SBC Computing Reviews, 2026, 05:1
ISSN: 2966-3938 • doi: 10.5753/rocs.2026.6779

REVIEW

Women in Business Process Management: a Systematic Mapping

Thayna Minuzzo [Universidade Federal do Rio Grande do Sul | *thaynaminuzzo@gmail.com*]
Guilherme Rego Rockembach [Universidade Federal do Rio Grande do Sul | *guilherme.rockembach@ufrgs.br*]
Rachelle Bianchi Sganderla [Universidade Federal do Rio Grande do Sul | *kelly.sganderla@iprocess.com.br*]
Pedro Vitor Alexandre [Universidade Federal do Rio Grande do Sul | *pedrovit0r.me@gmail.com*]
Encarna Sosa Sánchez [University of Extremadura | *esosa@unex.es*]
Lucinéia Heloisa Thom [Universidade Federal do Rio Grande do Sul | *lucineia@inf.ufrgs.br*]

Instituto de Informática, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500 - Agronomia, Porto Alegre - RS, 91509-900, Brazil.

Abstract. In recent years, Business Process Management (BPM) has increasingly incorporated non-technical dimensions, recognizing cultural and social factors as critical to the success of BPM projects. Against this backdrop, this study investigates women's participation in the BPM research domain, a field where their presence has traditionally been lower, especially in technology-related areas. Employing a systematic mapping methodology, the authors analyzed 5,029 scientific papers to assess patterns of authorship. The results demonstrate that women's scientific production in BPM remains considerably below that of men, accounting for less than one-third of all identified authors. Moreover, this disparity exhibits a slight decreasing trend over time, revealing only modest progress in female participation. By providing empirical evidence of gender diversity trends in BPM academia, this research contributes valuable insights into women's evolving role within the discipline and underscores the need for further efforts to foster inclusivity in this community.

Keywords: Business Process Management, BPM, Systematic Mapping, Gender Diversity, Women in BPM, Women in Technology

Received: 30 October 2025 • **Accepted:** 08 July 2026 • **Published:** 25 August 2026

1 Introduction

As a business management discipline, Business Process Management (BPM) is characterized by interdisciplinary practices where people and organizational culture are fundamental for the success of BPM programs [Dumas *et al.*, 2018]. The implementation of BPM projects requires a diverse set of skills and competences, which could be achieved with a broad variety of professionals [Müller *et al.*, 2016]. Furthermore, the "people" dimension is considered the key for the organization to be able to achieve high levels of BPM maturity [Kerpedzhiev *et al.*, 2021].

BPM demonstrates great potential to drive innovation [Schmiedel and vom Brocke, 2015], the same potential found in environments with the presence of diversity [Jones *et al.*, 2020]. A diverse workforce employs individuals that vary by many different characteristics, including, but not limited to, distinct cultures, races, ages, genders, religions, physical conditions, sexual orientations, socioeconomic aspects and nationalities. Therefore, diversity is naturally assumed to be beneficial to the innovation process, since it brings an alternation of viewpoints. Within the aspect of gender diversity, existent evidence shows gender heterogeneous teams produce higher quality technical and scientific outputs [Campbell *et al.*, 2013; Yang *et al.*, 2022].

Despite progress in promoting inclusion, gender disparity in the fields of science, technology, engineering, and mathematics (STEM) remains significant. In the 2020/2021 academic year in Spain, only 8% of all female university students were enrolled in Engineering courses, compared to 29.5% of all male university students, according to recent European statistical reports. [Arrazola *et al.*, 2025]. The BPM community is not immune to the challenges presented by the lack of diversity and female underrepresentation. Up to now, there have been no studies analyzing female participation in academia in the context of BPM. Since data about the mentioned phenomena is still sparse, the present study seeks to investigate the extent of female representation in academic publications related to BPM, while also raising awareness of gender inequality and its implications for the field.

The general goal of this study is to assess female participation in the BPM scientific production. In particular, the following research questions are answered:

RQ1 - *How is the gender distribution of authors in the BPM area?*

RQ2 - *What is the profile of women researching and publishing in the BPM area?*

To achieve the general goal of this study, we performed a Systematic Mapping Study (SMS), searching for the scientific production following the methods and guidelines proposed by [Petersen *et al.*, 2015]. In order to analyze the presence of women in the BPM community, well-known academic databases are used as data sources. Searches on the mentioned libraries are performed, and among the academic studies obtained, particular aspects are investigated, mainly related to the authors. These aspects correspond to the authors' gender, the authorship distribution, the order of the authors (analysis of the first and last authors), the years of the publications, the country of their affiliations and the BPM fields of the publications.

By this means, we expect to raise awareness and re-

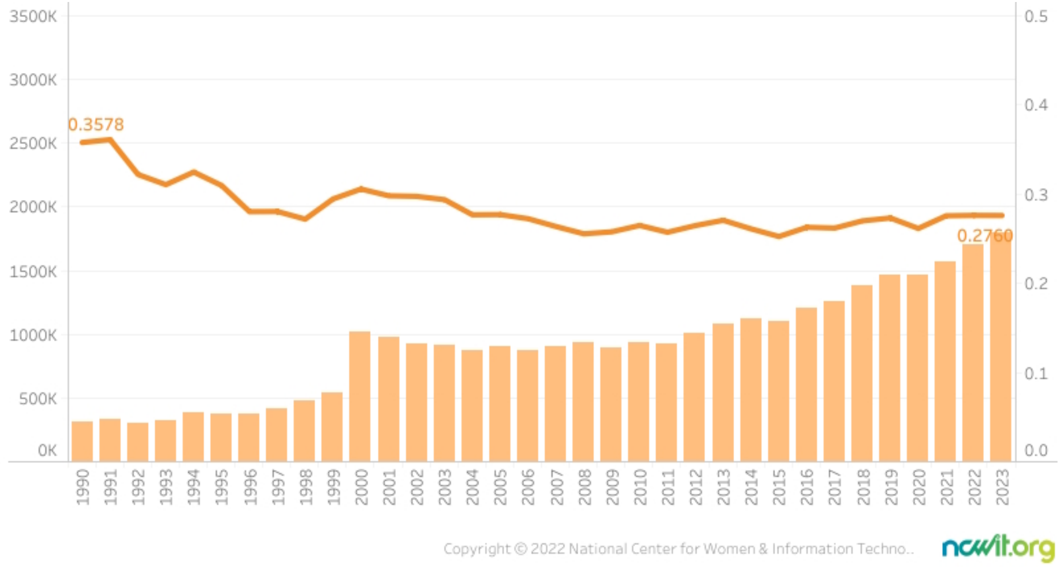


Figure 1. Women in computing occupations from 1985 to 2023 in the United States, according to National Center for Women & Information Technology

Figure 2. Women corresponding authors according to ISSA2, by science field (2018)

flections that motivate actions and policies to attract female professionals and researchers to the BPM area, promoting a more diverse environment that stimulates innovation.

This paper is organized as follows: Section 2 presents a research background, offering more context on the problem regarding gender diversity in computing area. Section 3 presents related works, Section 4 explains the methodology applied and Section 5 presents the results obtained, responding to each of the research questions. Finally, Section 6 presents the conclusions of this study and possible future works that may derive from this.

2 Research background

The absence of diversity in the workplace is a consequence of systemic discrimination, defined as unequal treatment based on criteria irrelevant to the activity or productivity involved [Becker, 1957]. In economic literature, this is often categorized into two models: the neoclassical model of discrimination by preference, where bias is a conscious choice [Becker, 1957], and statistical discrimination, where generalizations about certain groups lead to unintentional but systemic exclusion [[Phelps, 1972; Arrow, 1973]]. Regardless of the underlying mechanism, these biases lead to a less diverse work environment, stifling the innovation potential fueled by varied perspectives, “brainpower and creativity fuel innovation” [Trauth, 2011].

Historically, women were a critical part of the computing industry’s technical workforce until the late 1960s. This circumstances shifted by 1970, when personal computers began to be marketed almost exclusively to men [Hicks, 2018]. Currently, researchers are studying possible barriers regarding the participation of women in the area. According to [Kordaki and Berdousis, 2020], there are several factors involved, including school education, family, computer games, peer and work culture, role models, stereotypes, among others.

Concerning the academic field, researches show that women are underrepresented as first and last authors. The gap is especially large in authorship positions associated with seniority, and prestigious journals have fewer women authors [Holman *et al.*, 2018]. A potential reason is the gender bias women authors face. An experiment performed by [Knobloch-Westerwick *et al.*, 2013] shows that publications, where the identity of authors was experimentally manipulated, were associated with greater quality by participants when these publications were supposedly written by men.

Funding is also a critical factor that impacts the participation of women in scientific careers, where they face challenges in the competition for limited resources. A report shows that National Institutes of Health (NIH) research grants to women have increased over the past two decades around the world, from 23% in 1998 to 34% in 2019 [NIH, 2020], but it is still a topic to be addressed.

Gender inequity across the fields of STEM is a reality that has been increasingly acknowledged in recent years. Despite raising awareness, the change is slow, and factors that negatively impact women in this area are studied [Llorens *et al.*, 2021; Cheryan *et al.*, 2017]. The percentage of computing occupations held by women has been declining, since 1991, when it reached a peak of 36%, as shows Figure 1. Also, women are 1.5 times more likely than men to leave the STEM higher education pipeline [Ellis *et al.*, 2016]. This problem is emphasized for advanced career stages, when the participation of men increases and women decreases. This phenomenon perceived as an evolution of two groups in opposed directions is referred to as "scissors effect".

In science, this perception is confirmed by one of the findings of the most recent International Survey of Scientific Authors (ISSA2) performed globally by the Organisation for Economic Co-operation and Development (OECD) in 2018 [Bello and Galindo-Rueda, 2020]. Results show that women are underrepresented in research careers as they are in many other sectors, representing less than 1/3 of all researchers. Figure 2 shows the women correspondent authors by science field, highlighting business and computing science fields, the areas most commonly related to BPM publications.

3 Related works

While gender diversity is an established research topic in technology, studies examining female participation in academy specifically within BPM community remain limited. In order to position our study, we define three primary criteria for comparison: Context (Industry/Profession vs. Academia), Analysis Unit, and Methodological Scope.

Specifically, the Analysis Unit distinguishes the core focus of the evaluated studies: whether they investigate the professional *skills* of the industry workforce, the topical *content* of academic publications, or the *authorship* demographics of the researchers themselves. Table 1 provides a systematic comparison of these works, adding its primary limitations.

The primary research regarding gender in the BPM domain were studies by [Gorbacheva *et al.*, 2015] and [Gorbacheva *et al.*, 2016]. These papers offer an exploratory perspective on professional competencies. In their analysis of over 10,000 LinkedIn profiles from the United States, Canada, the United Kingdom, and Australia using text mining techniques, the authors revealed that women have poor representation among BPM professionals, particularly in technological roles. Although these papers provide valuable insights into the labor market and occupational stereotypes, their focus is narrow, as they overlook the researchers themselves and their academic production. Furthermore, the au-

Table 1. Systematic Comparison of Related Works in Gender Diversity

Study	Context	Methodology	Analysis Unit	Primary Limitations
[Gorbacheva *et al.*, 2015]	BPM Profession	Data Mining	Skills[1]	Relies on self-reported LinkedIn data; restricted to four English-speaking countries.
[Gorbacheva *et al.*, 2016]	BPM Profession	Data Mining	Skills[1]	Data cannot be independently verified; biased toward Western cultures.
[Kohl Silveira and Prikladnicki, 2019]	Software Engineering	SMS	Content[2]	Limited to Agile literature; does not investigate demographic backgrounds of authors.
[Rodríguez-Pérez *et al.*, 2021]	SE Profession	SLR	Content[2]	Search restricted to titles only; excluded non-CS venues and management perspectives.
[Felizardo *et al.*, 2021]	EBSE Academia	SMS	Authorship[3]	Restricted to secondary studies; utilizes automated binary gender inference tools.
[Aleem and Ahmed, 2023]	Software Engineering	SLR	Content[2]	Missing indexing systems like Scopus; focuses purely on how EDI is studied.
[Alarcón-Alarcón *et al.*, 2024]	SE Profession	SLR	Content[2]	Scarcity of explicit productivity metrics; lacks mapping of role distribution.
This Study	**BPM Academia**	**SMS**	**Authorship**[3]	Focuses on academic producers; does not capture internal industry dynamics.

[1] **Skills** refers to the evaluation of professional competencies and the demographics of the industry workforce.
[2] **Content** focuses on the topical findings, research themes, and study results reported within the literature.
[3] **Authorship** examines the demographic backgrounds and representation of the researchers who are producing the academic work.

thors acknowledge that their results are limited to specific Western, English-speaking cultures, meaning the findings might differ globally.

Given the technical and historical proximity between BPM and SE, we also analyzed several Systematic Literature Reviews (SLRs) and mappings evaluating diversity in the SE field. [Rodríguez-Pérez *et al.*, 2021] examined how developers perceive diversity in Open Source communities, though their analysis was constrained by search strings restricted to paper titles. Similarly, [Alarcón-Alarcón *et al.*, 2024] compiled evidence on gender roles and team dynamics, finding that women are often assigned to less technical roles due to unconscious bias. In the broader context of technical fields, studies such as [Kohl Silveira and Prikladnicki, 2019] and [Aleem and Ahmed, 2023] utilize systematic methodologies to evaluate Equity, Diversity, and Inclusion (EDI) principles within development teams. While these studies provide substantial benefits by mapping the state of diversity practices, a key limitation is that they analyze the *content* of the papers—such as research methods and EDI factors—rather than investigating the demographic backgrounds of the authors shaping the academic discourse.

Our study differentiates itself from the existing literature by shifting the unit of analysis from the industry workforce and paper content to the academic producers of knowledge. [Felizardo *et al.*, 2021] conducted a similar SMS within the subfield of Evidence-Based Software Engineering (EBSE), identifying an increase in female participation concentrated in Europe and Latin America. However, their study was limited to secondary studies and utilized automated binary gender inference without manual verification of cultural biases.

Ultimately, while prior BPM studies focus on the "supply side" of industry skills, ours directly addresses the unexplored gap regarding the representation and career progression of female researchers within the BPM academic community itself. In terms of methodological identity, this work is classified as a Systematic Mapping Study (SMS) following the taxonomy proposed by [Ralph and Baltes, 2022], as it focuses on categorizing and identifying gaps within the authorship pipeline of a specific research area.

4 Methodology

The SMS performed to achieve the general goal of this study aimed to identify and analyze publications from women authors in the BPM field. The SMS was conducted following the protocol proposed by [Petersen *et al.*, 2008] and updated by [Petersen *et al.*, 2015]. It consists of (1) formulation of research questions, (2) conduct search, (3) selection of publications, (4) extraction of data, and (5) mapping process by analysis of data. Figure 3 describes the protocol applied to this research.

By following a structured protocol with the parameters presented in this section, this study can be reproduced and compared to new SMS performed in the future, allowing an assessment of the evolution of female BPM authorship in time.

4.1 Formulation of the Research Questions

This study aims to assess women participation in the BPM area and their distinct contributions. Thus, the following research questions were defined to guide the systematic mapping:

RQ1 - *How is the gender distribution of authors in the BPM area?* This question intends to identify the gender gap between women and men in BPM research publications. This question encompasses distinct analyses, including the number of male and female authors over the years, first and last author distribution, gender authorship distribution, among others.

RQ2 - *What is the profile of women researching and publishing in the BPM area?* With this question, we investigate who are the women contributing to the academic field of BPM, that is, their professional profile, and how those contributions are distributed through the years. Also, the intent is to cluster these studies based on their different BPM fields as well as geographic location given by their place of work and professional profiles (e.g., their country and affiliation).

4.2 Conduct Search

Search String: Since the intent of this study is not to be limited to a specific BPM subject, the search query to retrieve the papers of interest needs to be generic enough to cover a variety of BPM topics. The search fields and research query are defined as follows:

- Search fields: Title AND keywords;
- Search string: (BPM OR "Business Process Management").
- OR
- Search field: Abstract;
- Search string: "Business Process Management".

The approach of using distinct search strings for different search fields was applied because using the string "BPM OR 'Business Process Management' " uniformly across title, keywords, and paper abstract brought many out-of-scope results. The acronym "BPM" suffers from severe semantic collision in academic literature, frequently returning studies related to medicine (e.g., beats per minute) or physics (e.g., beam propagation method). We observed that most of the unrelated results occurred due to the inclusion of the "BPM" term in the paper's abstract, thus adapting the search query to specific fields. This calibration ensured high relevance without significantly compromising the recall of valid studies.

Data Sources: The queries are applied to four electronic libraries considered highly relevant for conducting SMS in BPM: IEEE Xplore (`https://ieeexplore.ieee.org/`), Scopus (`https://www.scopus.com/`), ACM Digital Library (`https://dl.acm.org/`), and SBC-OpenLib (`https://sol.sbc.org.br/`.

The selection of these specific databases was intentional to mitigate potential database bias. Scopus was included to provide broad, multidisciplinary coverage of high-impact journals, while IEEE and ACM were selected to capture deeply technical, conference-oriented profiles that are characteristic of the Software Engineering and Information Systems domains. Finally, SBC-OpenLib ensures the inclusion

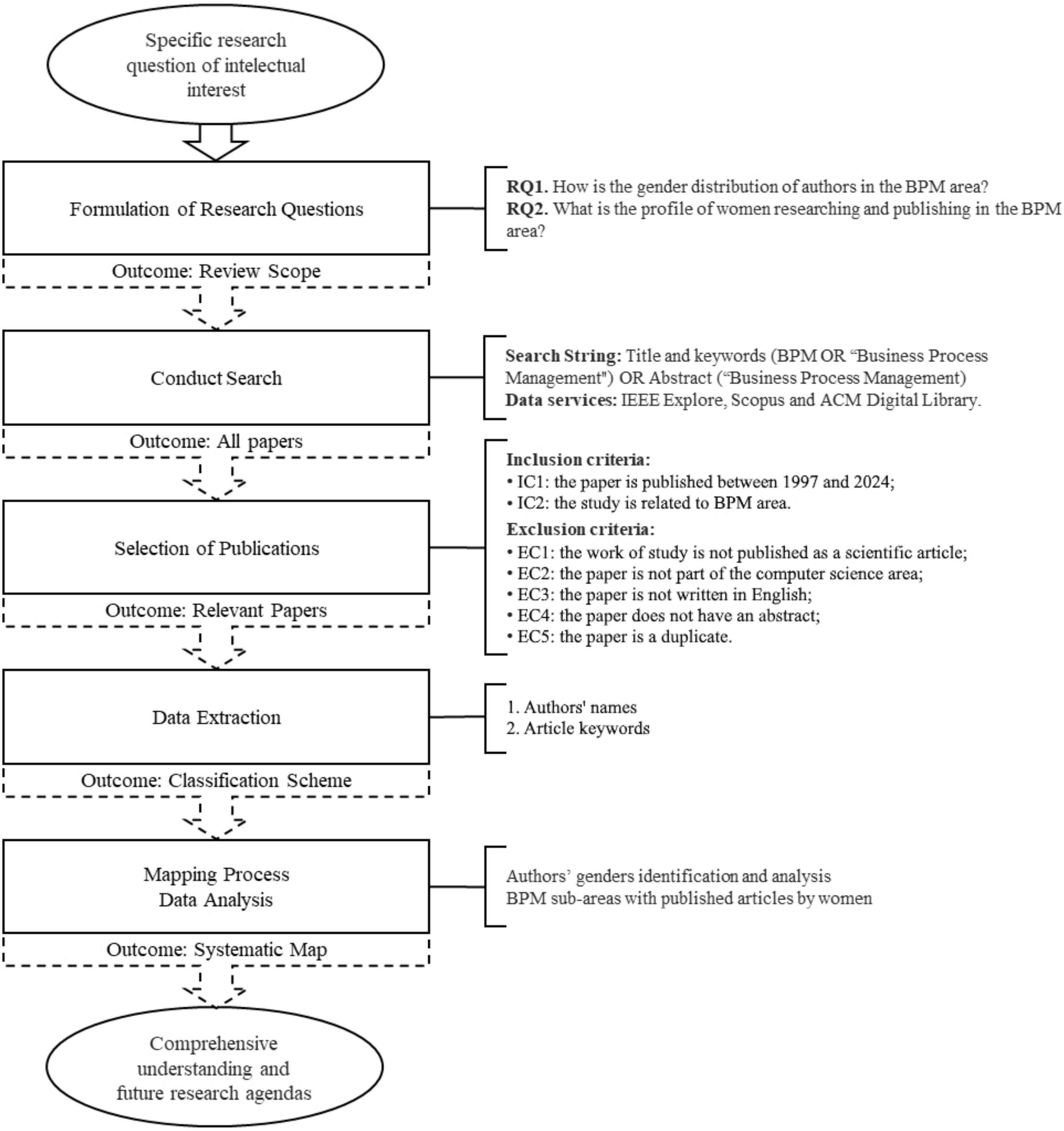


Figure 3. Protocol for Systematic Mapping Study

of regional Latin American contributions that might be overlooked by global indexes.

It was defined that the time range of the analysis would begin from the identification of the first article on BPM authored by at least one woman. An intermediate search was conducted, resulting in 7,077 articles, applying only the exclusion criteria for duplicates and articles without an abstract. The search covered publications from the 1950s to 2024, with the following time coverage by database: Scopus (from 1957), IEEE (from 1979), and ACM (from 1995). The earliest article on BPM with female authorship was identified in 1997.

The queries are applied to each academic database using the provided search engine mechanisms.

4.3 Selection of Publications

The inclusion criteria (IC) and exclusion criteria (EC) are used to exclude studies that are not relevant to answering the research questions.

Exclusion Criteria: The ECs are used to filter the irrelevant studies from the sources in terms of format, publication details and access. These are the defined criteria: EC1: the work of study is not published as a scientific article; EC2: the paper is not part of the computer science area; EC3: the paper is not written in English; EC4: the paper does not have an abstract; EC5: the paper is a duplicate.

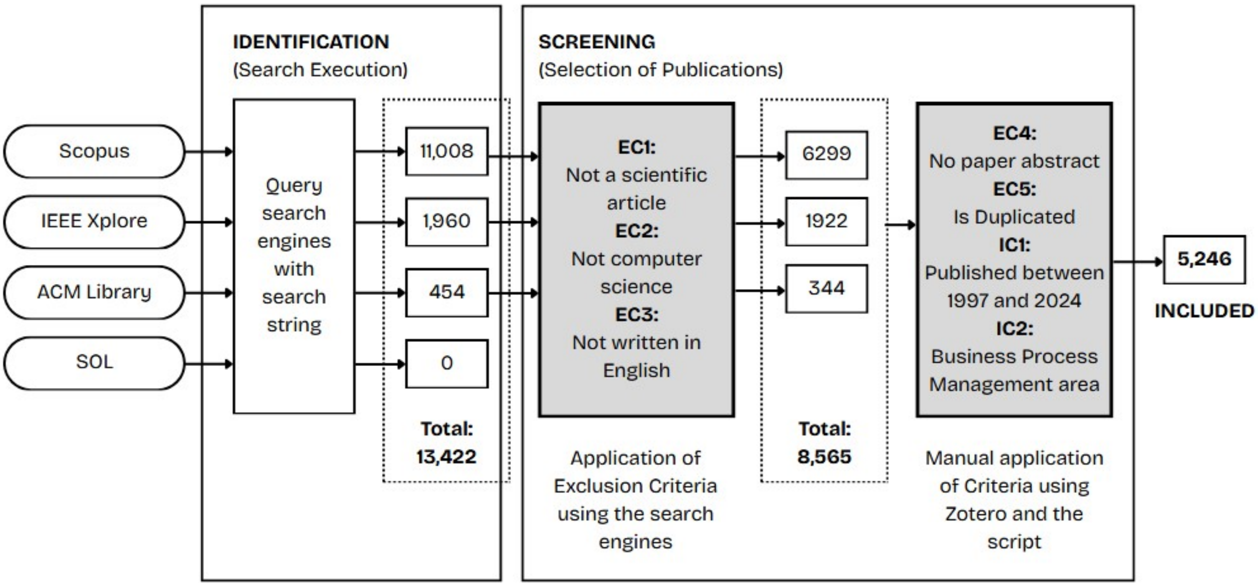


Figure 4. Overview of the Systematic Mapping Protocol and Article Screening Flow

The EC4 criteria is applied directly to the search engines of the academic libraries using the provided filters. From 13,422 studies obtained from the query search engines on online libraries, the filters applied with the first 3 exclusion criteria resulted in an amount of 8,565 papers, as shown in Figure 4.

For the EC4 and EC5 criteria, Zotero (`https://www.zotero.org/`), which is a library management tool, was used to eliminate both the duplicated papers and the papers with no abstract. These two criteria were applied simultaneously as the inclusion criteria inside the tool.

Inclusion Criteria: The resulting studies from the application of EC were submitted to ICs. Studies were included in the systematic mapping if they meet both of the following criteria: IC1: The study was published between the years 1997 and 2024; IC2: The study is related to the BPM area.

To minimize subjectivity during the application of IC2, a rigorous manual disambiguation protocol was executed using the Zotero reference manager. The primary objective was to filter out "false positives" caused by acronym collisions common in literature (e.g., "Beats per Minute" in cardiology/music, or "Beam Propagation Method" in physics). Titles, abstracts, and keywords were manually screened to exclude papers containing the acronym "BPM" that did not explicitly relate to Business Process Management.

Figure 4 presents the results of the application of manual EC/IC criteria, showing that the 8,565 papers resulting from the first process were reduced to 5,246 items.

4.4 Data Extraction

To achieve the findings for RQ1 and RQ2, it was required to collect data regarding identification of authors, the subject of study in BPM field and their location. This data is available in the metadata attributes that complement the catalog of studies and can be retrieved from the BibTeX files generated automatically by the digital libraries. Data extracted from the BibTeX files were then processed by an automated script written in Python, resulting in a database with the list of authors' names (to assess the identification of authors) and author keywords (to assess the subjects of study).

4.5 Data Analysis

Considering the large amount of data collected from the systematic mapping process, a set of tools is used to automate the reporting and analysis of the results. Among these tools, the majority corresponds to automated scripts, and they can be found on a remote repository: `https://github.com/bpm-research-lab-ufrgs/Women-in-BPM`.

Gender classification is an essential part of this study and it corresponds to the first phase of answering the defined research questions. The chosen approach evaluates gender based on the given name of the authors. Since there was a considerable set of articles to be evaluated, this study used automated tools to guarantee correctness and replicability.

Four different tools were considered to evaluate the authors' gender based on the names: Gender API (`https://gender-api.com/`), Genderize.io (`https://genderize.io/`), NameAPI (`https://www.nameapi.org/`), and NamSor (`https://www.namsor.com/`). All of them are APIs with large databases of names already classified and used to predict the gender of a person given their name. These tools can classify the gender for a given set of authors' names. Most of them generate as output the accuracy of the information response. [Santamaría and Mihaljević, 2018] provides a comparative study between the majority of the mentioned tools and it was also considered in order to decide the most appropriate API for this study. Finally, the NamSor tool was chosen, as it delivered the best results among the free alternatives tested. A more detailed comparison between the tools and their provided features can be seen in Table 2.

After selecting the most appropriate API for this study, it was necessary to define an acceptance threshold, discard-

Table 2. Comparison between gender assessment APIs (2025)

Feature	Gender API	genderize.io	NameAPI	NamSor
Database size	6 million+	1 billion+	3 million+	4.4 million+
Free tier quota	100/month	no	10,000/month	1,000/month
Regular data updates	yes	yes	yes	yes
Unstructured strings	yes	yes	yes	yes
Non-Latin support	yes	yes	yes	yes
Geo-localization	yes	yes	yes	yes
Locale-aware prediction	yes	yes	yes	yes
Assignment type	probabilistic	probability + count	categorical + confidence	probabilistic scale
Machine Learning based	yes	no	no	yes
API available	yes	yes	yes	yes

ing studies in which the authors' gender was assessed with a low accuracy rate based on the API's response. In general, the accuracy rate is lower for gender-neutral given names (e.g., "Andrea"). For this reason, gender classification results were considered only when the confidence level exceeded 75%. Authors whose names were abbreviated (e.g., "W. Smith") were also excluded, since gender classification in such cases is imprecise. In our tests, confidence levels for abbreviated names did not exceed 75%, frequently remaining around 50%, as expected.

Considering the cultural diversity of names, different strategies were evaluated for handling hyphenated first names, which are common in Hispanic and Asian naming conventions. Three approaches were tested: (1) removing the hyphen and including both parts of the name; (2) using only the first part; and (3) keeping the hyphen as is. Through testing, it was observed that some names were correctly recognized when the hyphen was retained—especially Hispanic names; however, this was not the case for names of East Asian origin. Option (2) yielded the most consistent results in the majority of cases and was therefore adopted, as it generally produced significantly better results.

Table 3. Overview of articles and authors in BPM (1997–2024)

Year	Articles	Authors
1997	7	17
1998	6 (-14.3%)	23 (+35.3%)
1999	4 (-33.3%)	11 (-52.2%)
2000	4 (0.0%)	12 (+9.1%)
2001	7 (+75.0%)	18 (+50.0%)
2002	5 (-28.6%)	13 (-27.8%)
2003	10 (+100.0%)	27 (+107.7%)
2004	18 (+80.0%)	49 (+81.5%)
2005	22 (+22.2%)	80 (+63.3%)
2006	35 (+59.1%)	97 (+21.2%)
2007	58 (+65.7%)	155 (+59.8%)
2008	104 (+79.3%)	326 (+110.3%)
2009	118 (+13.5%)	367 (+12.6%)
2010	223 (+89.0%)	652 (+77.7%)
2011	385 (+72.6%)	1,165 (+78.7%)
2012	324 (-15.8%)	979 (-16.0%)
2013	384 (+18.5%)	1,170 (+19.5%)
2014	371 (-3.4%)	1,151 (-1.6%)
2015	344 (-7.3%)	1,067 (-7.3%)
2016	362 (+5.2%)	1,128 (+5.7%)
2017	325 (-10.2%)	1,024 (-9.2%)
2018	373 (+14.8%)	1,257 (+22.8%)
2019	365 (-2.1%)	1,162 (-7.6%)
2020	261 (-28.5%)	902 (-22.4%)
2021	262 (+0.4%)	892 (-1.1%)
2022	297 (+13.4%)	997 (+11.8%)
2023	297 (0.0%)	1,125 (+12.8%)
2024	275 (-7.4%)	956 (-15.0%)
Total	**5,246**	**16,822** (not exclusive)

From the total set of 5,246 articles collected, the authors' gender was classified in 5,029, while 217 (4.1%) were excluded because they did not list any authors. We developed a Python script to compute the number of distinct authors over a specified time range. In this case study, only first names were extracted by the tool, as surnames were not used to evaluate gender through the API. Therefore, first names were normalized using a Python library before comparison, and any middle or last names were ignored—except in cases where two names were separated by a hyphen. After processing the article data and the defined time range, we identified a cluster of not exclusive 16,822 authors who published work in the BPM field between 1997 and 2024.

To ensure statistical robustness and validate the data collection process, a series of quantitative tests were performed using Python (*pandas* [McKinney *et al.*, 2010], *scipy*, and *statsmodels* [Seabold and Perktold, 2010]), adopting a significance level of $\alpha = 0.05$. First, to verify that the demographic findings were not skewed by methodological limitations over time, a linear regression was applied to test the stability of the rate of unidentified authors (unknown gender) between 2010 and 2024. Subsequently, a Simple Linear Regression was applied to validate the historical publication growth trend [Montgomery *et al.*, 2021]. To verify if the proportion of male and female authors changed significantly over time, a Chi-Square Test of Independence (χ^2) was used [Agresti, 2018]. Finally, to empirically test the hypothesis of whether the COVID-19 pandemic disproportionately affected productivity by gender, a Generalized Linear Model (GLM) with a Poisson distribution was employed [Cameron and Trivedi, 2013], specifically evaluating the interaction term between the author's gender and the pandemic period.

5 Results

This section presents and discusses the main findings of the study, providing insights and evidence to address the research questions previously outlined.

5.1 Research Question 1: How is the gender distribution of authors in the BPM area?

To answer this question, we conducted a quantitative analysis of the gender distribution of authors. First, table 3 presents the number of articles and the number of authors per year. Authors are counted individually, even if the same author has more than one article published in the same year. In this representation, the total number of authors also includes repetitions of authors over the years.

Before analyzing the demographic trends, the reliability of the gender identification tool was verified. A linear regression confirmed the absence of a significant trend in the rate of unclassified authors between 2010 and 2024 ($p = 0.5713$), validating that the observed demographic shifts are genuine. Furthermore, regarding the temporal evolution of scientific production, a simple linear regression demonstrated a continuous and highly significant growth in the overall number of publications across the analyzed period ($p < 0.001$).

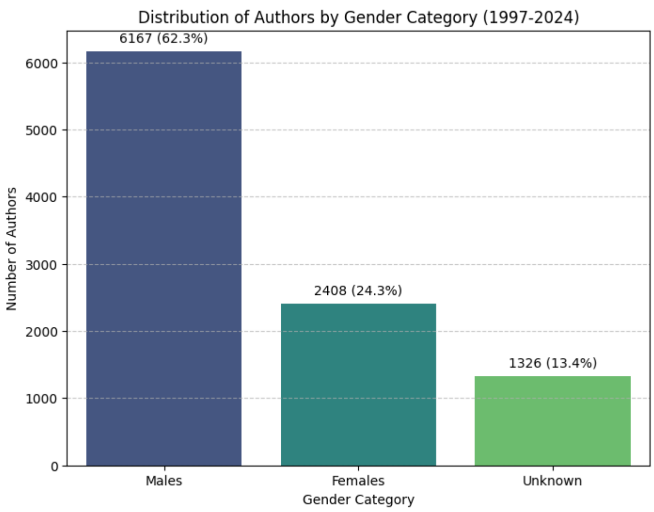


Figure 5. Absolute numbers of male and female authors

The figure 5 represents the number of unique authors who published their work between 1997 and 2024. The list of authors was curated to avoid duplicated values. From the 9,901 authors assessed, 6,167 are evaluated as men (62.3%), 2,408 are women (24.3%), and the tool was not able to classify gender for 1,326 authors (13.4%). These figures rely on an automated binary gender classification (male/female) inferred from first names. As further discussed in Section 5.2, this methodological decision may introduce classification biases and inherently excludes non-binary identities, warranting caution when interpreting the absolute findings. Even so, this data reveals that the number of male authors is almost three times the number of female authors.

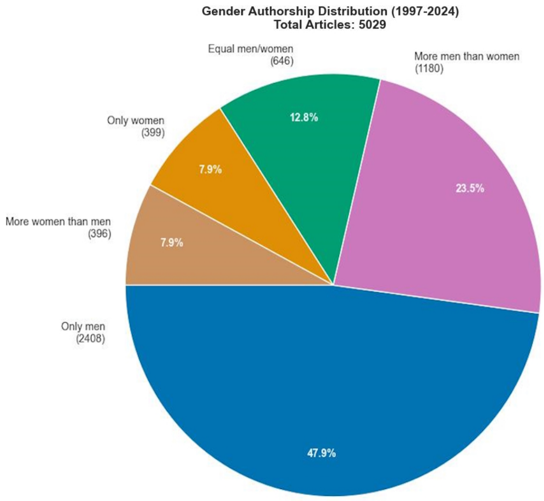


Figure 6. Gender authorship distribution

Another interesting perspective to investigate is how the BPM authors organize themselves to publish their studies and Figure 6 shows the gender authorship distribution. From 5,029 articles published between 1997 and 2024, only 399 were published exclusively by women (7.9%), while the number of articles published exclusively by men is more than six times bigger (47.9%). The low number of female-only authored papers may suggest both an underrepresentation and a greater reliance on collaboration with male colleagues for inclusion in publication networks.

It is a positive factor to notice that a considerable amount of publications analyzed have female and male authors collaborating together, although it is yet under half of the cases. This collaborative scenario should be stimulated, since BPM is an interdisciplinary area, and scientific research can benefit from the engagement of people with distinct perspectives.

Analyzing the gender authorship distribution throughout the years in Figures 7 and 8, it can be observed that despite the slight difference between the total number of authors each year, the number of male authors is always far superior to the number of female authors. The most significant disparity can be seen in 2018, with a proportion of 3.62 male authors for each female author. In order to complement this analysis, Figure 9 presents the normalized percentage distribution of male and female authors in each year. Although the presence of men is greater than that of women in all years assessed, it is possible to notice that the difference between genders tends to decrease slightly over the years, especially when the last seven years are observed. A Chi-Square Test of Independence confirmed that the proportion between male and female authors did not remain static over time ($\chi^2 = 1193.9$, $p < 0.001$), providing mathematical evidence of this structural distributional shift. In 2024, female authors accounted

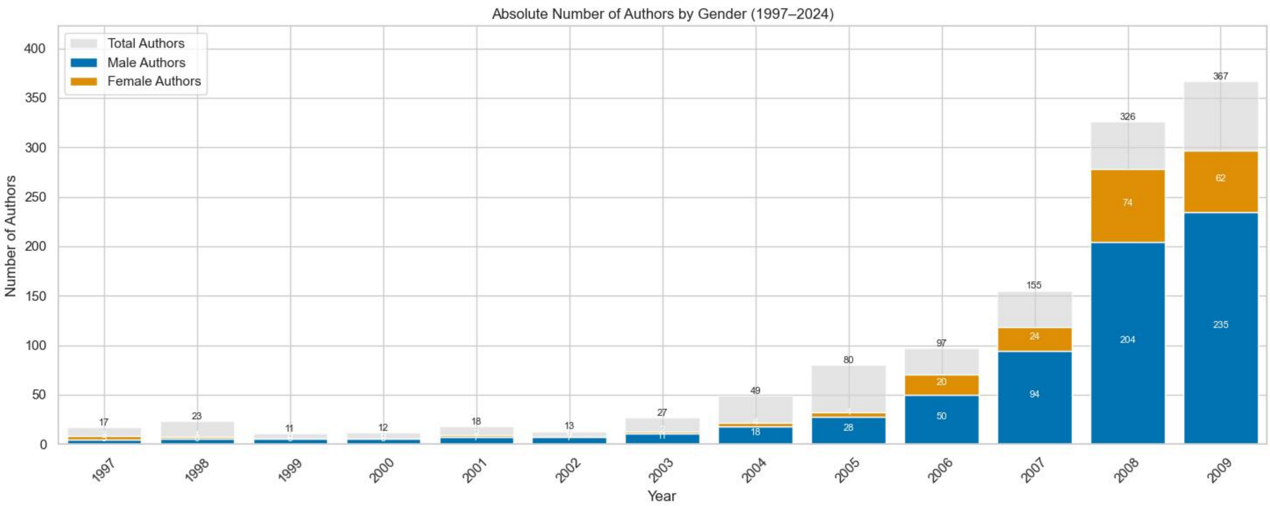


Figure 7. Absolute number of authors over the years (Part 1)

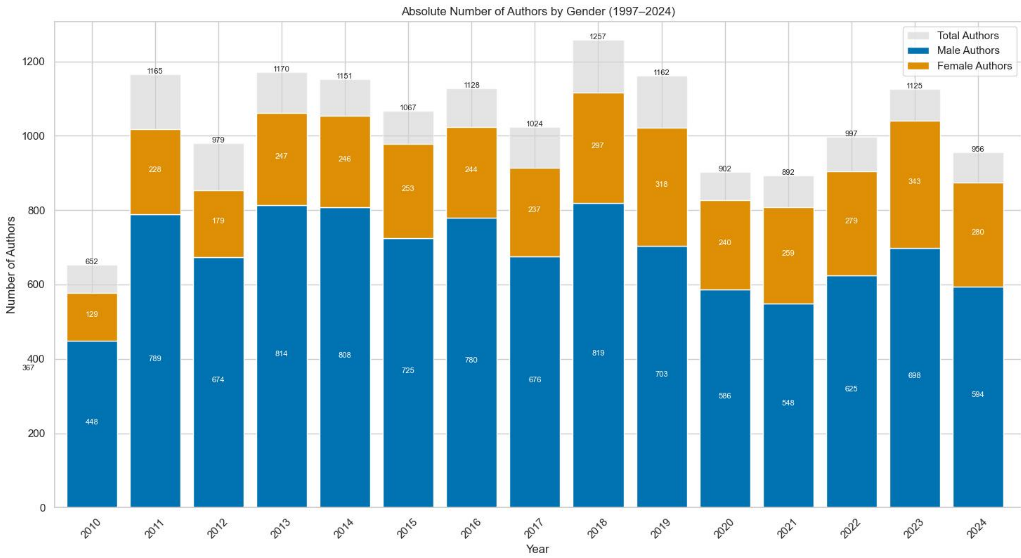


Figure 8. Absolute number of authors over the years (Part 2)

for 32% of the total, which represents an increase of 20.3 percentage points compared to 2018, the year with the lowest female representation. Considering all 28 years analyzed in this study, women contributed to an average of 22.11% of the publications.

Many factors may have contributed to the growing trend of women's participation. Initially, we hypothesized that the stay-at-home policies implemented during the COVID-19 pandemic might have distinctly affected women's productivity by altering the dynamics of reconciling scientific work with family affairs. However, empirical testing using the GLM revealed that while overall publications significantly increased during the pandemic period ($p < 0.001$), the interaction between gender and the pandemic was not statistically significant ($p = 0.278$). This indicates that the pandemic did not disproportionately favor or hinder female authors in our dataset; rather, the publication growth was proportional across genders. Consequently, this sustained increase in women's participation is more likely driven by other fundamental factors, such as the growing interest of women in computing and BPM research, public and institutional policies promoting diversity, and the strengthening of support networks for women in science.

Another interesting analysis can be done by comparing the first and last authors' genders over the years. Figures 10 and 11 present, respectively, the absolute numbers of the first and last author. Although there is no rule to define the order of authors in scientific publications, in general, the community understands the last author as the academic advisor and the first author as the mentee. This pattern was considered for the authors' name order analysis.

It is possible to visualize a significant difference between the number of male and female advisors and between male and female mentees. In both cases, the number of men is always higher than the number of women every year. Comparing exclusively the data distribution between the first and last female author, the last female authors is always lower than the first female authors, with an average of 18.13% less

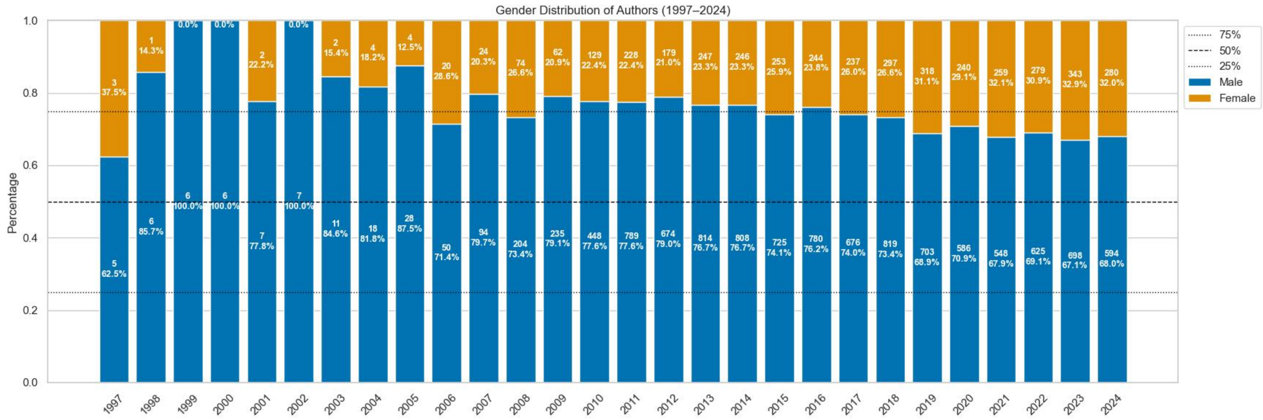


Figure 9. Gender percentage distribution over the years

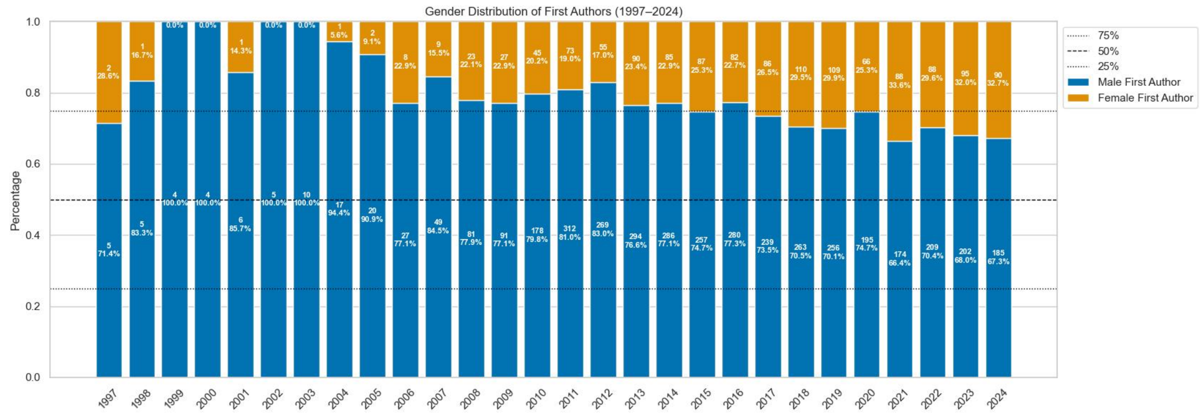


Figure 10. First author gender over the years

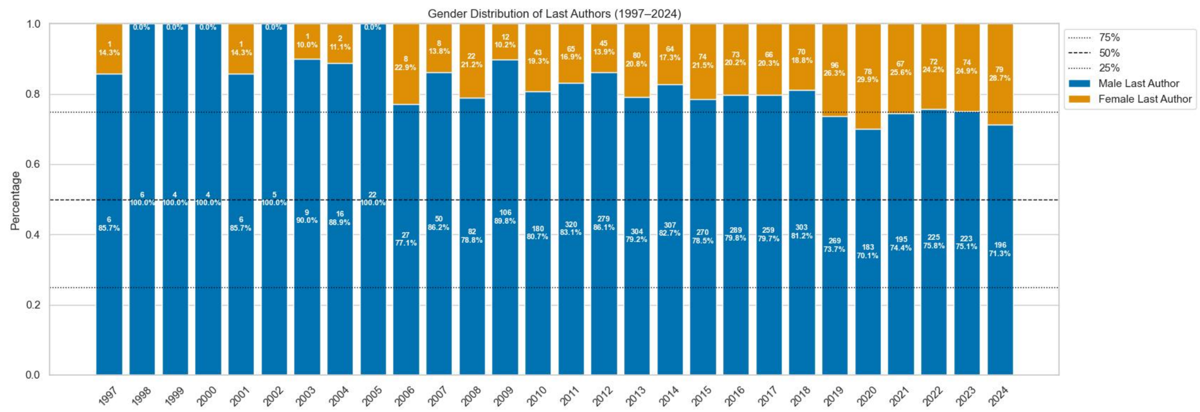


Figure 11. Last author gender over the years

participation. A possible interpretation of this data is of a potential increase of women in courses related to computer science and BPM areas, signalizing a possible increase in the number of female advisors. The lower proportion of women as final authors may also reinforce gender disparities in senior positions and reflect structural barriers to academic career progression. The number of female first authors registered in 2024 is the highest, more than 10% bigger than in 1997, the first year analyzed.

5.2 Research Question 2: What is the profile of women researching and publishing in the BPM area?

With this question, the goal is to investigate and understand more about the profile of female BPM researchers. To this purpose, just the articles with at least one woman as author are considered. As already seen in the previous section, 2,621 articles were collected that match this criterion, which corresponds to 49,96% of the total of 5,246 articles.

The first analysis corresponds to the different BPM topics addressed by the articles and which ones are most recurrent. These topics are extracted from the article keywords

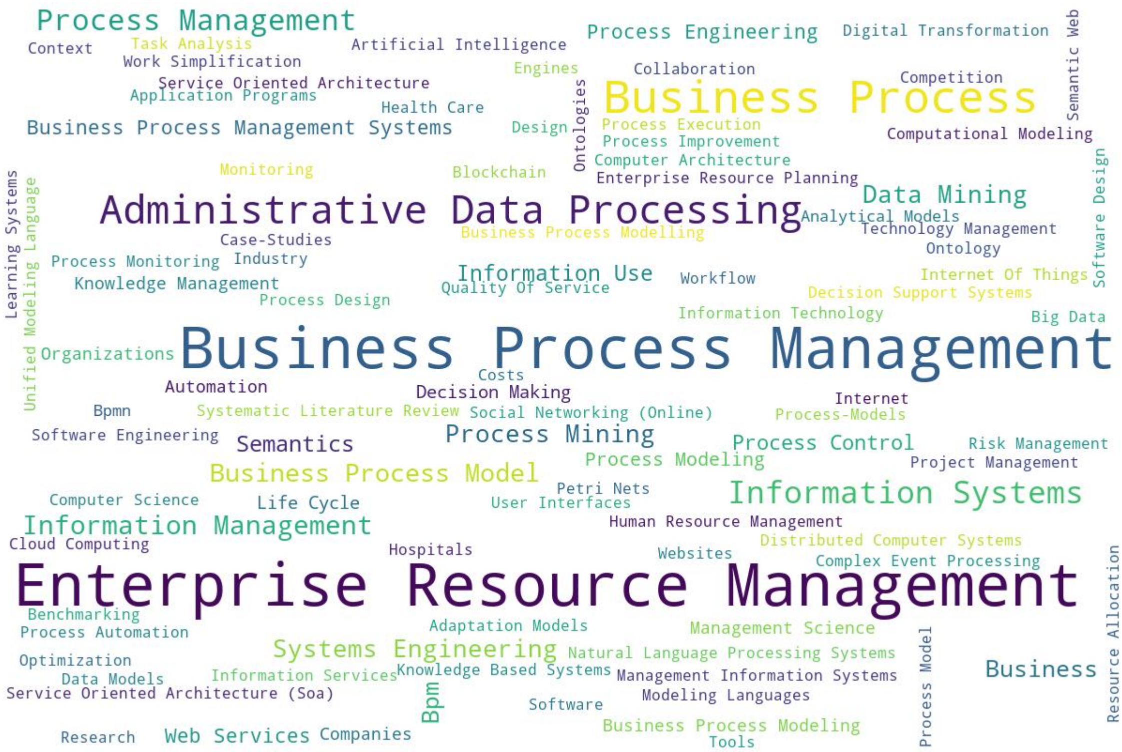


Figure 12. Word cloud with most common BPM topics

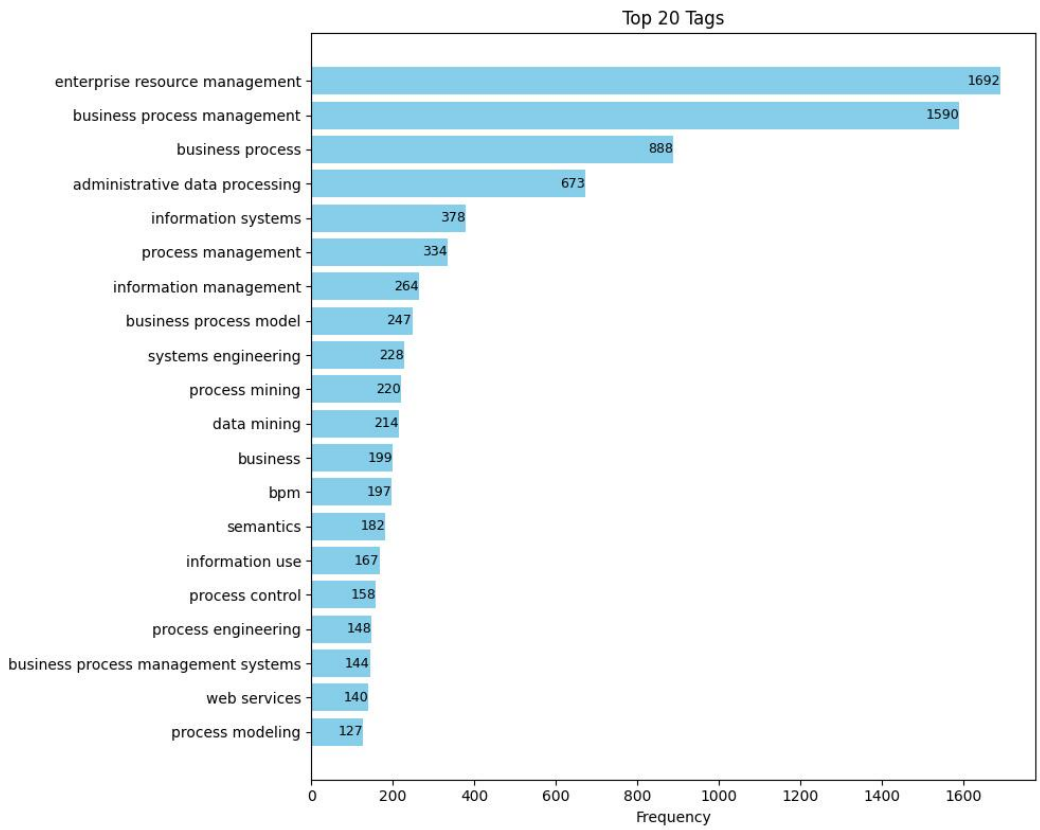


Figure 13. Main BPM topics addressed on articles

metadata. To accomplish this analysis, we developed another Python script. Basically, the script receives as input a CSV file with the result file from the articles analysis and can be settled to extract only the ones with women participation and generates as output a file with all the distinct keywords identified and also the number of times each keyword was found. The mentioned script also generates a word cloud visualization, that can be seen in Figure 12. Basically, a word cloud is a visual representation object for text processing, which shows the most frequent occurrences with bigger letters.

