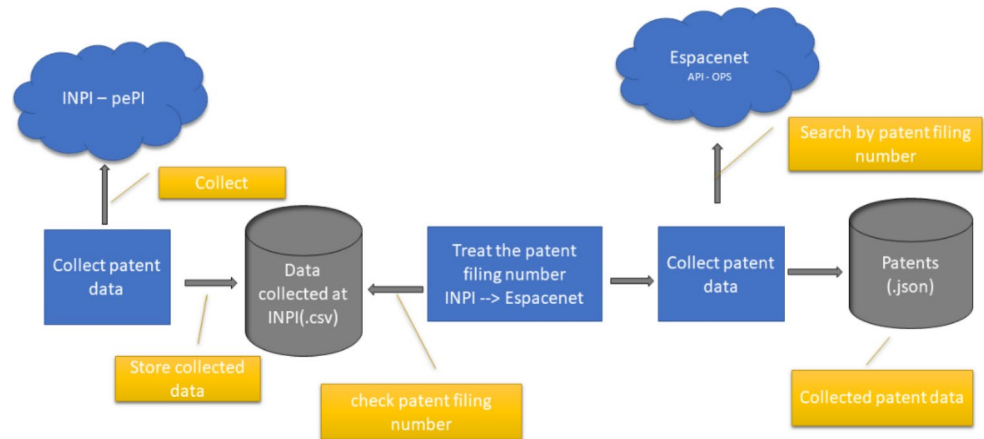
The patent data acquisition process was divided into two stages, (1) initially the data collection at the INPI and processing of the patent filing numbers, and later, (2) the validation and collection of patent data from the patents extracted at the Espacenet repository. Figure 2 presents the schema created for the data collection process.

To collect patent data at the INPI, the patent search tool pePI (Research in Industrial Property) maintained by the INPI was used, where it is possible to consult patent documents by informing login and password or by anonymous access. What differs between the two forms of identification is that choosing to inform the login and password will allow access to more services, such as the availability of documents in PDF format, among others, but to achieve the objective of this work, anonymous access is the enough, therefore, the same was used.

When entering in the search field “(22) Deposit Date”, the start date “01/01/1900” and end date “12/31/2020” and

⁸ <https://lattes.cnpq.br/>.

Fig. 2 Overview of the data collection process



selecting the “search” option, the system returns a page with the list of 862,726 patents spread over 8,627 pages displaying 100 records per page. To optimize data collection an algorithm was proposed to enable a computational process to automate the collection, consisting of 5 steps:

1. Perform anonymous login to retrieve the credentials needed to perform the search;
2. Access the advanced search, informing the credentials obtained in the previous step;
3. On the advanced search screen, enter the start date 01/01/1900 and the end date 12/31/2020 in the field “(22) Deposit Date” and trigger the search event;
4. Scroll through the entire patent listing displayed on the result page.
 - a. For each patent, access the details page:
 - i. Parse the HTML (HyperText Markup Language) content of the detail page and retrieve the information: “Order number”, “Deposit date”, “Publication date”, “Depositor Title”, “Inventor”, and “ICP Classification”.
 - ii. Store the retrieved information in a CSV (Comma-separated-values) file;
 - iii. Back to patent listing;
5. Repeat step 4 for all Search Results Pages

This entire strategy was coded using the Python programming language through web scraping and web crawler techniques.

During the tests of the developed algorithm, it was possible to identify a limitation in this approach; due to the large volume of data; for reasons of platform security, the credentials expire after a specific time. To bypass this limitation, monthly periods were used for the “Deposit Date” filter. Therefore, storing the data in CSV files, one file for

each year, the collection was performed between April and June 2020.

With the data collection at the INPI completed, the next step was to identify each patent collected at the INPI on Espacenet and extract its available data. Only the set of patents that are identified on Espacenet will be considered due to the completeness and consistency of the data.

Espacenet is a worldwide intelligent search service that provides free access to information on inventions and technical developments from 1782 to the present. Its query interface is simple and intuitive, making it accessible even to inexperienced users, currently containing data from more than 120 million patent documents from all over the world [32]. The Platform offers intelligent search features, in which it is possible to enter the desired term where it is searched in several fields of the patent, being able to enter up to 10 words separated by space. The service was designed to be used by humans, not allowing automatic queries or batch retrieval; when this is necessary, the use of OPS (Open Patent Services) is recommended.

OPS is a web service that provides access to data stored in the EPO (European Patent Office) database through web services using RESTful architecture. Using XML (eXtensible Markup Language) and JSON (JavaScript Object Notation) standards to format the response data to requests according to the parameterization. Consequently, developing self-extracting applications and robots to download large volumes of data becomes viable.

