XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING

Mapping Retractions with Open Data: A Brazilian Perspective

First A. Author[1], Second B. Author[1], Third C. Author[2]

[1]*Dept. of Something, University of Somewhere*
Address, Zip-Code, State/Province, Country
somebody1@somewhere.edu, somebody2@elsewhere.com
[2]*Dept. of Something Else, University of Somewhere Else*
Address, Zip-Code, State/Province, Country
somebody3@somewhere.edu

Abstract. This study investigates the phenomenon of scientific retractions in Brazil using an open data-based approach, integrating information from the Retraction Watch database and the Lattes Platform CVs. The main objective is to analyze the volume, temporal evolution, most recurrent topics and profiles of the authors involved in retracted articles indexed in national scientific production. To this end, a data cross-referencing methodology based on the digital identifier (DOI) was adopted, allowing the identification of records present in both platforms. The results show a significant growth in retractions in recent years, with emphasis on the year 2024, in addition to highlighting the concentration in areas such as Cell Biology, Biochemistry and Genetics. By using open access data sources and replicable methodologies, the study contributes to the principles of Open Science, promoting transparency, data reuse and the formulation of evidence-based public policies. The proposed analysis also provides support for strengthening scientific integrity, by alerting to the persistence of the circulation of retracted articles and the importance of effective systems for correcting and disseminating this information.

Keywords: Retractions, Scientific Production, Lattes Platform.

1 Introduction

The phenomenon of scientific article retractions has been widely discussed in international academic literature, largely due to growing concerns about research integrity and the implementation of more rigorous ethical oversight mechanisms. Retractions generally result from methodological or technical errors, or unethical behavior such as plagiarism, manipulation, or fabrication of data. As [1] explains, the retraction process is formalized post-publication when serious flaws that compromise the validity of the conclusions presented are detected.

However, the correction of these articles is not always effective in practice. Many retracted articles continue to be cited neutrally or even without indication of invalidation, which contributes to the perpetuation of misinformation in the scientific literature. In response to these concerns, Retraction Watch has emerged as an essential global initiative focused on collecting, cataloging, and analyzing retracted articles. Since its creation in 2010, the platform has played a fundamental role in raising awareness about the impact of retractions and promoting a culture of transparency in science. Using databases like Retraction Watch allows for the identification of flaws and aids in the development of preventive and corrective policies that can minimize the negative effects of retractions. [2] advocates, for example, the adoption of technologies that automatically track invalidated articles, strengthening control over academic integrity.

Given the above, this study aims to investigate, through bibliometric analysis, the main causes and characteristics of article retractions in Brazilian scientific production, as well as the profiles of their authors. Based on data extracted from Retraction Watch and supplementary information from other sources, such as CVs from the Lattes Platform, we hope to obtain a comprehensive and unprecedented portrait of scientific retractions in the country.

2 Methodology

This study used data related to scientific retractions and CVs registered on the CNPq Lattes Platform. Regarding retraction data, after Crossref acquired Retraction Watch, the data became available in an open, standardized, and free format. This change allowed for simpler and more transparent access to retraction information.

Regarding CV data from the Lattes Platform, the LattesDataXplorer framework [3] was used to extract information of interest for the research, particularly data on scientific publications and author profiles. After extracting and collecting the data from both databases, it is necessary to perform a link to identify the retracted articles contained in the Lattes Platform.

With both datasets in hand, a cross-reference can be performed to identify the retracted articles registered in the Lattes Platform CV database. For this identification, the article's DOI from the Retraction Watch database was used, as well as its verification against the DOI list of all publications extracted from the CVs. For each article identified in both databases, all information was grouped into a single record to create a data extract for later analysis.

The choice to use open data, from both Retraction Watch and the Lattes Platform, reinforces the commitment to the principles of Open Science, promoting reproducibility and transparency of research. Furthermore, it demonstrates the potential for reusing scientific data for analyses that directly impact scientific integrity policies.

3 Results

Data on retractions were collected from Crossref's GitLab database. The dataset was collected in January 2025, totaling 59,531 records. Resumes from the Lattes Platform were collected in December 2024, allowing for the analysis of approximately 8.6 million resumes.

After linking the data between the databases, it was possible to analyze the retractions present in the Lattes Platform resumes. A total of 895 retractions were identified, with 2024 accounting for the largest number of articles, with a total of 249 records (Figure 1).

