

# Extraction and validation of scientific data for the identification of publications in open access journals

## *Extração e validação de dados científicos para identificação de publicações em periódicos de acesso aberto*

Patrícia Mascarenhas DIAS<sup>1</sup>  0000-0002-8448-6874

Thiago Magela RODRIGUES DIAS<sup>1</sup>  0000-0001-5057-9936

Gray Farias MOITA<sup>1</sup>  0000-0002-6510-1019

Adilson Luiz PINTO<sup>2</sup>  0000-0002-4142-2061

### Abstract

This work aims to identify articles published in open access journals registered in the Lattes Platform curricula. Currently, the curricular data from the Lattes Platform has been the source of several studies that adopt bibliometric metrics to understand scientific evolution in Brazil. However, when registering a publication in a curriculum, only basic information from the magazine is informed. Therefore, to quantify the publications made in open access journals, a strategy is proposed that uses data from the Directory of Open Access Journals, validating the publications and, thus, obtaining a process that allows identifying which publications were made in this communication format. Therefore, using the proposed methodology, which has a low computational cost, it was possible to quantify in an unprecedented way the publications of Brazilians in open access journals registered in the curricula of the Lattes Platform. In addition, having the authors' data of the publications, it was possible to understand how these authors are geographically distributed in the country, what their areas of operation are and how they have published the results of their research in open access. It is noteworthy that they are generally published in Portuguese and English. Over the last few years, the representativeness of publications in open access journals has increased.

**Keywords:** Data integration. Directory of Open Access Journals. Lattes Platform. Open access.

### Resumo

*Este trabalho tem como objetivo identificar artigos publicados em periódicos de acesso aberto cadastrados nos currículos da Plataforma Lattes. Atualmente, os dados curriculares da Plataforma Lattes têm sido fonte de diversos estudos que adotam métricas bibliométricas para compreender a evolução científica no Brasil. Porém, ao registrar uma publicação em um currículo, apenas informações básicas da revista são informadas. Portanto, para quantificar as publicações que foram feitas em periódicos de acesso aberto, é proposta uma estratégia que utiliza dados do Directory of Open Access Journals, validando as publicações e, assim, obtendo um processo que permite identificar quais publicações foram feitas neste formato de comunicação. Com isso, utilizando a metodologia proposta, que*

<sup>1</sup> Centro Federal de Educação Tecnológica de Minas Gerais, Programa de Pós-Graduação em Modelagem Matemática e Computacional. Av. Amazonas, 7675, Nova Gameleira, 30510-000, Belo Horizonte, MG, Brasil. *Correspondência para/*Correspondence to: T. M. RODRIGUES DIAS. E-mail: <thiagomagela@cefetmg.br>.

<sup>2</sup> Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Departamento de Ciência da Informação. Florianópolis, SC, Brasil.

Received on June 16, 2021 and approved on November 17, 2021.

*Como citar este artigo/How to cite this article*

Dias, P. M. *et al.* Extraction and validation of scientific data for the identification of publications in open access journals. *Transinformação*, v. 34, e210040, 2022. <https://doi.org/10.1590/2318-0889202234e210040>



*possui baixo custo computacional, foi possível quantificar de forma inédita as publicações de brasileiros em periódicos de acesso aberto registradas nos currículos da Plataforma Lattes. Além disso, de posse dos dados dos autores das publicações, foi possível compreender como estes autores estão distribuídos geograficamente no país, quais suas áreas de atuação e como eles têm publicado os resultados de suas pesquisas em acesso aberto. Já com relação as publicações, destaca-se que em geral estão publicadas no idioma Português e Inglês e que ao longo dos últimos anos, a representatividade das publicações em periódicos de acesso aberto vem aumentando.*

**Palavras-chave:** *Integração de dados. Directory of Open Access Journals. Plataforma Lattes. Acesso aberto.*

## Introduction

Understanding the evolution of a country's scientific and technological development is extremely. It makes it possible to identify how research progress in different areas of knowledge has evolved historically. In addition, this understanding makes it possible to identify the main research topics, the profile of the researchers and their scientific collaborations, which can serve as a basis for various policies to promote scientific research.

Due to information and communication technology, the traditional printed format of science communication gradually gives way to new electronic formats. In the context of bibliometric research and studies, scientific communication emerges today as a central element at various levels of discussion. Therefore, the scientific journal appears as an important mechanism for communicating research results.

Mueller (1999) states that the scientific journal performs at least four essential functions: certification of science with the support of the scientific community; communication channel between scientists and wider dissemination of science; scientific file or memory and record of authorship of scientific discovery.

In another study, Mueller (2006) highlights that the scientific journal is characterized as the most desired and efficient channel for the dissemination of research results, in addition to being the one that gives greater visibility to the researcher who discloses his results, considering that such medium is accepted and validated by the scientific community. For the author, although the scientific journal is not, even today, the predominant vehicle in all areas of knowledge, scientific journals indexed and referenced by peers have become the center of the traditional scientific communication system. Its preferred channel status was granted by the scientific community, which delegates the task of confirming the authorship of the scientific discovery and giving prestige to scientists and scientific journals through an evaluation system based on indicators, such as the number of publications, indexes of international citation and visibility.

According to several studies, journals, mainly electronic format grow since the last decade. It can be said that journals, in all areas of knowledge, have the role of being a filter for the recognition of works that have been accepted. For Rodrigues and Oliveira (2012), publication in a magazine recognized by the area is the most accepted way to register the work's originality and confirm that the works were reliable enough to overcome the skepticism of the scientific community.

In this context, in the early years of the 21<sup>st</sup> century the Open Access Movement, whose definition is "[...] to make available to any internet user to read, download, copy, distribute, print, search or reference the full text of articles or use them for other purposes without any barriers, as long as the work is properly recognized and cited", encouraged the appearance of journals in this format (Leta; Costa; Mena-Chalco, 2017, p. 3).