The retrieval of data referring to each patent is made possible using the patent search available at OPS, using the patent application number as a selection criterion. The order number is important for patent identification at INPI and Espacenet, as each patent has its unique filing number. The composition of the patent filing application number at the INPI has two different formats: one was used for older patents and another is currently adopted. Since January 2, 2012, the new form has been assigned to new patent

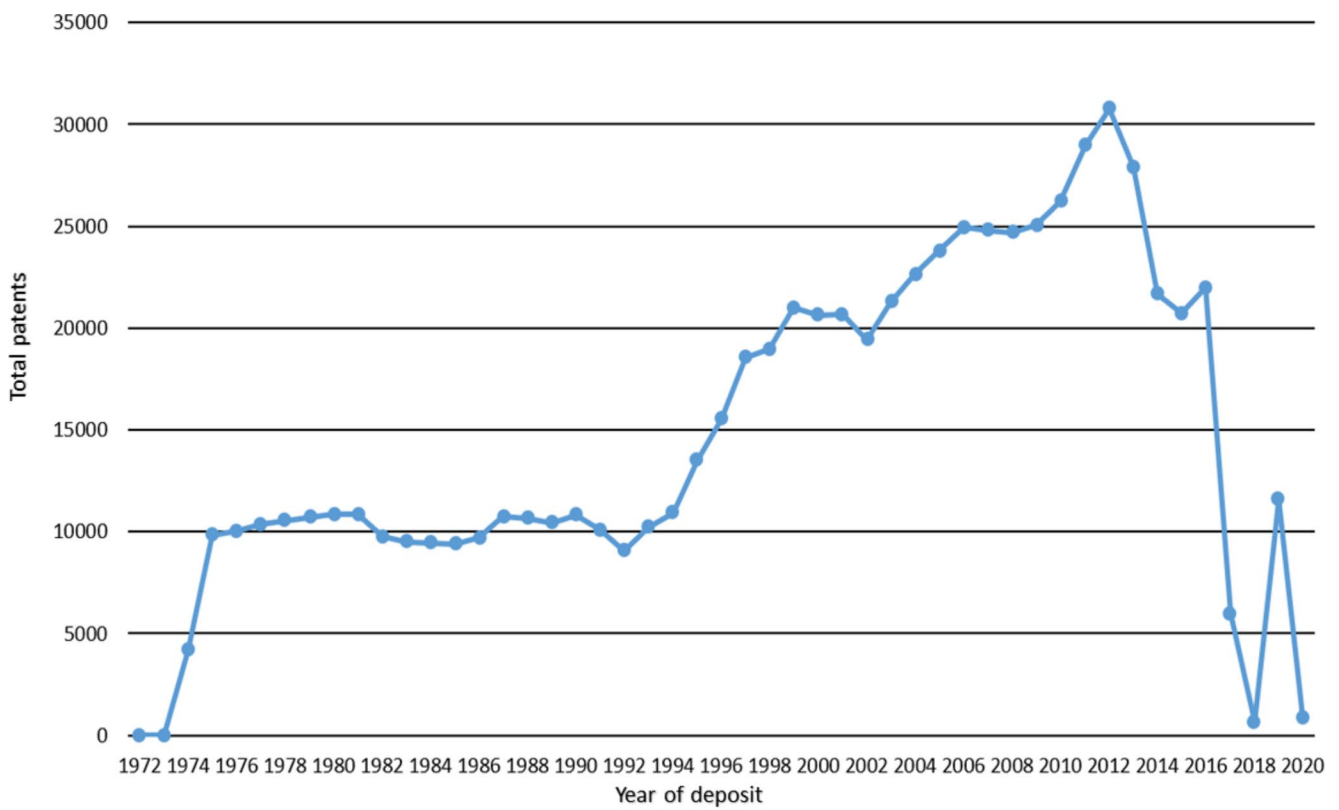
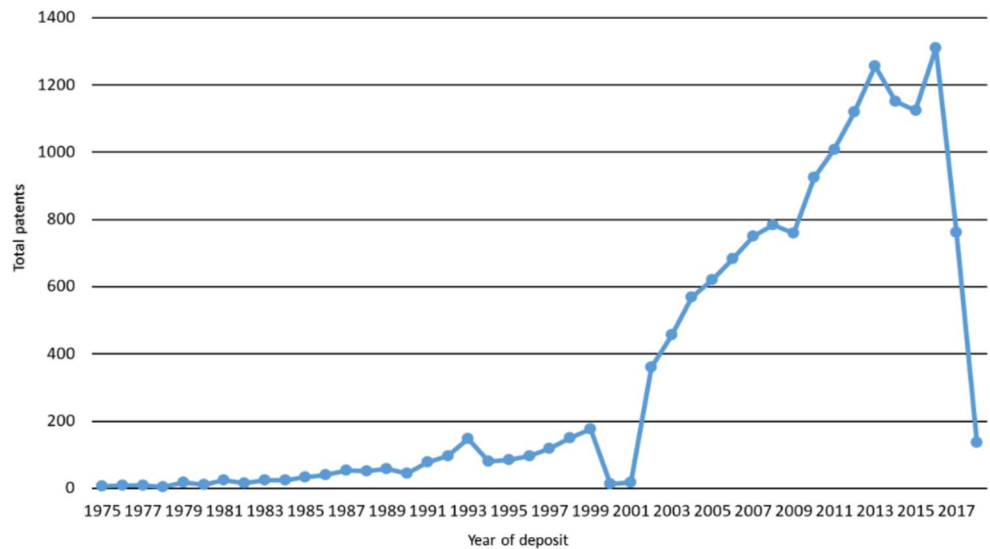


