

Evolution in ontology research in Brazil: A metric study of ONTOBRAS series

Sara Maciel Soares^{1,2,†} and Fernanda Farinelli^{1,2,*†}

¹ Faculty of Information Science, University of Brasilia, Campus Darcy Ribeiro - DF, Brasilia, Brazil - 70297-400

² Brazilian Institute of Science and Technology, SAUS Q 5, L 6, Bl H, Brasília, DF.

Abstract

The Seminar on Ontology Research in Brazil (ONTOBRAS), in its 17th edition in 2024, provides a space for the exchange of knowledge about theories, methodologies, and applications of ontologies. It attracts researchers and professionals in the areas of Information Science and Computer Science to share their research and acquire new knowledge. However, there is a hiatus regarding the investigation of research presented throughout the editions of ONTOBRAS. This work seeks to understand the context of the scientific production of this seminar series. The study identified the most influential authors, the main between these authors and their institutions, and mapped the evolution of scientific production on ontologies over the seminar editions. A more comprehensive work, detailing these findings, will be published soon.

Keywords

Bibliometrics, Metric Studies, Scientific Communication, Scientific Production on Ontologies.

1. Introduction

Ontology studies focus on theories, methodologies, and practices for presenting a set of concepts and their relationships within a specific domain. It plays a key role in organizing, managing, and retrieving knowledge while aiding human and machine information processing. This subject of study is essential across various fields, including computer science, information science, philosophy, and healthcare, where it addresses challenges related to knowledge modeling, retrieval, representation, and classification [1].

The Seminar on Ontology Research in Brazil (ONTOBRAS) aims to exchange knowledge about ontology theories, methodologies, and applications. It brings together researchers and professionals, mainly from information and computer science, to share findings and explore new developments in ontology. Despite its role in bridging theory and practice in ontology research, a comprehensive analysis of ONTOBRAS's contributions is still needed. Key questions remain about the most influential authors, participating institutions, collaborative networks, and the evolution of scientific output across its editions.

The evolution of scientific output at ONTOBRAS remains underexplored, with limited analysis of shifts in research interests and interdisciplinary collaboration. Understanding these dynamics is key to anticipating trends and fostering growth. This paper synthesizes a part of the bibliometric analysis of ONTOBRAS contributions from 2011 to 2023. It identifies key trends by focusing on the most productive authors, institutions, and collaborative networks. Through this analysis, the paper provides insights into the development and evolution of ontology research in Brazil.

Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS 2024) and 8th Doctoral and Masters Consortium on Ontologies (WTDO 2024), Vitória, Brazil, October 7-10, 2024

* Corresponding author.

† These authors contributed equally.

✉ saramacisoares@gmail.com (S. M. Soares); fernanda.farinelli@unb.br (F. Farinelli)

ORCID 0009-0002-9701-9375 (S. M. Soares); 0000-0003-2338-8872 (F. Farinelli)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Theoretical Background

This section presents the theoretical framework, emphasizing key concepts in metric studies, particularly bibliometric and scientometric indicators. Bibliometrics and scientometrics provide tools to measure and evaluate the scope and impact of research, using quantitative data to identify publication patterns, knowledge dissemination, collaboration networks among researchers, institutions, and countries, and trends across different fields. The foundational principles of these studies guide the use of indicators such as publication counts and collaboration networks to evaluate productivity and impact across various scientific domains.

Scientific development generates a vast body of records that validate knowledge and support the flow of scientific production, forming the foundation for new research. This process has been accelerated by technological advancements, incorporating new methods of production and communication into science. Studies in information metrics focus on visualizing, analyzing, and evaluating the evolution of scientific activity and its outputs, aiding in planning, monitoring, and assessing scientific processes [2-4].