Despite the numerous benefits that open access journals provide, there is a need for a joint effort so that the main element of the whole process, scientific information, is accessible to all interested parties. To this end, some initiatives have already been undertaken, such as the creation of digital repositories to store and organize scientific literature under international interoperability standards and the search for awareness of the main actors involved in the process of production, publishing and evaluation of scientific information, to make such content available in digital environments open to the general public.

Neubert, Rodrigues and Goulart (2012) state that open access assumes an important role in the entire context of scientific activity, as it allows the researcher to have access to the results of other studies without the cost barriers and difficulties of access, in addition to promoting visibility and dissemination the results of the scientific activities of each researcher and each university.

Open access scientific publication is part of a broader scenario in favor of opening knowledge in general (open access, open data, open educational resources, free software, open licenses) and is essentially a movement towards the design of information and knowledge as public goods (Furnival; Silva-Jerez, 2017).

Governments of several countries recognize that open access to data, information and knowledge contributes decisively to advances in scientific research and innovation, in addition to maximizing the value derived from public investments, bringing benefits to the economy and society and inserting countries developing in the global science system, contributing to its economic and social development (Organização para a Cooperação e Desenvolvimento Econômico, 2004).

It is worth mentioning that there is generally a limited number of resources to promote research and a large number of researchers or institutions interested in these resources. Therefore, the broader and more accurate this understanding of scientific production, the greater the possibility of determining resources correctly. However, this type of assessment is a highly complex task, as it involves the analysis of different characteristics, both quantitative and qualitative. In addition, there is no consensus on which measures or characteristics should be considered to assess of scientific productivity (Digiampietri, 2015).

Bearing in mind that a large part of scientific research in the country is financed with public resources, usually in public educational institutions or research centers, it is expected that such studies will be disseminated without any barrier, mainly financial. In this context, coupled with the advantages that open access publications have, such as availability, visibility and accessibility, several efforts are being made to ensure that more and more scientific articles are published in open access journals.

Therefore, understanding how the publications of a specific group of researchers have been carried out in open access journals, makes it possible to identify an overview of the current stage of this type of communication in Brazil. It also allows verifying if this type of publication tends to be more frequent in certain areas of knowledge.

This type of study is an essential mechanism to evaluate the evolution of publications in open access journals by Brazilian researchers, verifying whether the incentive policies for publication in this communication format have achieved satisfactory results.

## Related works

In the work of Silva and Alcará (2008), the authors analyzed the policies of open access to scientific information and the proposals for action, emphasizing on government initiatives in different countries. It was identified that the movement of free access to scientific information was already a concern officially registered in several countries, although with different degrees of development. Among these differences are the policy determinations themselves, as some oblige public institutions and researchers to make their research results available in open access. In contrast, others only suggest the involvement and participation of these researchers and institutions in the movement.

Oliveira and Chalhub (2009) analyzed the Ibero-American scientific journals that joined the open access movement, members of the Directory of Open Access Journal (DOAJ). Specifically, this study aimed to identify the journals of the Ibero-American region included in the repository, their publishing institutions, their periodicity and their coverage areas, in addition to verifying their insertion in the free access movement, through the analysis of their participation in other virtual spaces. The results showed an increased adhesion of scientific journals in the

region. Internationally, Brazil and Spain occupy the second and fourth positions, respectively. Teaching and research units represent the majority of publishing institutions. It also highlights the diversity of the areas of the journals, ranging from Engineering to Linguistics, with an expressive participation of Medicine. It is concluded that specific data from a directory seem to strengthen the legitimacy of the free access movement by the actors involved in the scientific communication system.

Chalhub and Pinheiro (2011), identified the main open access scientific communication channels used by researchers, and the factors involved in adhering to the self-archiving of their scientific production are analyzed. The objective of the work was to identify the main channels of scientific communication in open access used by researchers from public universities in the State of *Rio de Janeiro*. They used a list with 47 National Council for Scientific and Technological Development (CNPq) Advisory Committees for Research Productivity Scholarships. Stratified probabilistic sampling by knowledge area was carried out, following the division by Advisory Committee (Agrarian Sciences, Biological Sciences, Exact and Earth Sciences, Science of the Health, Human Sciences, Applied Social Sciences, Engineering, and Linguistics, Letters and Arts). From the selection of researchers contemplated by the CNPq Research Productivity Scholarship program in 2010, whose list is available on the website of this federal agency, those linked to public universities in the State of *Rio de Janeiro* with postgraduate courses and *stricto sensu* graduation were identified. After identifying the e-mail addresses of those selected, correspondence was sent containing the form with closed and open questions attached to the following categories: informational behavior, open access publication and adherence to institutional repository.

In general, the results of the research point to a change in the attitude of these researchers about publication of research results in open access channels. Some present areas publications informal channels of scientific communication, such as electronic journals, and self-archiving in institutional or thematic repositories. Others are more part of individual or group research initiatives, often anticipating institutional policies. The researchers were unanimous regarding the advantages of open access publishing, and the majority pointed out the democratization of knowledge as the main advantage of this adhesion. In addition to this aspect, the benefit of communication between peers – “exchanges”, “partnerships” and “dialogues” – also appears in the speeches of the researchers in the knowledge production process. It also signaled the importance of using this open communication channel at two different times: for the researcher to access the information for their research and make their results available, allowing them greater visibility and impact.

To explore the national and international scenario and thus present an investigation that seeks a technological solution to effect open access to research data, Pavão, Rocha and Gabriel Junior (2018) propose a methodology divided into five stages: (a) identification of practices of open access to research data in Brazilian institutions; (b) mapping your users and their needs; (c) proposal for a web portal to bring together the national community; (d) survey of services and technological solutions existing in the international scenario for the sharing of research data; and (e) proposing recommendations to support the creation of research data repositories in national institutions and their aggregation to a research network with open access to research data. As a result, international initiatives and strategies are proposed to create a research data repository and communities of practice around the subject.