Fig. 3 Time evolution of patent filing per year

Fig. 4 Patent filing in Lattes Platform curricula



applications (invention and utility model), industrial design and geographical indication.

The format assigned to patents filed up to 12/31/2011 is composed of ZZ XXXXXXX-D, where ZZ refers to the nature of the protection, XXXXXXX is an annual serial number consisting of 7 digits, and finally, D, which is the verifying digit.

The new format established aims to meet the INPI's international integration policy, meeting the standards internationally suggested by WIPO St133 published by WIPO (World Intellectual Property Organization). This new format has the following structure BR ZZ AAAA XXXXXX D CP, where BR is the identification of the country, ZZ is the nature of the protection, YYYY year of entry into the INPI, XXXXXX numbering that corresponds to the order of filing

Fig. 5 Training of tenderers

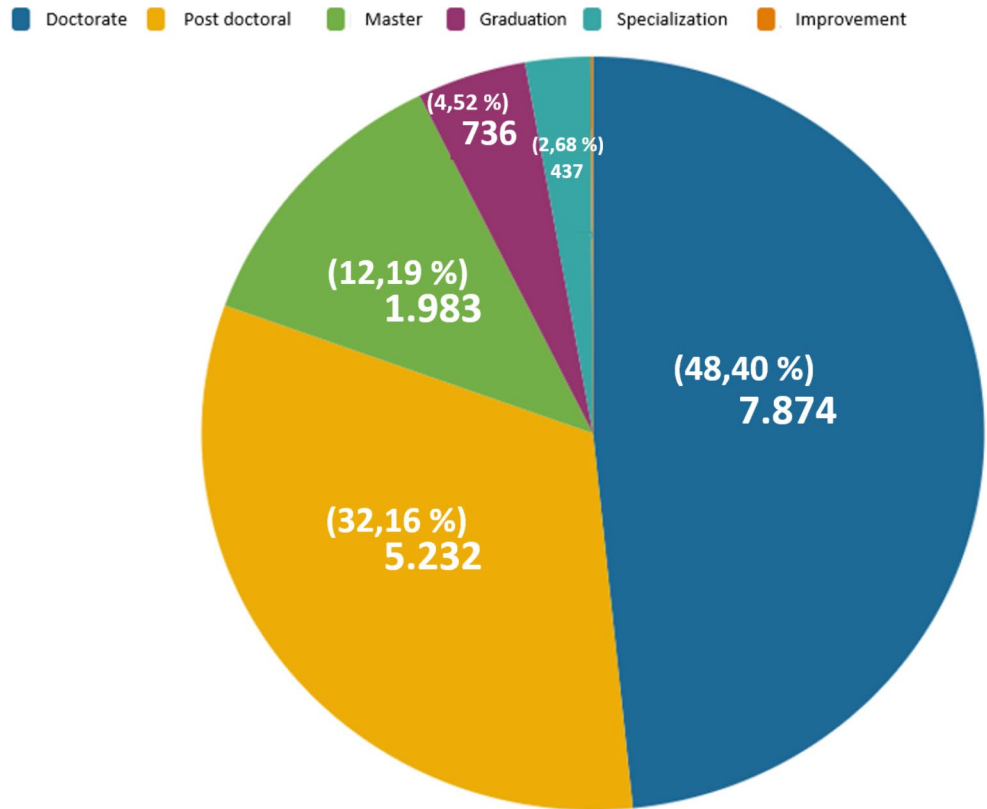
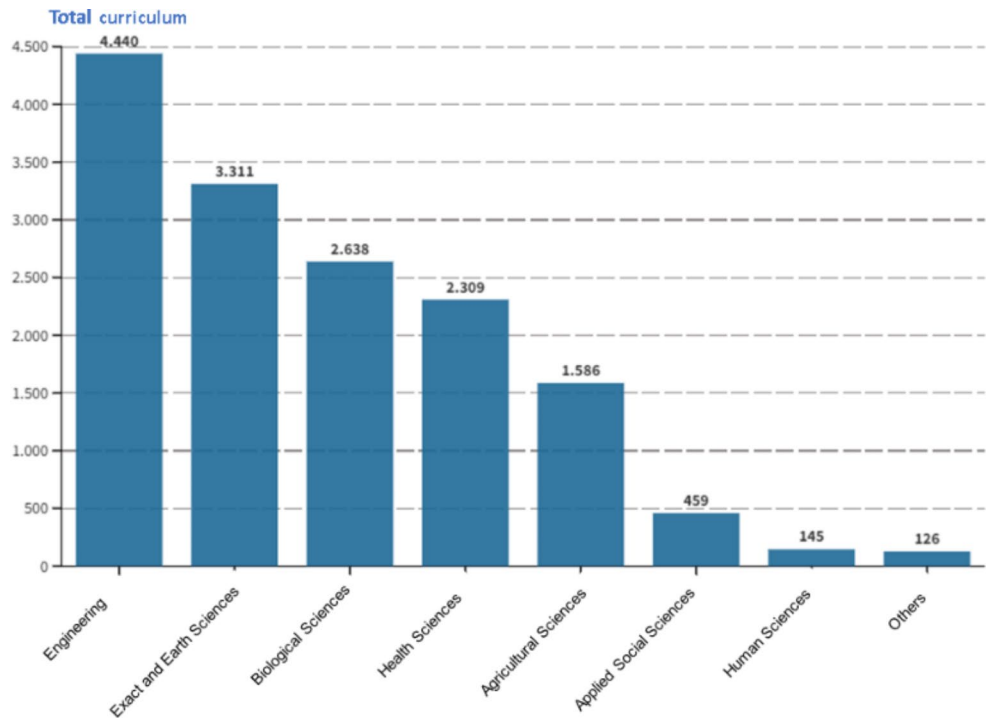


Fig. 6 Large scope of activity for patent applicants

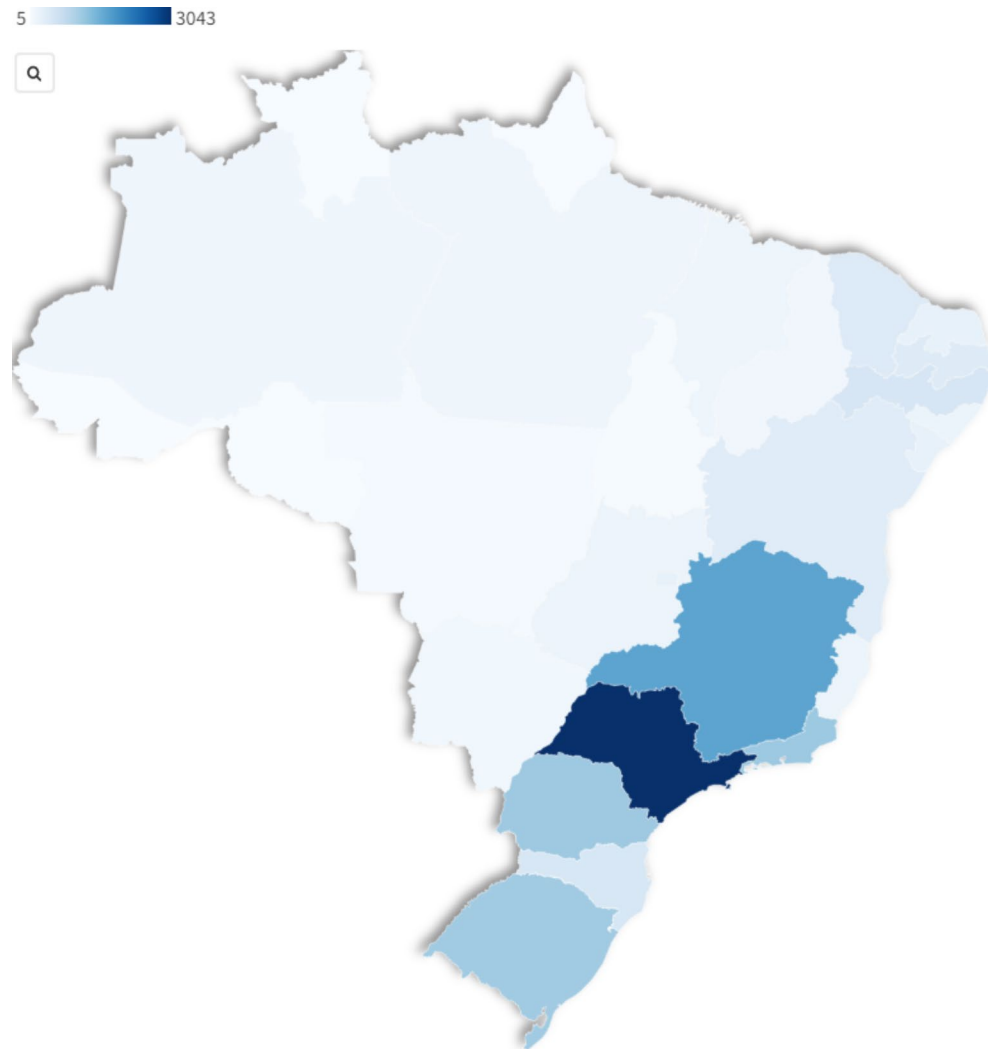


of applications composed of 6 digits, D the check digit and finally CP that corresponds to the publication code, the legal status of the request with the INPI.

Espacenet adopts the international standard WIPO St13 to store information about patents, which implies processing

the filing application number before performing the search on Espacenet.

Based on a set of defined rules, it was developed using the Python programming language. This algorithm goes through all the patents collected at the INPI and applies all

Fig. 7 Distribution of patent applicants by state

Source: IGBE

the prescribed rules, storing the results in CSV files, one file for each year of filing.

After processing the patent filing numbers, an algorithm was developed using the Python programming language, which runs through all the CSV files with the results of processing the patent filing numbers and making use of the services available in the OPS, performs the consultation of each patent, using the previously processed patent application numbers as search criteria, storing each patent located in the Espacenet repository, in a file in .json format.

After 241 h of running the algorithm, it was possible to retrieve data from 722,347 patents successfully identified on Espacenet, about 83% of the 862,726 patents collected at the INPI. The collection was carried out between July to December 2020 and January and February 2021.