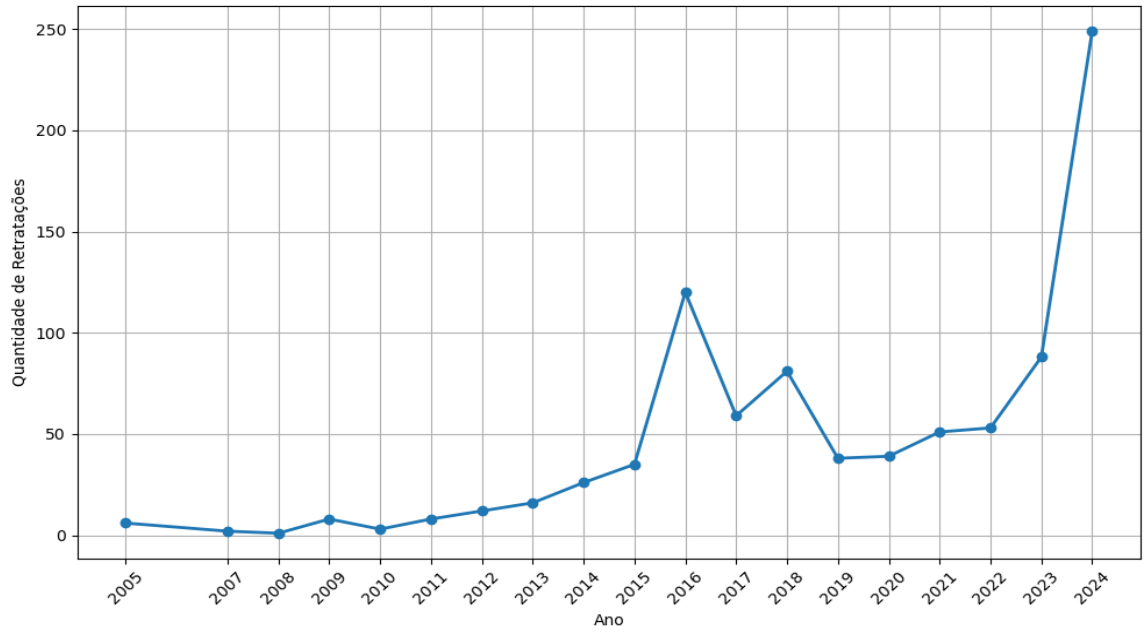


Figure 1 – Temporal evolution of the number of articles retracted in the Lattes Platform resumes

As can be seen, significant growth has occurred over the years, peaking in 2024. This increase may be associated with increased monitoring and open availability of data on retractions, strengthening Open Science practices by making this information accessible and auditable.

Retractions by research topic refer to the categorization of scientific article retractions based on the areas of knowledge covered in the original studies. Each article can be associated with one or more disciplines, such as biology, chemistry, physics, or social sciences, reflecting the central theme of the research. In total, 84 topics were identified across different areas of knowledge. Figure 2 presents the ten subjects with the highest number of retractions. Among them, the following stand out: Cell Biology (260), Biochemistry (229), Zoology (196), Genetics (163), and Toxicology (160).

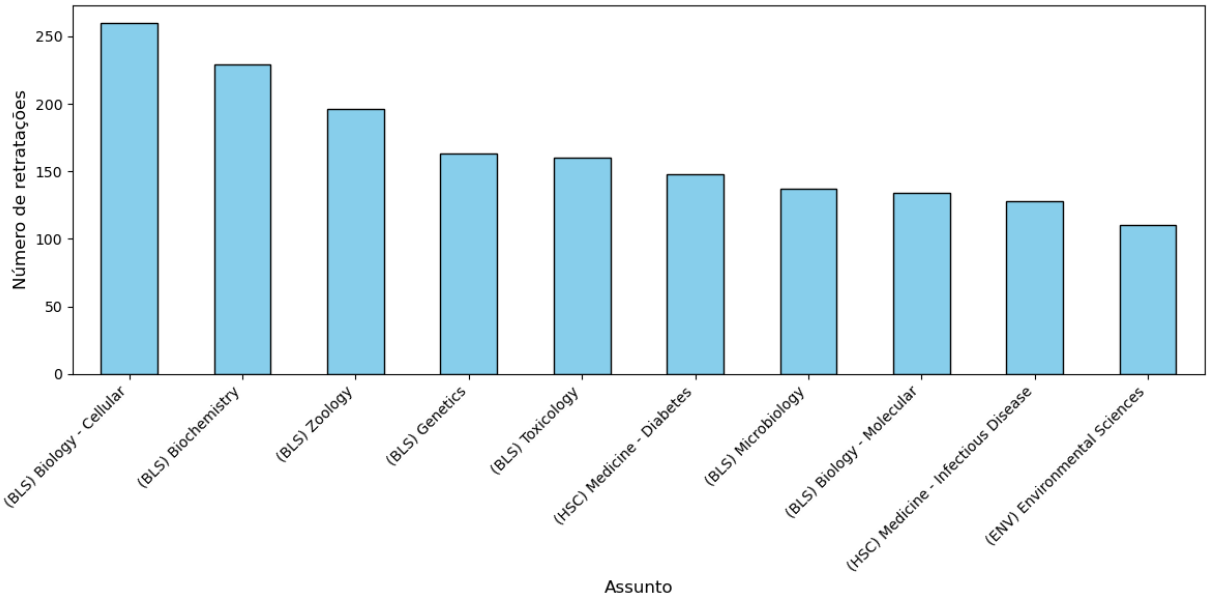


Figure 2 – Topics with the highest number of retractions.

The ten research topics with the highest number of articles covered are highlighted, with emphasis on the areas of Cell Biology, Biochemistry, and Genetics. The predominance of these areas suggests the need for more robust scientific integrity policies in these fields. Openly making these data available favors the development of preventive strategies targeting more susceptible areas, reinforcing the principles of Open Science.

4 Conclusions

This study revealed that, although scientific retractions represent an important corrective action, they still face challenges in their effective dissemination and recognition within the scientific community. The integration of open data, such as that provided by Retraction Watch and the Lattes Platform, enables more transparent and comprehensive analyses of national scientific production, aligning with the principles of Open Science. By identifying patterns, fields, and author profiles, the work contributes to strengthening academic integrity policies and reinforces the importance of transparency and open access as pillars of scientific credibility. Such actions are essential to prevent the perpetuation of errors and promote a more ethical, responsible, and accessible scientific culture.

References

[1] Silveira, E. A vida após a morte da ciência invalidada. Revista Questão de Ciência (2022). Disponível em: https://www.revistaquestaodeciencia.com.br/questao-de-fato/2022/05/19/vida-apos-morte-da-ciencia-retratada. Acesso em: 18 fev. 2025.

[2] Andrade, R. O. “Artigos Zumbis”. Pesquisa FAPESP, nº 322, dezembro de 2022, pp. 32-35.

[3] Dias, T. M. R. (2016). Um Estudo Sobre a Produção Científica Brasileira a partir de dados da Plataforma Lattes. 2016. 181 f. Tese (Doutorado) - Curso de Programa de Pós-graduação, Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte.