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

Infrastructure for Open Science: analysis of the national coverage of persistent identifiers in Brazil

Davi F. S. Guimarães[1], Thiago M. R. Dias[2], Patrícia Mascarenhas Dias[3], Dênnyo A. Souza[1], Guilherme M. Dias[4].

[1]*Technical High School Program, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)*
Av. Amazonas, 5.253, Nova Suiça, Belo Horizonte, 30.421-169, Minas Gerais, Brazil
guimaraes.davi.fs@gmail.com, dennyoarcanjo@gmail.com
[2] *Dept. of Computing, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)*
Av. Amazonas, 5.253, Nova Suiça, Belo Horizonte, 30.421-169, Minas Gerais, Brazil
thiagomagela@cefetmg.br
[3] *Dept. of Computing, Universiade Estadual de Minas Gerais (UEMG)*
Av. Presidente Antônio Carlos, 7545, São Luiz, Belo Horizonte, 31.275-083, Minas Gerais, Brazil
patriciamdias@gmail.com
[4] *Dept. of Computing, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)*
Av. Amazonas, 5.253, Nova Suiça, Belo Horizonte, 30.421-169, Minas Gerais, Brazil
guilhermedias2501@gmail.com

Abstract. The adoption of persistent identifiers is one of the pillars for the advancement of Open Science, as it promotes interoperability, traceability, and reproducibility of research. This article presents a quantitative analysis of the presence of persistent identifiers in CVs registered on the Lattes Platform, focusing on ORCID for authors and DOI for publications. More than 9 million CVs were analyzed, revealing that only 377,385 researchers had informed their ORCID. Regarding publications, of the 12.8 million publications in journals, 46.6% contained DOI; this percentage was significantly lower in articles published in conference proceedings (2.4%), book chapters (14.3%), and books (6.4%). The results point to the need for greater institutional and cultural incentives for the use of persistent identifiers, as an essential condition for strengthening Open Science in Brazil.

Keywords: Persistent Identifiers. Scientific Production, Lattes Platform.

1 Introduction

Open Science has established itself as a new paradigm in scientific development and communication, promoting transparency, collaboration, and unrestricted access to data, methods, and research results. In a context where scientific production grows exponentially and the challenges related to reproducibility and data integrity become increasingly complex, the adoption of open practices emerges as an essential strategy for strengthening the research ecosystem. In this scenario, persistent identifiers play a fundamental role. They function as unique and persistent markers that allow for the precise tracking of authors and publications over time, even in the face of institutional changes, name variations, and repository reorganizations.

Among the main persistent identifiers used in the scientific community are ORCID (Open Researcher and Contributor ID), intended for identifying researchers, and DOI (Digital Object Identifier), used to identify scientific publications and digital objects in general. Although they are distinct technologies, ORCID and DOI complement each other within a common governance and standardization framework for scientific communication [1]. These mechanisms enable automated integration between systems, facilitate authorship recognition, prevent duplicate records, and promote interoperability between different databases and digital platforms. In Brazil, however, the adoption of these identifiers is still limited and underexplored. Therefore, the research questions what

is the level of national coverage of persistent identifiers in Brazil?

Therefore, this study aims to map the insertion of persistent identifiers by Brazilian researchers in data repositories, with special attention to the presence of ORCID in CVs registered on the Lattes Platform and DOI in registered publications. The relevance of this research lies in its ability to reveal the current status of adoption of these instruments in Brazil and, from this, contribute to actions aimed at consolidating a more transparent, connected, and interoperable scientific infrastructure. and understanding how researchers have used them is essential to guide public policies, institutional strategies and the development of technologies aimed at Open Science.

2 Methodology

To achieve the proposed objective, a large-scale analysis was conducted of the curriculum data set from the Lattes Platform, one of the largest and most important databases of academic and scientific curricula in Brazil [2], maintained by the National Council for Scientific and Technological Development (CNPq). The analyzed database comprises over 9 million curricula, covering researchers from different fields of knowledge, levels of training, and institutional affiliations.

The methodological process involved the automated extraction of structured information from the curricula using the LattesDataXplorer framework [2], with a specific focus on two central points: (1) the presence of ORCID identifiers in the authors' profiles, and (2) the presence of DOIs in registered publications. For this final analysis, publications were categorized into four types: articles in scientific journals, articles in conference proceedings, book chapters, and complete books. After data collection, a quantitative analysis was conducted to identify the frequency of persistent identifier insertion across different categories and profiles.

This methodological approach allowed us not only to verify the presence of identifiers but also to observe adoption patterns across different types of publications, providing a comprehensive overview of the Brazilian reality regarding the use of these Open Science support tools.

3 Results

The results reveal a still-incipient adoption of persistent identifiers among Brazilian researchers, especially with regard to ORCID. Of the more than 9 million CVs analyzed, only 377,385 had a registered ORCID identifier, representing less than 5% of the total. This data highlights that most researchers have not yet incorporated this important identification mechanism into their academic profiles, which can negatively impact the visibility, traceability, and interoperability of their scientific output.

A total of 12,881,308 articles were identified in scientific journals, of which 5,998,784 (approximately 46.6%) had a DOI provided. This result indicates that almost half of publications of this type already adopt the persistent identifier, reflecting the journal editorial system's greater adoption of standardization and data tracking practices.

Regarding scientific publications, the figures also point to partial inclusion of the DOI identifier, with significant variations between publication types (Figure 1).

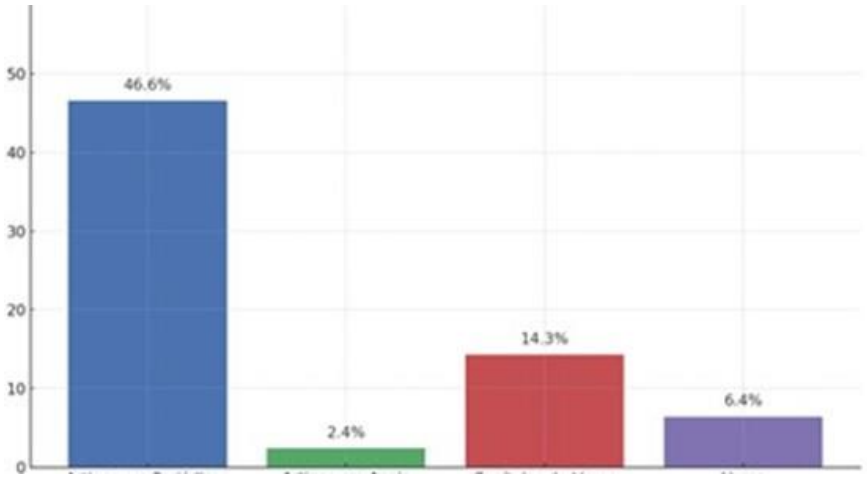


Figure 1 – Number of publications with DOI.

Figure 1 presents the percentage distribution of the different types of bibliographic production identified in the analyzed set. It can be seen that articles published in scientific journals represent the majority of the production, corresponding to 46.6% of the total. Next, book chapters stand out, with 14.3%, and complete books, with 6.4%. Articles published in conference proceedings correspond to only 2.4% of the total. These results indicate a predominance of publications in scientific journals, which may reflect both the appreciation of this type of production in academic evaluation processes and the greater visibility and impact these publications tend to provide.

Figure 2 presents the number of publications with and without DOI in temporal order.