The next step of this work was to collect CVs registered on the Lattes Platform with patent information, such as the filing application number or the patent title, to validate such

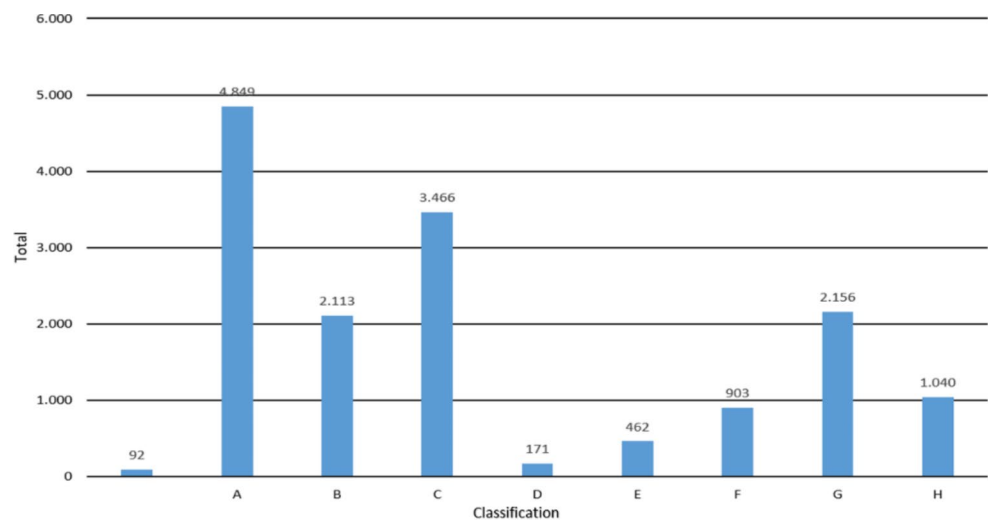
information with the set collected from Espacenet. Collecting and selecting curriculum data from the Lattes Platform was carried out using the LattesDataXplorer framework [27].

5 Results

As a result, initially, 722,347 patents were identified in the Espacenet repository, about 83% of the set of patents collected at the INPI. A hypothesis for unidentified patents is given by the fact that they have not yet been made available in the Espacenet repository, or by problems in identifying the correct format of the patent application number.

With the set of patent data retrieved from the Espacenet repository, an annual patent filing was analyzed, presenting data between 1972 and 2020, with the year with the highest number of patents filed being 2012, with a total of 30,774

Fig. 8 Classification of registered patents in Lattes Platform curricula. **A** - Human needs; **B** - Processing operations, transport; **C** - Chemistry; Metallurgy; **D** - Textiles, Paper; **E** - Fixed constructions; **F** - Mechanical engineering, Lighting, Heating, Weapons, Explosions; **G** - Physics; **H** - Electricity



deposit requests (Fig. 3). It is also worth noting that there was a continuous growth in deposits until 2012. And afterward, a significant drop.

In order to evaluate the representativeness of patent records registered in the Lattes Platform curricula and duly identified in Espacenet, an analysis was conducted using a verification between the sets. Of the total of 72,256 records with patent data extracted from 29,514 curricula, after processing and cleaning the data referring to patents extracted from the Lattes Platform curricula, it was possible to identify 15,252 patents from the set extracted on Espacenet. Figure 4 shows the distribution of patents identified in the CVs by filing year.

The patents reported in the curricula registered on the Lattes Platform were deposited between the years 1975 to 2018, with a greater concentration between 2002 and 2016, highlighting an increase in the number of patent deposits between 2000 and 2018. It is noteworthy that the set of CVs used in the analysis was collected in 2019, which justifies the absence of patents in recent years; a sudden drop is possible in 2018, as some patents may not have been registered by their proponents. It is important to point out that the Lattes Platform was created in 1999, which may explain the low number of patents between 1975 and 1999.

A relevant analysis referring to the technical production recorded in the Lattes Platform curricula consists of identifying the academic background of the proponents. This factor makes it possible to outline the profile of researchers who contribute to national technological development and the level of training that most favor the growth of the number of patent deposits in Brazil. Figure 5 shows the distribution of researchers considering the highest level of informed academic training; the selection criterion adopted was to count all CVs that have at least one valid patent on Espacenet and that have educational information registered in their CV, considering the higher level.

In Fig. 5, the values referring to improvement were omitted due to their low representativeness and for better data visualization. Evaluating the profile of patent applicants is important to understand the profile of these applicants, identify specialists in certain areas of knowledge, and point out possible partnerships for developing new technologies. According to Dias [27], doctors are responsible for most of the scientific production registered in the curricula of the Lattes Platform, and this phenomenon is repeated in technical production; doctors and post-doctors together correspond to 80.4% of the total depositor's patents analyzed. Although postdoctoral is not a level of academic training, individuals were classified in this category for better visualization and classification.

In the CVs of the Lattes Platform, individuals can inform in their data what their primary areas of expertise are. This fact makes it possible to determine which area of knowledge is most representative given the profile of the proponents analyzed and, consequently, to assess which area of knowledge is characterized as the driving force behind national technological development. Figure 6 presents the significant areas of expertise and their respective number of individuals. The selection criterion adopted was to count all CVs with at least one valid patent on Espacenet and information on the area of expertise registered in their CV. As a curriculum can have up to six records of areas of activity, for this analysis, the first area informed was considered.

Technological development is directly linked to the most diverse areas of knowledge, which implies the formation of multidisciplinary teams with a high level of expertise. However, the engineering area stands out in technological development in the analyzed set.

In addition to analyzing the profile of individuals with patents informed in their CVs registered on the Lattes Platform, the analysis of their professional addresses allows us to verify how these applicants are geographically

distributed in the country. Figure 7 shows the distribution by the state where the set of analyzed individuals are working professionally.

It is possible to observe a large concentration in the southeast and south regions of Brazil, this fact can be explained by the fact that they are the most economically developed regions and because they house the largest public universities in the country and, consequently, concentrate a higher percentage of researchers.

Given this, several analyses can be made possible to better understand the retrieved dataset. According to its nature and purpose, each patent receives a classification according to the international patent classification system, the IPC (International Patent Classification), most of the patents produced by researchers with information extracted from their curricula, about 32%, are classified as “A-Human Needs”. Figure 8 presents a graph with the patent classifications reported in the Lattes Platform curricula.

In a general context, most Brazilian patents receive the classifications A - Human needs (32%); C - Chemistry; Metallurgy (23%); and G - Physics (14%).

Given the above, several new analyzes can be carried out, mainly through the integration of Espacenet data repositories and the data recorded in the Lattes Platform curricula. Such analyses will be necessary to understand better how technical production has evolved in the country and the profile of researchers who have deposited patents.

6 Final Considerations

From the results obtained by the extracted data set, it was possible to verify the great viability and scientific value of adopting patent information as a source of data for analyses of the technical production of a country, region or area of knowledge, characterized as paramount importance for understanding the national technological scenario. The group of Brazilian patents identified in Espacenet is described as a significant portion of the entire set of data registered at the INPI, given the high complexity in placing them in Espacenet, due to the lack of a standard for converting the deposit numbers of patents registered until the year 2011.

Assessing the representativeness of the patents registered in the Lattes Platform curricula in international repositories is presented as an important alternative, considering that such identification may even enable a validation of the data recorded in the curricula. Only 72,256 patent records were identified in the more than 6.8 million CVs analyzed. Few CVs present patent information; that is, only 1% of CVs have patent records, highlighting the need for a study to understand this type of production better.

Based on the results obtained, the analysis of the collaboration network becomes feasible in future works. Due to the fact that works are being carried out more and more in a collaborative way, an analysis based on collaborations in the proposal of patents becomes relevant. The purpose of the study is to be able to identify the collaboration and cooperation networks, as well as to measure the efficiency of the constituted bonds.

Declarations

Conflict of Interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

- Amadei JRP, Torkomian ALV (2009) University patents: analysis of the deposits of paulista public universities (1995–2006). 38:9–18 *Ciência da Informação*. <https://doi.org/10.18225/ci.inf.v38i2.1241>
- Mueller SPM, Perucchi V (2014) Universities and the production of patents: topics of interest to the student of technological information. *Perspect. Inform. Sci*, 19(2), 15–36. Available in: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362014000200003&lng=pt&tlng=pt. Access in: jan. 2023
- Ribeiro SPMJC (2014) O estado da arte da patentometria em periódicos inter-nacionais da ciência da informação. Encontro Brasileiro de bibliometria e cientometria
- Souza ASN, Bandeira AA, Ubirajara WM, Holanda FSR (2021) Mapeamento de depósitos de propriedade intelectual no INPI por discentes do programa de pós-graduação em Ciência da propriedade intelectual da universidade Federal de Sergipe. <https://doi.org/10.51722/S2318-3403202100011344>
- Cativelli AS (2020) Metric indicators of patent value: construction of a Value Index using brazilian green patents. Thesis (Doctorate in Information Science) - Federal University of Santa Catarina - Education Science Center. <https://doi.org/10.13140/RG.2.2.13682.12481>
- Sebrae D, de Patente, Brasil (2021) Available in: <https://www.sebrae.com.br/sites/PortalSebrae/artigos/definicao-de-patente,230a634e2ca62410VgnVCM100000b272010aRCRD>. Access in: jan. 2023
- Brasil (1996) Código de propriedade industrial - lei 9279/96. Diário Oficial da União, mai. Available in: https://www.planalto.gov.br/ccivil_03/leis/l9279.htm. Access in: jan. 2023
- INPI. Instituto Nacional da Propriedade Industrial. Brasil (2021) Available in: <https://www.gov.br/inpi/pt-br>. Access in: jan. 2023
- Mueller SPM, Perucchi V (2014) Universidades e a produção de patentes: tópicos de interesse para o estudioso da informação tecnológica. *Perspectivas em Ciência da Informação* 19:15–36. <https://doi.org/10.1590/1981-5344/1828>
- WIPO. Sistema de Classificação de patentes. 2020. Available in: <http://ipc.inpi.gov.br/classificacoes/ipc/ipcpub/?notion=scheme&version=20200101&symbol=none&menulang=pt&lang=pt&viewmode=f&fipcp=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=02n&initial=A&cwid=none&tree=no&searchmode=smart>. Access in: jan. 2023.
- Brandão FG (2016) Democratization of information from the use of institutional digital repositories: from scientific communication to technological patent information. Thesis (Master's degree) Universidade Regional Integrada do Alto Uruguai e

- das Missões, set. 2016. Available in: <https://lume.ufrgs.br/handle/10183/179853>. Access in: 24 mar. de 2021
12. Scartassini FVB et al (2018) Patentometric study of Brazilian patents via pct. Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (XIX ENANCIB), n. XIX ENANCIB, Available in: <http://hdl.handle.net/20.500.11959/brapci/102164>. Access in: jan. 2023
 13. Menezes MS (2023) Potentiometric study related to pulmonary tuberculosis therapy. Thesis (Master's degree) - Fundação Oswaldo Cruz, Programa de Pós-graduação Profissional em Gestão, Pesquisa e Desenvolvimento na Indústria Farmacêutica, do Instituto de Tecnologia em Fármacos, 2019. Available in: <https://www.arca.fiocruz.br/handle/icict/37031>. Access in:
 14. Fagundes MC, de Almada Garcia PA, da Silva Motta G, Armond-de-Melo DR (2014) Perfil Tecnológico da CSN: um estudo patentométrico. RAI Revista de Administração e Inovação 11(1):276–294. <https://doi.org/10.5773/rai.v11i1.1307>
 15. Nascimento R, Speziali M (2020) PATENTOMETRIA. Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN), 4(1). Available in: <https://desafioonline.ufms.br/index.php/EIGEDIN/article/view/11705>. Access in: jan. 2023
 16. Narin F (1995) Patents as indicators for the evaluation of industrial research output. *Scientometrics* 34(3):489–496. <https://doi.org/10.1007/bf02018015>
 17. GUZMAN SANCHEZ MV (1999) Patentometría: herramienta para el análisis de oportunidades tecnológicas. 1999. 130 f (Doctoral dissertation, Tese (Doutorado em Gerência de Informação Tecnológica)—Facultad de Economía, Universidad de La Habana, La Habana)
 18. Aristodemou L, Tietze F (2018) The state-of-the-art on intellectual property analytics (IPA): a literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Inf* 55:37–51. <https://doi.org/10.1016/j.wpi.2018.07.002>
 19. QUONIAM LM, KNISS, C. T., MAZIERI MR (2014) The patent, object of research in Information Science and Communication. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*; v. 19, n. 39 ; 243–268, 24(2), 268 – 243. DOI: <https://doi.org/10.5007/1518-2924.2014v19n39p243>
 20. Morais SP, Garcia JCR (2014) O estado da arte da patentometria em periódicos internacionais da ciência da informação. Encontro Brasileiro de Bibliometria e Cientometria, 4. Available in: <http://hdl.handle.net/20.500.11959/brapci/45400>. Access in: jan. 2023
 21. Cativelli AS (2023) Patents of Brazilian Public Universities: Analysis of Concessions. Dissertation (Professional Master's Dissertation in Management of Units of Information) - University of the State of Santa Catarina, 2016. Available in: https://www.udesc.br/arquivos/faed/id_cpmenu/1441/dissertacao_adriana_15689010225034_1441.pdf. Access in:
 22. Mendes MLS, de Melo DRA (2017) Technology assessment: a methodological proposal/Avaliacao tecnologica: uma proposta metodologica. *RAC-Revista de Administracao Contemporanea*, 21(4), 569–585. Available in: <https://link.gale.com/apps/doc/A502000376/AONE?u=anon-1910dca0&sid=googleScholar&xid=bad05f8f>. Access in: jan. 2023
 23. Speziali MG, Nascimento RDS (2021) *Química Nova* 43:1538–1548. <https://doi.org/10.21577/0100-4042.20170620>. Patentometria: uma ferramenta indispensável no estudo de desenvolvimento de tecnologias para a indústria química
 24. Cativelli AS, Pinto AL, Sanchez MLL (2021) Patent value index: measuring brazilian green patents based on family size, grant, and backward citation. *Iberoamerican J Sci Meas Communication* 1(1):004–004. <https://doi.org/10.47909/ijsmc.03>
 25. Reymond D (2021) Patents information for humanities research: could there be something? *Iberoamerican J Sci Meas Communication* 1(1):006–006. <https://doi.org/10.47909/ijsmc.02>
 26. Serrano BP, Junior JAG (2014) *Revista Gestão da Produção Operações e Sistemas* 9(1):101. <https://doi.org/10.15675/gepros.v0i1.1145>. Redes de inovação: mapeamento de inventores de patentes em uma empresa do setor de cosméticos
 27. Dias TMR A study of Brazilian scientific production based on data from the Lattes Platform. 181 p. Thesis (Doctorate in Mathematical and Computational Modeling) - Federal Center for Technological Education of Minas Gerais, September 2016. Available in: [url?sa=t& \(2023\) rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjh0riyucP-AhXDJLkGHUARCA0QFnoECAGQAQ&url=https%3A%2F%2Fsig.cefetmg.br%2Fsigaa%2FverArquiuvo%3FidArquivo%3D2033874%26key%3Dd8d1d2008e1ebe20f0f136527af3a222&usg=AOvVaw1W65ontFW5A019OGCx9900](https://repositorio.ufmg.br/handle/1808/11145). Access in: jan.
 28. Coimbra FS, Dias TMR (2021) Use of open data to analyze the publication of articles in scientific events. *Iberoamerican J Sci Meas Communication* 1(3):1–13. <https://doi.org/10.47909/ijsmc.123>
 29. Mascarenhas H, Dias TMR, Dias P (2021) Academic mobility of doctoral students in Brazil: an analysis based on Lattes platform. *Iberoamerican J Sci Meas Communication* 1(3):1–15. <https://doi.org/10.47909/ijsmc.53>
 30. Dias TM, Dias PM, Moita GF (2022) Collection and integration of scientific data for production analysis Brazilian scientific in open access journals. *Advanced Notes in Information Science*, 2, 34–42. DOI: <https://doi.org/10.47909/anis.978-9916-9760-3-6.105>
 31. Mascarenhas HAD, Dias TMR, Dias PM (2020) Adoption of Network Analysis Techniques to understand the training process in Brazil. *AWARI* 1(1):e004–e004. <https://doi.org/10.47909/awari.63>
 32. ESPECENET Espacenet patent search (2023) Available in: <https://worldwide.espacenet.com/patent/>. Access in:

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.