In the work of Taga (2016), research articles on open access indexed in the Scopus database are analyzed, published between the years 2001 and 2015, aiming to propose a classification scheme for the subthemes of open access, classifying the research articles on open access, check the evolution of the identified subthemes and organize a bibliography on open access. Three hundred forty-seven research articles were analyzed, and as a result a classification scheme on open access was proposed, with eight classification categories being established. The study revealed a continuous and growing interest from the research community in carrying out case studies dedicated to analyzing the development or evolution of open access specific groups, institutions, regions and periods.

For Minniti, Santoro and Belli (2018), the spread of the open access movement in Latin American and Caribbean countries, driven by the growth of regional and national initiatives such as the creation of digital magazine libraries in open access and the establishment of government policies of support, has provided evidence of the significant role of open access for the participation of these countries in global scientific production. This work maps open access publications from Latin America and Caribbean countries, through a bibliometric analysis of the publications indexed by Web of Science and Scientific Electronic Library Online from 2005 to 2017. It is found that the journals have intensified significantly in the period examined, and that although there is an increase in the number of publications in this format, its magnitude does not translate into a relative weight of open access in the total number of publications in some countries.

In the work of Pavan and Barbosa (2018), the authors analyze documents published in open access between the years 2012 and 2016 by authors with Brazilian affiliations and identify the profile of these publications. Data from 930 journals and 63,847 documents were collected from Web of Science. It is also noteworthy that an endogenous profile characterizes the Brazilian scientific production in open access. Policies are still necessary to encourage the publication of articles in open access, mainly in international journals.

Considering the works that analyze publications in open access journals, it is clear that most of them analyze small sets of individuals, in addition to using international data repositories, thus neglecting publications from some areas of knowledge and, therefore, not representing significantly the Brazilian production in open access as a whole.

Most studies that evaluate the open access movement do not have publications in this format as their primary object of study, but repositories or journals in open access. Therefore, although the works presented in this section are important to understand Brazilian researchers' existing initiatives and opinions as the main open access repositories in Brazil. A comprehensive study of Brazilian researchers who have published widely published papers in open access journals is necessary.

## Methodological Procedures

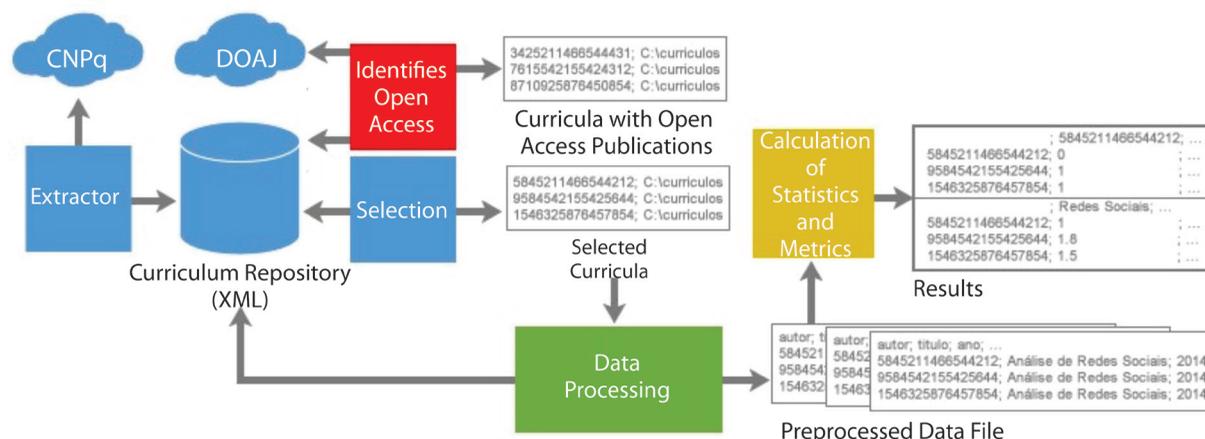
For the process of data extraction for the analyzes to be carried out in the context of this work, curricular data from the CNPq Lattes Platform were used. A large part of the funding notices for research projects, carried out by various funding agencies, use data registered in the applicants' curricula as one of the forms of evaluation of the proposals. Therefore, researchers have a great incentive to keep their curriculum information up to date. This makes Lattes Platform curricula an excellent source of data for analysis. For this same reason, several works have used the Lattes Platform as a data source for several studies on different topics, such as networks of scientific collaborations, productivity analysis, and academic genealogy (Dias, 2016; Digiampietri, 2015).

Considering that the majority of related works analyzed only specific groups of individuals, and considering that the manipulation of large amounts of curricula from the Lattes Platform is not a trivial task, since there are problems involving information retrieval and efficient algorithms for handling large volumes of data, LattesDataXplorer (Dias, 2016), a framework for data extraction and treatment, developed by the research group of this work was used.

LattesDataXplorer is responsible for encompassing a whole set of techniques and methods for collecting and processing data used in this work. It consists of components that allow it to collect and process the curricula. The process of extracting all curriculum data from the Lattes Platform is divided into components that aim to minimize the computational cost required for this task. With all the resumes stored locally in XML format, the possibility of manipulating the data with flexibility allows exploring all the collected data's potential. To analyze specific curricula groups, such as those formed by professors in a postgraduate program or in a particular institution, a component called "Selection" is used, which comprises subgroups of curricula based on information present in their curricula records (Dias, 2016).

Therefore, using LattesDataXplorer, it is possible to group a set of curricula based on desired parameters. In the search for the parameters in each of the curricula, regardless of the section in which they are found, these curricula are selected and form a group for analysis. The data for these curricula are organized into a list of selected curricula.