Combining principles from the sociology of science with methodologies from mathematics, statistics, and computing [2], these metrics create quantitative indicators to map and visualize areas of study [5], supporting decision-making and policy development. Bibliometrics, which studies the quantitative aspects of human knowledge production, and scientometrics, which quantifies scientific output and communication [6], are key subfields. Both are grounded in classic bibliometric laws, such as Bradford's, Lotka's, and Zipf's, which form the basis for analyzing document corpora.

Oliveira [5] emphasizes that metrics must be interpreted within their specific contexts and can be enhanced with epistemological, historical, and social insights. Science metrics analyze both inputs (resources enabling research) and outputs (scientific products). Output indicators, or productivity indicators, assess research dissemination and societal impact [3]. Recent years have seen a growing interest in these indicators, as they inform decisions on resource allocation and the development of public and institutional scientific policies [3,5].

3. Methodology

This study employed a descriptive and quantitative analysis using bibliometric and scientometric indicators to map the scientific production and trends related to ontologies presented at ONTOBRAS. The dataset included all papers from the 12 editions of the seminar (2010-2023), except for 2014 when the event was not held. The proceedings were published in the CEUR Workshop Proceedings, and the full list of proceedings can be accessed via the ONTOBRAS website.

Data collection involved downloading all available papers from these sources. Institution-related data and keywords were manually extracted from the papers. For the 2022 and 2023 editions, keywords were provided, while for earlier editions, they were generated from titles using natural language processing techniques. After data standardization, bibliometric and scientometric methods were applied to analyze articles, authors, institutions, key topics, and collaboration patterns.

The KNIME Analytics Platform (version 5.2) [7] was used for data collection and authorship network visualization. Python scripts generated graphs and performed additional analyses. Bibliometric and scientometric analyses, along with visualizations, were done using KNIME, Python, and Excel. The process involved: 1) collecting data from the CEUR-WS repository and ONTOBRAS website, 2) manually extracting institution data, 3) standardizing authorship, 4) creating visualizations, and 5) analyzing the data and indicators.

4. Results and discussion

This section summarizes the study's results, highlighting key data and its implications.

4.1. Works presented in 12 editions

A total of 301 papers were published in the ONTOBRAS proceedings, with 252 from the main track (Ontobras) and 49 from the WTDO, which started in 2017. The average number of papers per edition was 25.08, with a median of 24. The average number of papers per edition was 25.08, with a median of 24. The highest number of papers, 33, was presented in 2012 and 2018. However, there has been a significant decline since 2022, with only 14 papers published in 2023 (nearly 10 below the median), marking the lowest count in the event's history.

The decline in papers may partly stem from the lasting effects of the COVID-19 pandemic, which disrupted academic productivity and shifted research priorities, limiting collaboration and paper submissions [10-12]. Despite stable output in 2020 and 2021, the long-term impacts likely contributed to reduced submissions in recent editions.

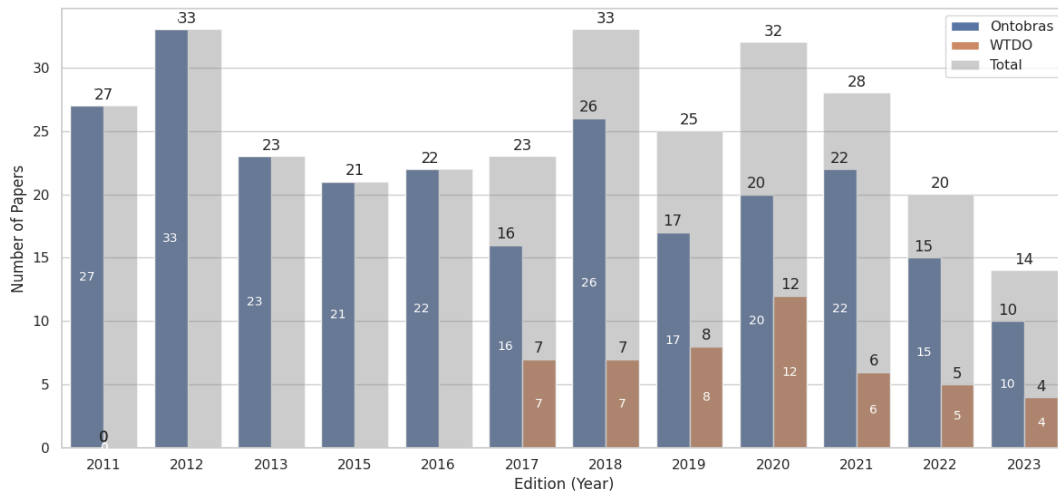


Figure 1: Number of papers per edition and track, from authors (2024).

4.2. Institutions

A total of 127 institutions contributed to ONTOBRAS over 12 editions, with 55.12% publishing only one paper. Brazilian institutions produced 78.8% of the papers, with the southern and southeastern regions leading, as 8 of the 11 most productive institutions are from these regions. The top third of institutions contributed 78.5% of the total output. Table 1 lists the 10 most productive institutions. ONTOBRAS's international reach is reflected in the participation of 36 foreign institutions, responsible for 64 papers (21.2%). Companies like Petrobras (5 papers), A. C. Camargo Cancer (2), IBM (2), and public bodies such as the Maranhão State Treasury (2) and Belo Horizonte City Hall (1) also contributed.

Table 1

The 10 most productive institutions in the ONTOBRAS series

Institution	Frequency per year													F	%	%AC
	2011	2012	2013	2015	2016	2017	2018	2019	2020	2021	2022	2023				
UFES	1	3	3	3	6	1	6	3	8	7	2	2	46	15,28	15,28	
UFRGS	8	3	4	4	0	0	3	3	3	5	3	1	38	12,62	27,91	
UFMG	4	4	2	1	0	3	5	3	6	2	1	0	32	10,63	38,54	
USP	2	2	4	1	1	4	2	1	4	1	2	0	24	7,97	46,51	

UFRJ	1	2	1	1	1	0	1	1	2	1	3	1	15	4,98	51,50
UNIRIO	0	4	1	0	1	3	3	0	1	2	0	0	15	4,98	56,48
UFSC	2	2	2	0	2	0	1	2	0	1	0	0	12	3,99	60,47
UnB	0	1	0	0	3	1	3	1	0	1	0	2	12	3,99	64,45
UFBA	3	0	0	1	1	1	2	1	1	0	1	0	11	3,65	68,11
PUCRS	2	2	0	3	1	2	0	1	0	0	0	0	11	3,65	71,76

N= 301. From authors (2024)

An analysis of institutional collaboration reveals that 174 papers (57.80%) were authored within the same institution, while 127 (42.20%) involved inter-institutional collaboration. The average number of institutions per paper was 1.55, indicating limited cross-institutional work. Of the 64 papers with foreign institutions, 30 (46.7%) featured international collaboration, with significant contributions from UFES (10 collaborations) and PUCRS (6 collaborations). Most papers (96.35%) were collaborative but typically within the same institution, suggesting potential for more inter-institutional partnerships. This is typical of emerging fields like applied ontology, where collaboration enhances research visibility and accuracy. The Southeast region dominates production, led by UFES, contributing 15.3% of the total output

4.3. Keywords cooccurrence

Figure 2 presents the most frequent n-grams (unigrams, bigrams, and trigrams) extracted from the titles of papers presented at ONTOBRAS over the years. Since keywords were only available for the 2022 and 2023 editions, the analysis focused on titles to identify recurring topics and trends across earlier editions.