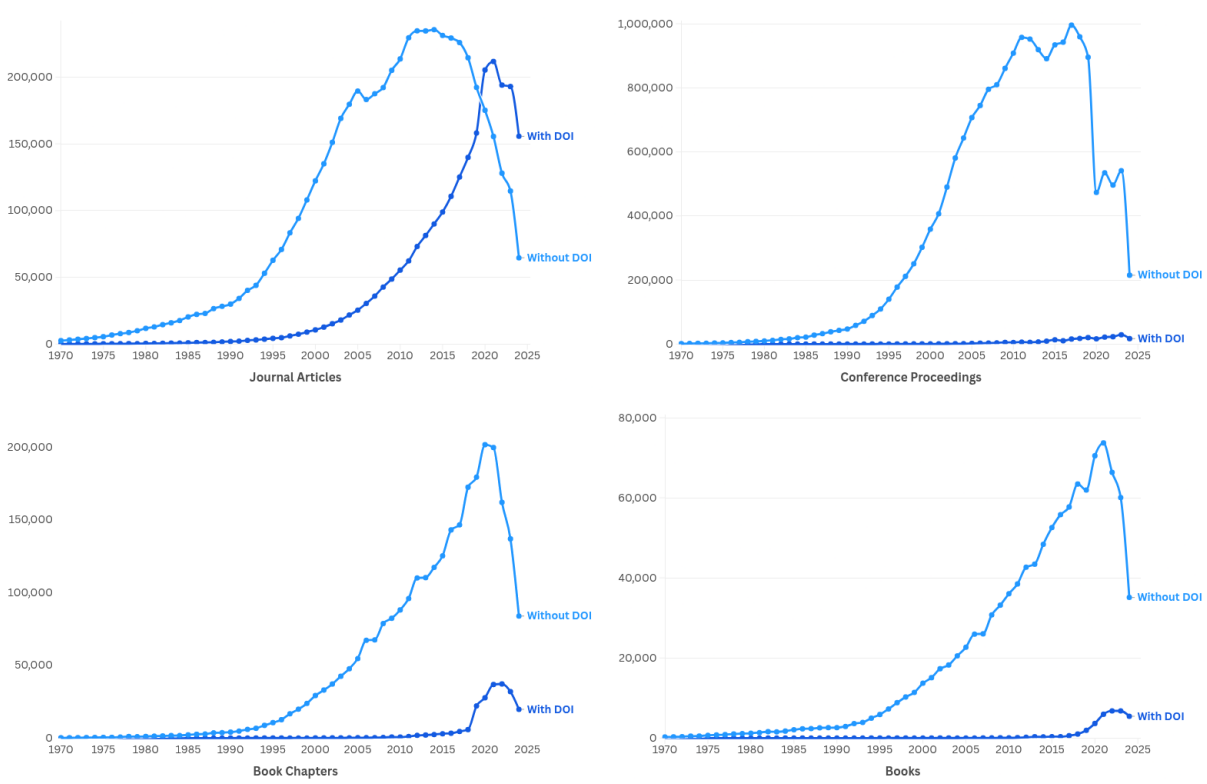


Figure 2 – Number of publications with and without DOI per year.

Figure 2 shows the evolution of the number of scientific publications between 1970 and 2025, differentiated by the presence or absence of a persistent DOI (Digital Object Identifier), for four document types: journal articles, conference proceedings, book chapters, and books.

Across all document types, there was a significant increase in publications starting in the 1990s, followed by a sharp decline after 2020, possibly associated with the lag in indexing and updating the most recent databases.

In journal articles, there was a strong predominance of records with DOI starting in 2010, reflecting the consolidation of this identifier as an international standard for scientific publications. A similar pattern, albeit on a smaller scale, is observed in book chapters and full books, where the use of DOIs only became more prominent in the 2010s.

On the other hand, conference proceedings exhibited a distinct pattern, with a low proportion of records with DOIs throughout the period, highlighting the uneven adoption of this identifier across different types of scientific production.

These results demonstrate that, although DOIs are widely consolidated in periodicals, their adoption in other types of documents (especially proceedings and books) is still limited and recent, reflecting different editorial policies and levels of internationalization in scientific publishing practices.

An analysis of the adoption of the persistent Orcid identifier and the level of training of Brazilian researchers can be seen in Figure 3..

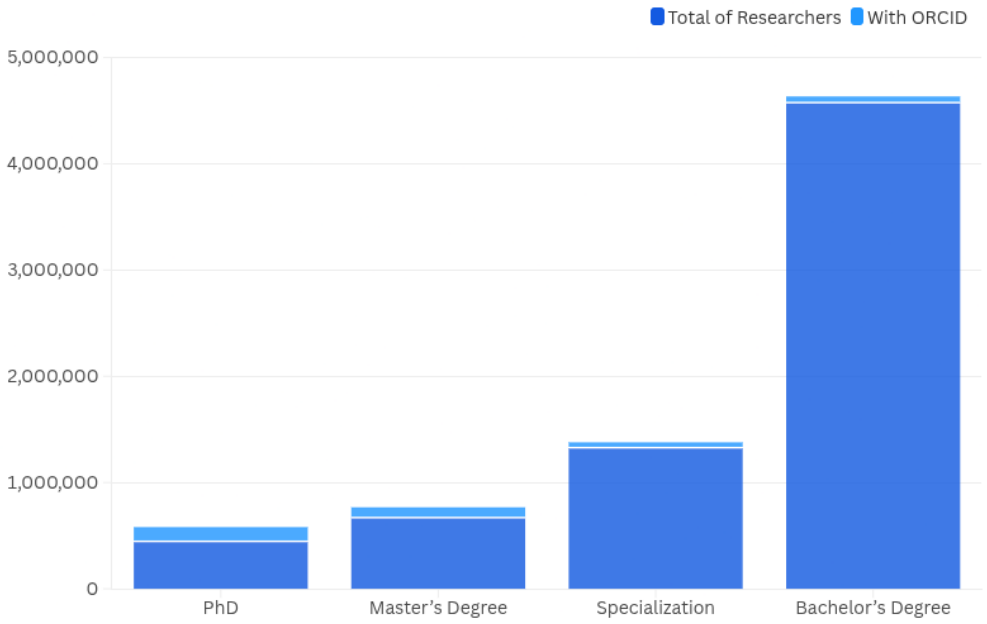


Figure 3 – ORCID adoption by highest degree.

Figure 3 shows the distribution of researchers at different levels of education (Doctorate, Master's, Specialization, and Undergraduate), also indicating the proportion of those with an ORCID identifier.

The largest concentration of researchers is at the Bachelor's Degree level, significantly higher than at other levels. Next are the groups with Specialization, Master's, and Doctorate degrees, respectively.

Despite the high total number of researchers at all levels, the proportion with an ORCID registration is relatively small, as evidenced by the narrow upper band on each bar. ORCID registration is slightly more significant among researchers with Doctorates and Master's degrees, suggesting greater adoption among those more involved in research and scientific publishing activities.

These data reinforce the importance of encouraging ORCID registration and use as a fundamental practice for persistent author identification, academic data integration, and interoperability between scientific information systems, especially among researchers at early and intermediate levels of education.

4 Conclusion

The analysis presented in this study shows that the adoption of persistent identifiers in Brazil, while growing, is still far from reaching ideal levels that would foster full integration into the Open Science ecosystem. The low incidence of ORCID in academic curricula and the limited presence of DOIs in publications, especially in conference proceedings, chapters, and books, reveal the urgent need for coordinated institutional actions to promote the systematic use of these tools.

Persistent identifiers are structuring elements for interoperability between data repositories, facilitating the integration, reuse, and reliability of scientific information. Their adoption directly contributes to avoiding problems such as authorship ambiguity, duplicate records, and loss of links between publications and authors. Furthermore, they enable the development of more accurate metrics, favoring the evaluation of scientific production in a fairer and more transparent manner.

In this sense, we recommend strengthening public policies focused on Open Science, training researchers on the use of identifiers, and encouraging integration between institutional platforms and open data systems. Future work will deepen the analysis of the presence of persistent identifiers in the country by gender and areas of knowledge.

5 Acknowledgments

This work was supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

References

[1] Béquet, G. (2022). Identifiers and open science: Elements for a socio-technical analysis of ORCID. International Journal of Knowledge Engineering and Management, Florianópolis, v. 11, n. 29, p. 125-16

[2] Lane, J. (2010). Let's make science metrics more scientific. Nature, v. 464, n. 7288, p. 488-489.

[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.