As already explained, a curriculum registered in the Lattes Platform can contain various information capable of helping to understand the evolution of Brazilian science from different perspectives. However, to serve the purposes of this work, only data from publication of articles in open access journals were considered. Given this, an extension of LattesDataXplorer was proposed with the inclusion of nonexistent a priori components, which would evaluate for each article published in a journal (6,985,179), of each of the individuals (5,901,161) (data collections in October 2018), if the journal in which that article had been published as open access (Figure 1). Therefore, only authors and publications in open access journals could be analyzed with the proposal for this extension.



**Figure 1** – LattesDataXplorer extended.

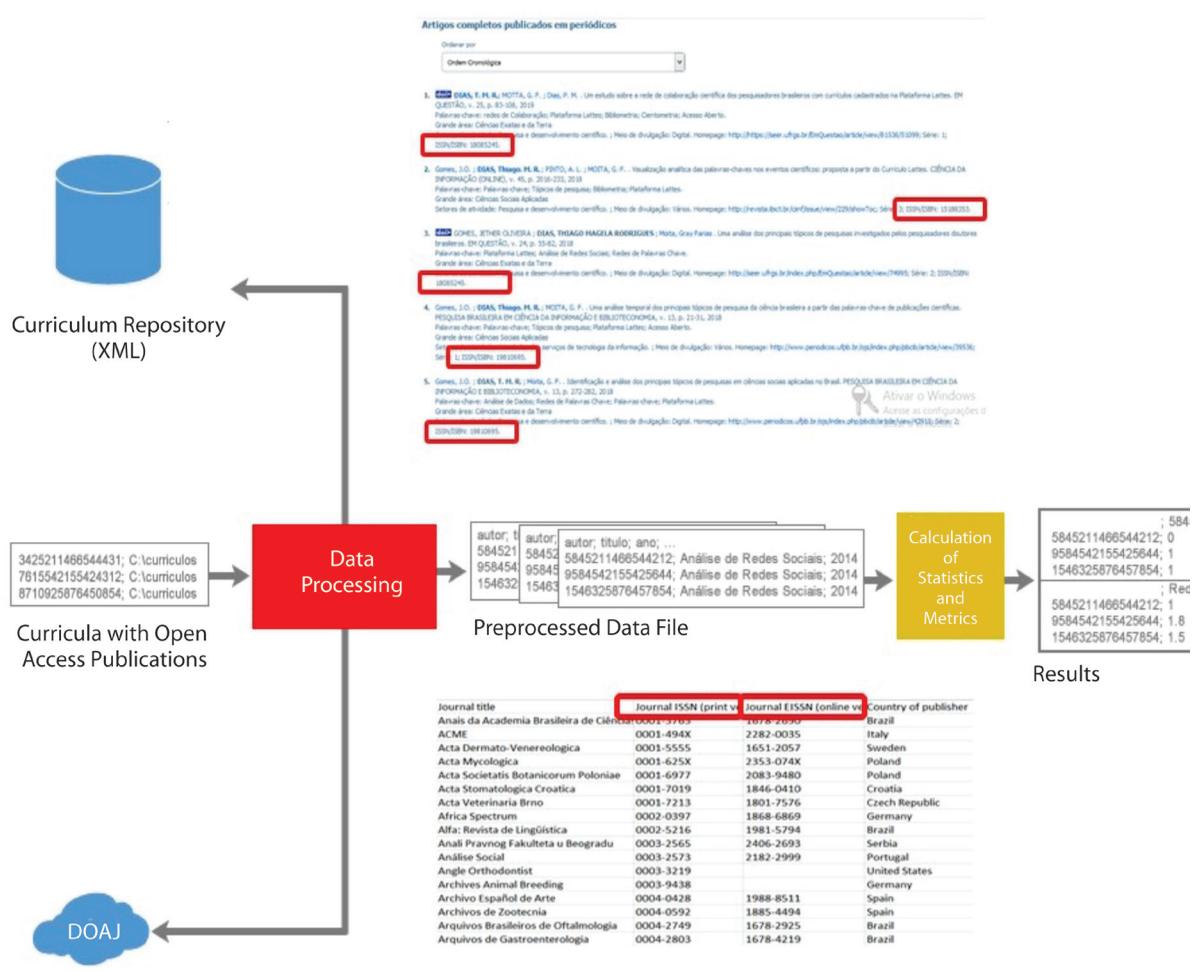
Source: Elaborated by the authors (2021).

Using LattesDataXplorer, all resumes registered on the Lattes Platform in October 2018 were collected and stored in the local repository. Then, the component developed and called “Identifies open access” was used to retrieve all open access journals registered on the DOAJ portal. This online directory indexes and provides access to open access journals. In February 2019, the DOAJ indexed 12,324 journals and 3,513,782 articles. DOAJ has been a source of data and reference on open access journals for several studies.

Collected the journals’ data on the DOAJ portal in October 2018, the same period of collection of the curricula for the analyses presented in the present work, 12,171 open access periodical titles were retrieved, containing data such as title, ISSN and eISSN, among other information.

To optimize the computational processing of curricula as much as possible, whenever a publication whose ISSN or eISSN of the journal was contained in the list of open access journals extracted from DOAJ, immediately the identifier of the curriculum under analysis was inserted in the list of curricula in open access, and the following curriculum of the set under analysis was evaluated.

After analyzing all the resumes that make up the local repository, a list containing all resumes with open access publications is generated. It becomes the basis for the “Data processing” component, which incorporates the methods proposed in this work (Figure 2).



**Figure 2** – Method for identifying publications in open access journals.  
Source: Elaborated by the authors (2021).

With the list of curricula that have articles in open access, the identification of publications in this format is performed with the processing of the curricula, using the “Data Processing” module of LattesDataXplorer, to generate the pre-processed data files that summarize information of interest and that will serve as the basis for calculating the metrics.

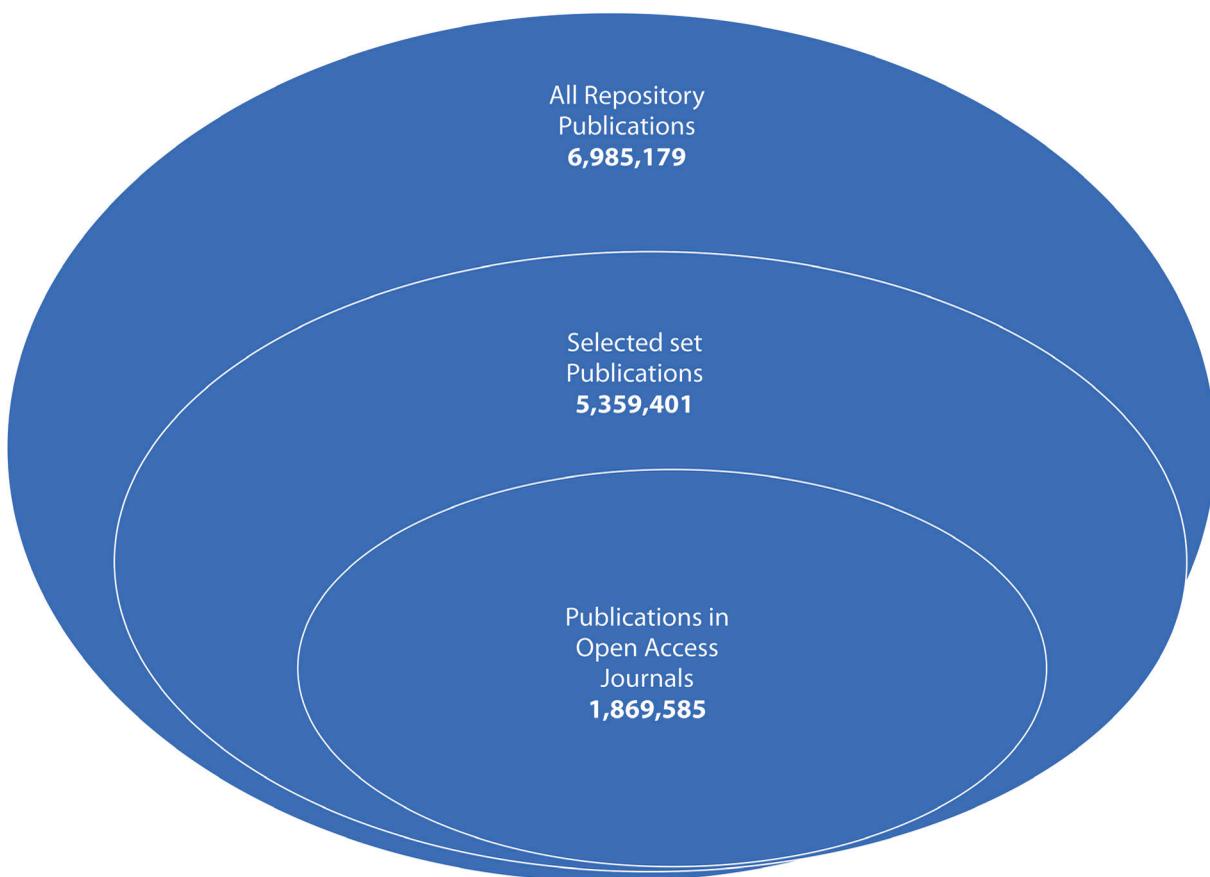
In addition to general data on researchers with open access publications that will compose some of the archives, such as academic training, areas of expertise, guidelines and professional practice, each of the articles recorded in the section “Complete articles published in journals” were analyzed. Of each curriculum contained in the “List of curricula with publications in open access”. For each article in each curriculum, it was verified and analyzed whether the ISSN or eISSN of the publication was present in the list of journals recovered from DOAJ. Thus, it was possible to identify the entire number of articles in open access journals.

Of the total set of articles published in journals, considering the entire history of publications recorded in all resumes registered in the Lattes Platform (6,985,179 publications), a percentage of 26.76% (1,869,585) was published in open access journals, taking into account the list of journals recovered from DOAJ. This percentage of publications in open access is relevant, above all, for considering the entire publication history of each researcher. It is noticed that publications in open access journals have been receiving attention

and adherence by researchers year after year, presenting themselves as a trend in dissemination and scientific communication, especially in recent years. A temporal evaluation was carried out to assess the growth of publications year by year.

## Results

Using the extension proposed in this work for LattesDataXplorer, all authors who published at least one article in an open access-journal (370,431) were identified. These authors, despite being a small number of individuals concerning the whole set registered in the Lattes Platform (6.27%), have great representativeness when considering the total number of articles published in journals (approximately 76%) (Figure 3).



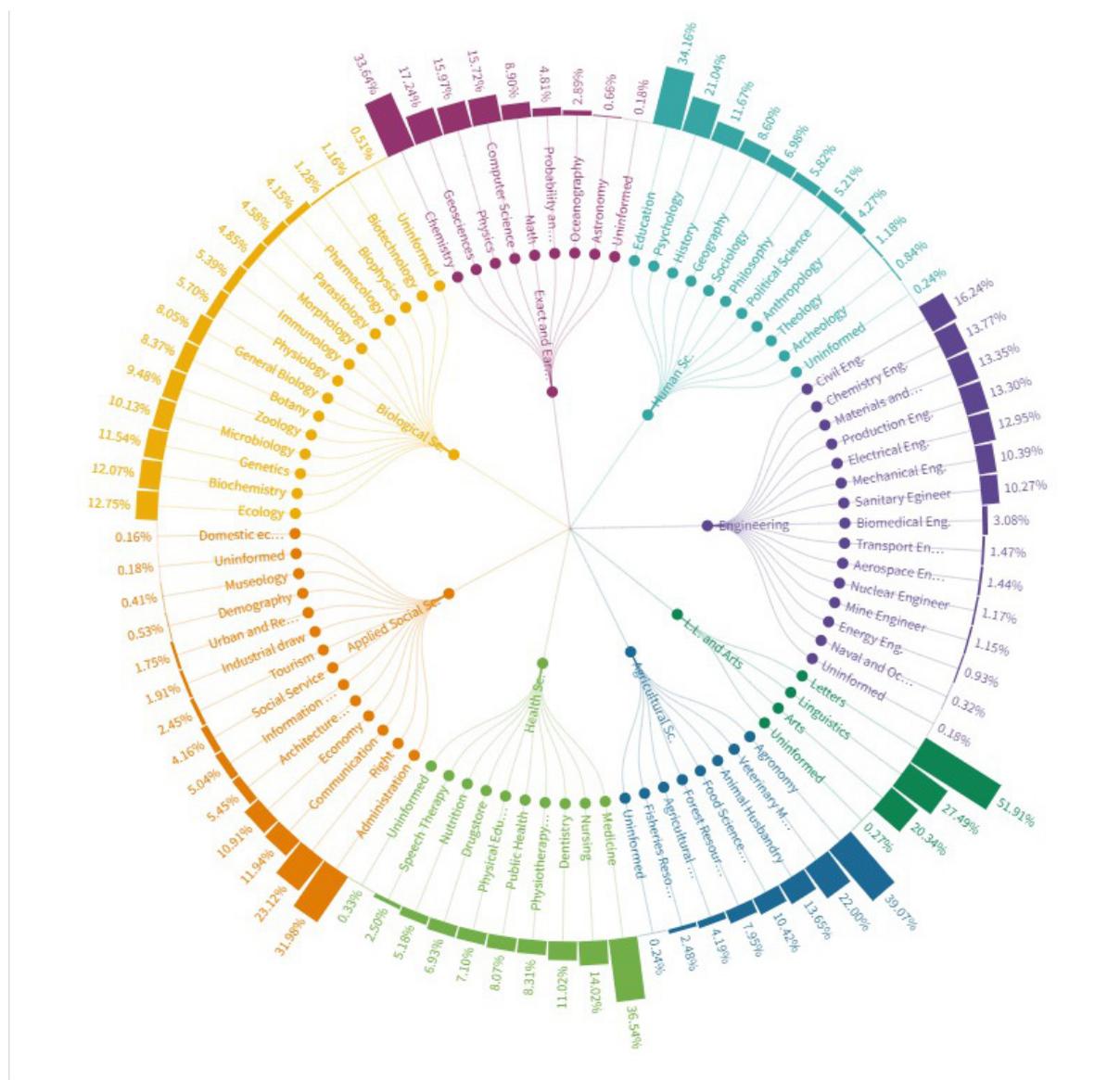
**Figure 3** – Number of publications in registered journals.

Source: Elaborated by the authors (2021).

Therefore, it is possible to note the representativeness of the set to be analyzed in this work. Considering that it includes a considerable portion of the authors who have published articles in journals in Brazil, the results presented may provide an unprecedented view on the evolution of articles in open access and serve as a basis for several other works.

In the curricula, it is possible to inform the areas, subareas and specialties in which a given individual operates. When analyzing the areas of activity of the group of individuals, it is possible to notice great diversity in

the distribution of curricula in each significant area and irregular distribution in the number of areas that each major area has. Therefore, an analysis based on the areas of activity is essential (Figure 4).



**Figure 4** – Distribution of authors by their areas of expertise. Source: Elaborated by the authors (2021).

As can be seen, there is no uniform distribution of the number of areas in each large area. The large area of Linguistics, Letters and Arts has only three areas, while the large areas of Biological Sciences and Engineering have 14 areas each. When registering the large areas of activity, the individual may not inform the “area” field. In these cases, individuals were also categorized as “Not Informed”. Due to the small number of individuals (0.81%) who reported “Others” as a large area, the analysis of their areas was not considered.

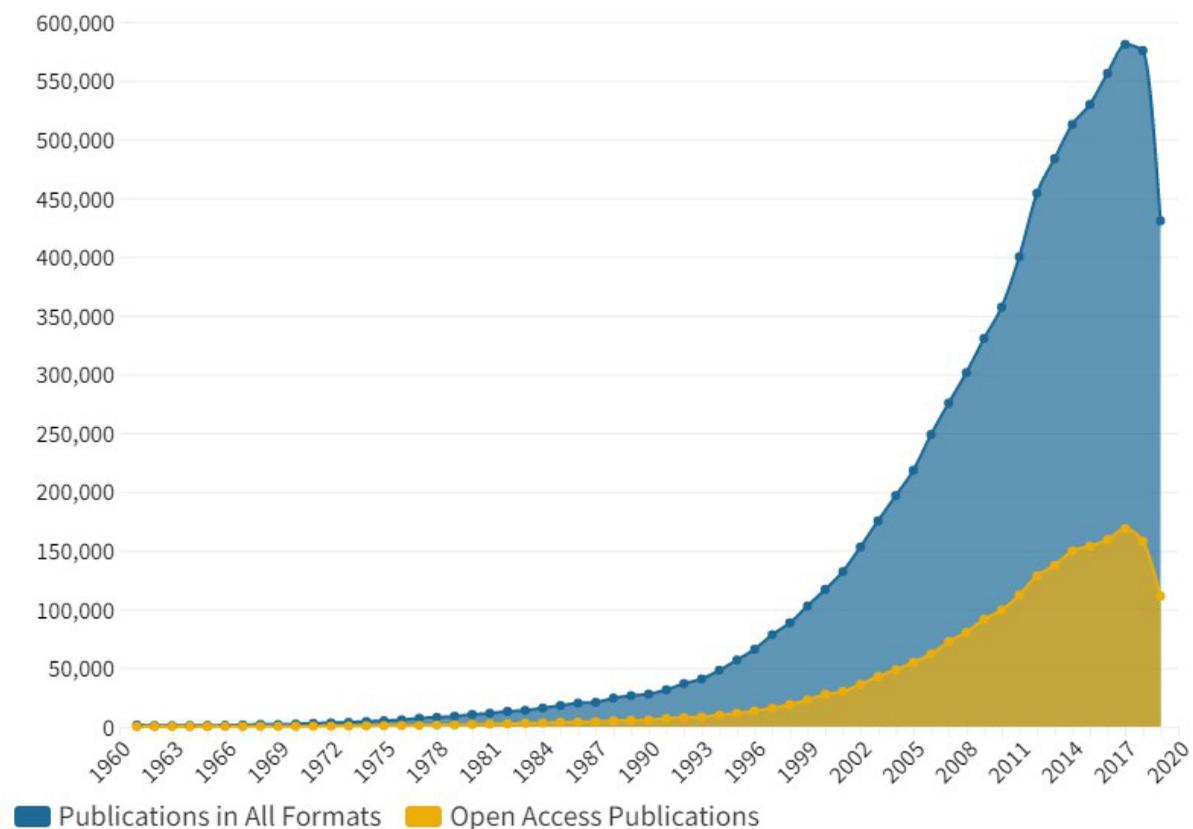
To verify the most representative areas of knowledge of the analyzed group, the area of Medicine (33,966) stands out, composing the large area of Health Sciences, the most representative, as already presented. The representativeness of the field of Medicine is so considerable that it alone has practically the same number of individuals as the total sum of the major areas of Linguistics, Letters and Arts and Engineering combined. Following,

the areas of Education (17,066), Agronomy (15,705), Nursing (12,734), Administration (11,717), Chemistry (10,490) and Psychology (10,511) stand out. These seven areas alone are responsible for housing approximately 30% of the individuals analyzed.

Considering information about the authors of the articles in open access journals, with resumes registered in the Lattes Platform, it is possible to verify how these authors are geographically distributed in the country.

Considering only the curricula that have a professional address informed, it is possible to observe that the vast majority of individuals are concentrated in the Southeast and South regions of the country, with emphasis on the state of *São Paulo*, which concentrates approximately 16% of the group of individuals, a very good percentage. Greater than the sum of individuals from several other states. Then, the states of *Rio de Janeiro*, *Minas Gerais* and *Rio Grande do Sul* stand out. The sum of the quantitative of these four states is greater than the sum of all other states combined. This scenario, not by chance, also coincides with national indices that present similar inequalities, such as levels of education and income distribution. In addition, a hypothesis to explain such distribution may be the large concentration of universities and research centers in these states. In addition to these authors, it was possible to identify 52 individuals from the analyzed group who have declared working in foreign institutions in their curricula with Colombia (7 authors) and Portugal (5 authors) being the most representative countries.

When analyzing the set of publications in the selected set, it is evident that analyzing articles published in journals is one of the main means of disseminating research results. In addition, articles published in journals are also frequently used as indicators for evaluating researchers, research institutions and even graduate programs. Therefore, in the context of this work, which aims to analyze publications in open access journals in Brazil, performing a temporal analysis of the evolution of this means of dissemination is necessary (Figure 5).



**Figure 5** – Temporal evolution of the production of articles in journals in Brazil.

Source: Elaborated by the authors (2021).

The most accentuated growth in articles in open access journals occurred since 2000, when this type of publication represented approximately 24.00% of the total set of articles published in journals. After this period, with the increase in the number of publications over the years, accentuated in 2011, the representativeness of articles in open access journals became more significant, with 25.02% of the total set in 2017. With the fall accentuated in the general quantity of articles in the last year analyzed, articles in open access journals started to have the highest percentage of the entire period analyzed, 25.83% of the total set, thus corroborating the trend and the growing interest of the scientific community to publish the results of their research openly.

An explanation hypothesis for the drop in the number of publications, especially in the last few years, may be the lack of updating of the curricula, since the authors may not have registered recently published works.

Information regarding the language of a given publication is entered by the individual at the time of registration of the article, selecting a language from a previously registered list. Therefore, spelling errors in the registration of this information do not exist; however, as it is not a mandatory registration, it is not uncommon for registrations in which the author does not inform this field, leaving the article without this information. Of the total set of articles in open access journals analyzed in this work, 85 papers do not have information on the language of the publication.

Consequently, as expected, the most representative language is Portuguese (63.76%) followed by less English (35.22%). In general, publications in Portuguese are articles published in national journals, the majority being, thus justifying their representativeness. The articles written in English are mostly published in international journals, with English being the language accepted by the vast majority of international journals.

With a lower quantity, Spanish, French and Italian appear. A total of 1,808 publications are worth mentioning, with other languages informed, such as German, Chinese and Japanese. Collaborations partly influence such publications by researchers fluent in these languages. Therefore, due to the predominance of publications in Portuguese, it is important to highlight the limitations that such publications may suffer related to visibility by the international scientific community.

Therefore, the data presented here, make it possible to have an overview of publications in open access journals in Brazil and a portrait of the group of individuals who have used this means of communication to disseminate the results of their research. Consequently, several other studies can be carried to understand better Brazilian researchers' communication and their adherence to open access.

## Conclusion

To draw a picture of the publication of articles in open access journals by Brazilian researchers, it was necessary to develop components that, incorporated into LattesDataXplorer, could enable the analyzes carried out in this project. Thus, the entire curriculum data repository of the Lattes Platform was analyzed, enabling an unprecedented study on the Brazilian production of articles in open access journals using data from DOAJ.

When analyzing the profile of researchers who have published articles in open access journals, it can be seen that the vast majority of these individuals have high levels of academic training, with completed masters and/or doctorates. It is noteworthy that a significant percentage of individuals are concentrated in certain areas of activity, such as Medicine, Education, Agronomy and Nursing, which may represent a trend in disseminating research results in these areas.