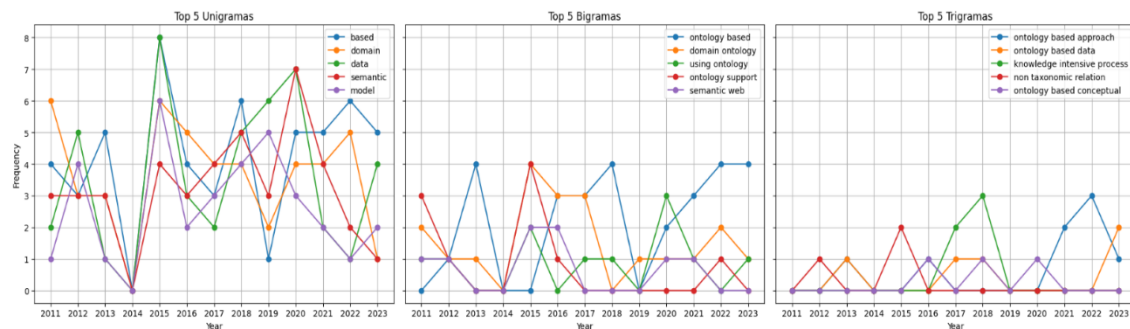


Figure 2: Top 5 frequent unigrams, bigrams, and trigrams across ONTOBRAS, from authors (2024)

The results highlight a strong focus on semantic integration and the practical use of ontologies. Including keywords in future ONTOBRAS editions will improve discoverability, facilitate indexing, and enable more effective tracking of trends and gaps through network visualizations. This will foster collaboration, guide future research, and contribute to a more structured and interconnected body of knowledge, ultimately enhancing the visibility and impact of ontology research.

4.4. Authorship and author collaboration

The 301 papers presented at ONTOBRAS were authored by 573 individuals. Table 2 highlights 13 authors who contributed to 10 or more papers, with the top 7 authors responsible for over a third of the total publications. The median number of papers per author was 1, with an average of 1.72. The most productive third of the authors contributed 61.2% of all papers. On average, 70.9 authors participated in each ONTOBRAS edition, with a median of 69. However, 2022 and 2023 saw the lowest number of contributing authors, reflecting a decline in paper submissions.

Table 2

Authors with participation in 13 or more papers about the total number of papers analyzed

Author	F	%	%AC
Mara Abel	27	8,97%	8,97%
Maurício Barcellos Almeida	16	5,32%	14,29%
Fernanda Araújo Baião	15	4,98%	19,27%
Giancarlo Guizzardi	13	4,32%	23,59%
Joel Luís Carbonera	13	4,32%	27,91%
Maria Luiza Machado Campos	12	3,99%	31,89%
Monalessa Perini Barcellos	12	3,99%	35,88%
João Paulo Andrade Almeida	11	3,65%	39,53%
Marcello Peixoto Bax	11	3,65%	43,19%
Fabício Henrique Rodrigues	10	3,32%	46,51%
Renata Silva Souza Guizzardi	10	3,32%	49,83%
Renata Vieira	10	3,32%	53,16%
Ricardo de Almeida Falbo	10	3,32%	56,48%

N= 301. From authors (2024)

Collaboration is a key driver of paper production at ONTOBRAS. Of the 301 papers, only 11 (3.65%) were single-authored, with 96.35% involved collaboration. The average number of authors per paper was 3.27, with a median of 3, reflecting a high rate of co-authorship, often within the same institution.

The cumulative analysis shows a concentration of contributions, with Mara Abel alone responsible for 8.97% of all papers. The top 13 authors contributed 56.48%, highlighting the dominance of a small group in the field. Figure 3 maps the author network, showing the collaborations of at least 3 co-authors in ONTOBRAS editions. The Louvain method identified 67 collaboration communities, with the largest, involving 70 authors and 180 collaborations, centered around Mara Abel, emphasizing her strong influence on ontology research at the event.

Scientific collaboration enhances the visibility, completeness, and accuracy of research, and is a sign of a field's maturity [8], especially in the emerging area of applied ontology. As collaboration increases, researchers bridging smaller institutional networks will be pivotal in expanding the field. Figueiredo and Almeida [9] underscore the importance of information flow within these networks for advancing ontology research.

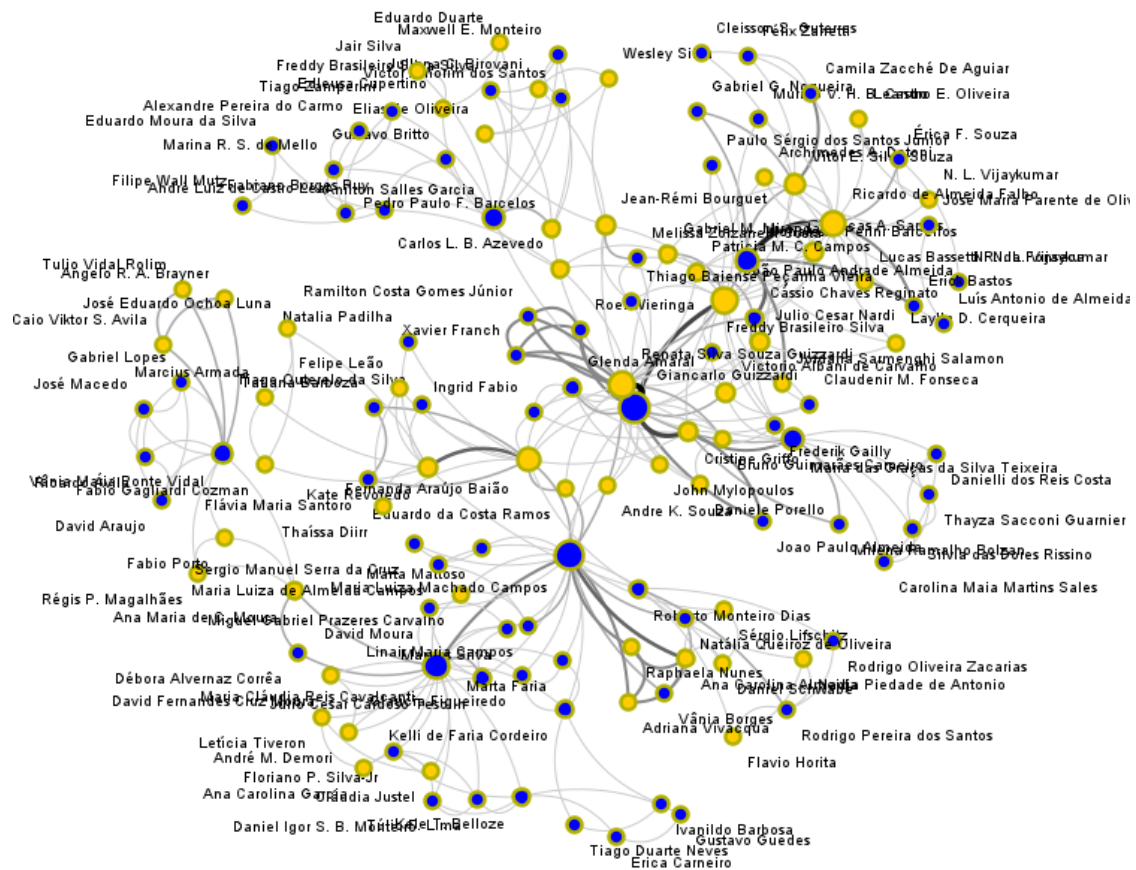


Figure 3: Authors connected with at least 3 other authors' network, from authors (2024).

5. Final remarks

The 17th Brazilian Ontology Research Seminar in 2024 solidifies its position as the country's leading event for ontology research, driving the field forward. delve deeper into emerging topics presented at ONTOBRAS, identify evolving production trends, and perform citation analyses to map foundational literature and networks of influence. Additionally, examining the recent decline in submissions may provide insights into whether this is a reflection of internal event dynamics or indicative of a broader stagnation in ontology research in Brazil, offering potential strategies for revitalizing participation.

This study sheds light on key aspects of ontology production at ONTOBRAS, from identifying the most productive authors and institutions to mapping regional concentrations, collaboration networks, and international participation. An expanded study in development will provide further details on these findings. These findings significantly contribute by outlining the formal and informal communication pathways that shape ontology research, guiding both researchers and organizers in enhancing collaboration, visibility, and impact within this growing field.

Acknowledgements

We would like to extend our heartfelt thanks to the University of Brasília (UnB) for their generous funding support for our research.

References

- [1] F. Farinelli and M. B. Almeida, "Ontologias biomédicas: teoria e prática," in *Livro de Minicursos do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2019)*, Niterói, RJ: Sociedade

- Brasileira de Computação, 2019, pp. 93–140. Accessed: Aug. 14, 2024. [Online]. Available: <https://books-sol.sbc.org.br/index.php/sbc/catalog/view/29/97/247>
- [2] M. C. C. Grácio, Estudos métricos da informação, In: Análises relacionais de citação para a identificação de domínios científicos: uma aplicação no campo dos Estudos Métricos da Informação no Brasil, Oficina Universitária, Marília, SP, Cultura Acadêmica, São Paulo, SP, 2020, pp. 19–75. doi: 10.36311/2020.978-65-86546-12-5.
- [3] D. P. Noronha, J. M. Maricato, Estudos métricos da informação: primeiras aproximações, *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Esp. (2008) 116–128. URL: <https://www.redalyc.org/articulo.oa?id=14709810>.
- [4] C. A. A. ARAÚJO, Bibliometria: evolução histórica e questões atuais, *Em Questão* 12 (2006) 11–32. URL: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/16>.
- [5] E. F. T. Oliveira, Estudos métricos da informação no Brasil: indicadores de produção, colaboração, impacto e visibilidade, Oficina Universitária, Marília, SP, Cultura Acadêmica, São Paulo, SP, 2018. doi: 10.36311/2018.978-85-7983-930-6.
- [6] L. Bufrem, Y. Prates, O saber científico registrado e as práticas de mensuração da informação, *Ciência da Informação* 34 (2005), pp. 9–25. doi: 10.1590/S0100-19652005000200002.
- [7] F. Farinelli, Revolucionando a pesquisa científica com a plataforma knime analytics, in: *Tecnologias utilizadas em pesquisas acadêmicas em Ciências Sociais Aplicadas*, IBICT, Brasília, DF, 2023, pp. 275–326. doi: 10.22477/9786589167938cap10.
- [8] C. M. Hilário, M. C. C. Grácio, Colaboração científica na temática “redes sociais”: análise bibliométrica do Enancib no período 2009 – 2010, *Revista EDICIC* 1 (2011), pp. 363–375. URL: <http://hdl.handle.net/11449/115334>.
- [9] F. C. Figueiredo, F. G. Almeida, Ontologias em Ciência da Informação: um estudo bibliométrico no Brasil, *Ciência da Informação* 46 (2017), pp. 23–33. doi: <https://doi.org/10.18225/ci.inf.v46i1.4011>
- [10] G. Abramo, C. A. D’Angelo, and I. Mele, “Impact of Covid-19 on research output by gender across countries,” *Scientometrics*, vol. 127, no. 12, pp. 6811–6826, Dec. 2022, doi: 10.1007/s11192-021-04245-x.
- [11] K. R. Myers *et al.*, “Unequal effects of the COVID-19 pandemic on scientists,” *Nat Hum Behav*, vol. 4, no. 9, pp. 880–883, Sep. 2020, doi: 10.1038/s41562-020-0921-y.
- [12] UNESCO, “COVID-19: reopening and reimagining universities, survey on higher education through the UNESCO National Commissions - UNESCO Digital Library,” ED/E30/HED/2021/01. Accessed: Sep. 08, 2024. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000378174>