The set of articles published in open access journals has as authors a total of 370,431 individuals, which represents approximately 6% of the total set of individuals with curricula registered in the Lattes Platform. It should be noted that this percentage of authors is much lower than the number of articles in open access journals, which

represent approximately 27% of the total number of articles published in journals of all individuals. This percentage is very close to that presented by other related works, which point out that only around 30% of the total scientific articles published in the world annually are available through open access channels. Therefore, it is identified here that the percentage of publications in open access journals in Brazil is slightly lower than the world average of publications in this format.

Finally, as a result of this study, it was possible to identify that the publication of articles in open access journals in Brazil has increased significantly, especially in recent years, and that the distribution of authors by their large areas of activity demonstrating that some of these areas stand out for the number of people linked to them, such as large areas of Health Sciences and Human Sciences.

In this context, based on the methodology used – the cross between data from the DOAJ and the Lattes Platform, it was possible to map and discover in an unprecedented way the characteristics of Brazilian authors who are publishing articles in open access journals in different areas of the knowledge.

## References

- Chalhub, T.; Pinheiro, L. V. R. *Acesso aberto à informação científica no Brasil: um estudo das universidades públicas do estado do Rio de Janeiro*. Report. Brasília: IBICT, 2010. Available from: [http://sitehistorico.ibict.br/capacitacao-e-ensino/pesquisa-em-ciencia-da-informacao/pos-doutorado/pesquisas-concluidas-1/acesso-aberto-a-informacao-cientifica-no-brasil-um-estudo-das-universidades-publicas-do-estado-do-rio-de-janeiro/Tania\\_Relatorio.pdf](http://sitehistorico.ibict.br/capacitacao-e-ensino/pesquisa-em-ciencia-da-informacao/pos-doutorado/pesquisas-concluidas-1/acesso-aberto-a-informacao-cientifica-no-brasil-um-estudo-das-universidades-publicas-do-estado-do-rio-de-janeiro/Tania_Relatorio.pdf). Access on: Mar. 10, 2020.
- Dias, T. M. R. *Um estudo sobre a produção científica brasileira a partir de dados da Plataforma Lattes*. 2016. 181 f. Tese (Doutorado em Modelagem Matemática e Computacional) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.
- Digiampietri, L. A. *Análise da Rede Social Acadêmica Brasileira*. 2015. 160 f. Tese (Doutorado em Informação e tecnologia) – Universidade de São Paulo, São Paulo, 2015.
- Furnival, A. C. M.; Silva-Jerez, N. S. Percepções de pesquisadores brasileiros sobre o acesso aberto à literatura científica. *Informação & Sociedade*, v. 27, n. 2, p. 153-166, 2017.
- Leta, J.; Costa, E. H. S.; Mena-Chalco, J. P. Artigos em periódicos de acesso aberto: um estudo com pesquisadores bolsistas de produtividade do CNPq. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, v. 11, p. 1-6, 2017.
- Minniti, S.; Santoro, V.; Belli, S. Mapping the development of open access in Latin America and Caribbean countries: an analysis of Web of Science Core Collection and Scielo Citation Index (2005-2017). *Scientometrics*, v. 117, n. 3, p. 1905-1930, 2018. Doi: <http://dx.doi.org/10.1007/s11192-018-2950-0>.
- Mueller, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. *Ciência da Informação*, v. 35, n. 2, p. 27-38, 2006.
- Mueller, S. P. M. O círculo vícios o que prende os periódicos nacionais. *Datagramazero*, v. 0, n. 4, p.1-8, 1999.
- Neubert, P. S.; Rodrigues, R. S.; Goulart, L. H. Periódicos da Ciência da Informação em acesso aberto: uma análise dos títulos listados no DOAJ e indexados na Scopus. *Liinc em Revista*, v.8, n.2, p.389-401, 2012. Doi: <http://doi.org/10.18617/liinc.v8i2.497>.
- Organização para a Cooperação e Desenvolvimento Econômico. Declaration on Access to Research Data from Public Funding. Anexo 1 do Science, Technology and Innovation for the 21st Century. *Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level*. Paris: OCDE, 2004. p. 29-30. Available at: <https://www.oecd.org/sti/sciencetechnologyandinnovationforthe21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-financialcommunique.htm>. Accessed on: Oct. 19, 2018.
- Oliveira, E. C. P.; Chalhub, T.: O movimento de acesso livre à informação e repercussões nas revistas científicas Ibero-americanas. In: Foro Ibero-Americano de Comunicação e Divulgação Científica, 1., 2009, Campinas. *Anais [...]*. Campinas: Ibict, 2009. p. 1-7.
- Pavan, C.; Barbosa, M. C. Article processing charge (APC) for publishing open access articles: the Brazilian scenario. *Scientometrics*, v. 117, n. 2, p. 805-823, 2018. Doi: <http://doi.org/10.1007/s11192-018-2896-2>.
- Pavão, C. G.; Rocha, R. P.; Gabriel Junior, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 16, n. 2, p. 329-343, 2018.
- Rodrigues, R. S.; Oliveira, A. B. Periódicos científicos na América Latina: títulos em acesso aberto indexados no ISI e SCOPUS. *Perspectivas em Ciência da Informação*, v. 17, n. 4, p. 76-99, 2012.
- Silva, T. E.; Alcará, A. R. Políticas de acesso aberto à informação científica: iniciativas governamentais. In: Encontro nacional de pesquisa em Ciência da Informação, 9., 2008, São Paulo. *Anais [...]*. São Paulo: USP, 2008. p. 1-14.
- Taga, V. *Acesso aberto: classificação dos artigos indexados na Scopus*. 2016. 179 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Santa Catarina, Florianópolis, 2016.