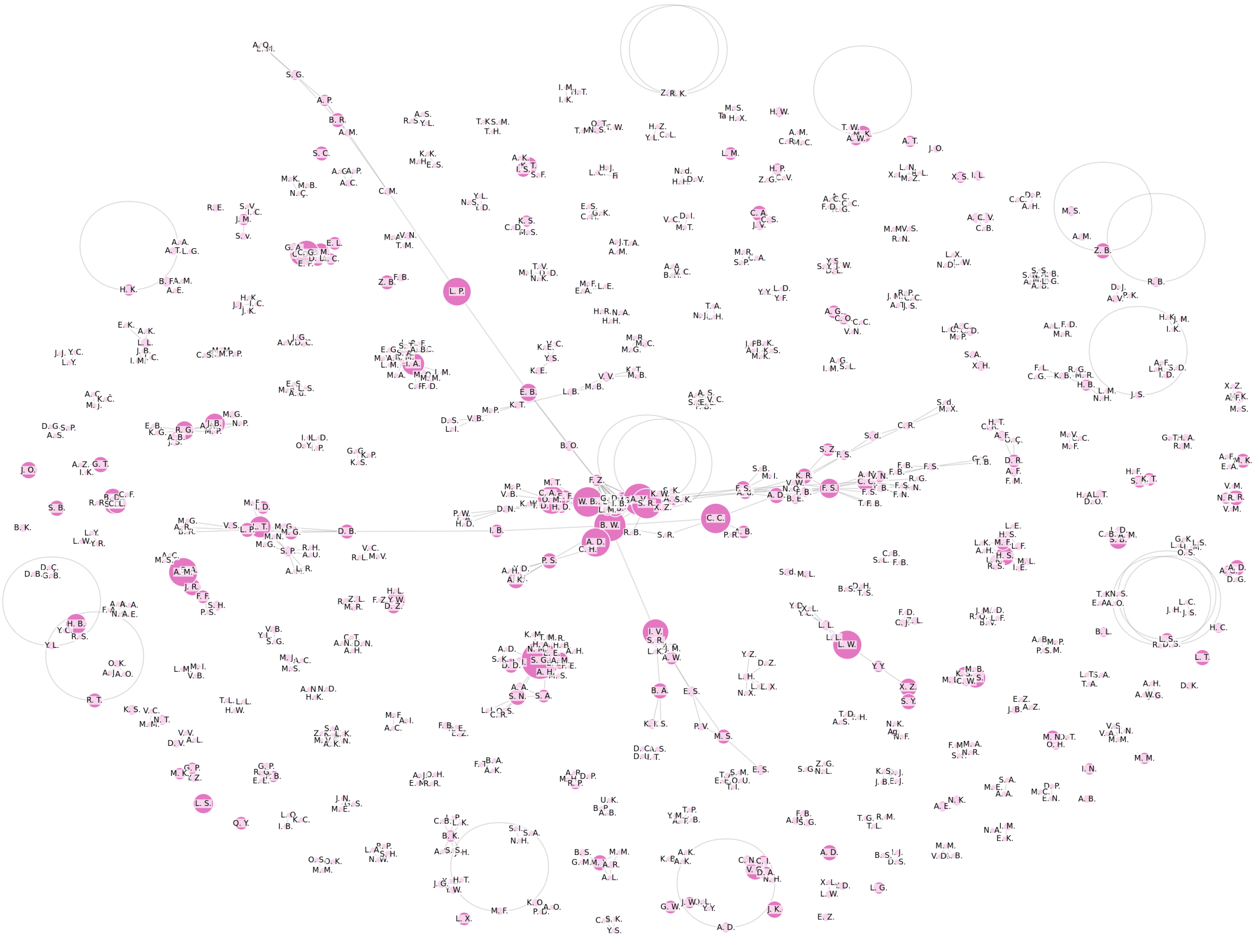


Figure 14. BPM Female-Only Co-authorship Network (Anonymized Initials)

Across all the articles investigated, we found 10,438 distinct keywords. All the keywords extracted from articles were normalized to avoid duplicates caused by case sensitivity. Figure 13 shows the first 20 most common topics. Surprisingly, "Business Process Management" was not the first keyword, appearing in second place in 60.66% of the articles, while other forms of referring to it, as "Business Process" (third place, with 33.88%) and "Process Management" (sixth place, with 12.74%) also entered the podium. The first place was "Enterprise Resource Management", with 64.55%, suggesting more focus on finance, HR, and supply chain areas, within a central database. It was followed in fourth place by "Administrative Data Processing", with 25.68%, and "Information Systems", with 14.42%. Although BPM is a multidisciplinary area, connecting knowledge and practices of business administration, computer science, and economics, almost all keywords are related to technology. This demonstrates a higher interest of female researchers in the computing-related field of study.

5.2.1 *Correlation Analysis*

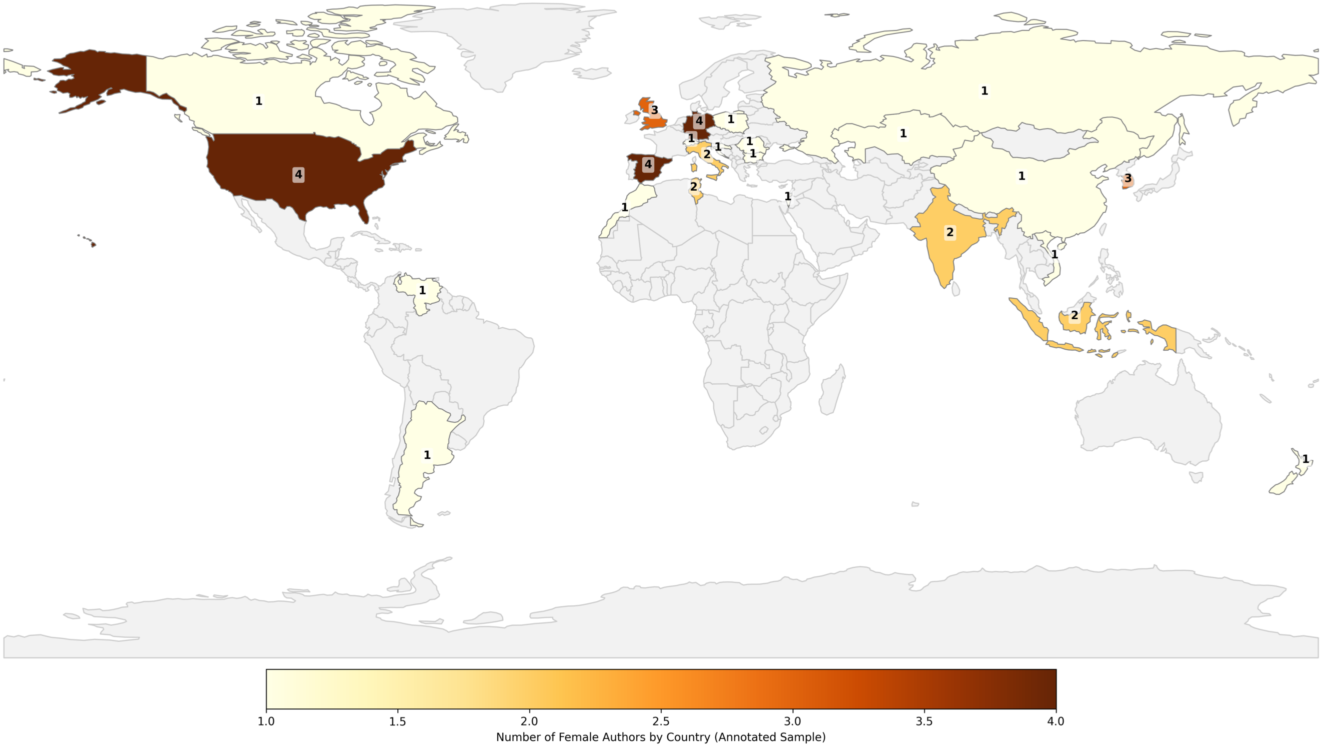
In order to further explore the collaboration trends between the female researchers in the field of BPM, an analysis of the co-authorship network was performed focusing only on the authors who were identified as female with a probability value ≥ 0.75. As can be seen from the topology obtained through this study, the involvement of women in BPM collaboration activities does not revolve around individual efforts; quite the contrary, it revolves around the presence of effective collaborative systems.

Out of the 2,408 unique women authors identified in the study, a substantial percentage (63.4%, $n = 1,528$) has engaged in co-authoring papers in collaboration with other women scientists, thus showing the strong structural homophily that is present in the BPM field, especially among the authors with most publishings. From a visual perspective, the core graph based on the selection of the nodes with a degree centrality threshold set at ≥ 1 clearly shows clusters of collaboration and "hubs" of researchers (with bigger sized nodes representing authors having high productivity) who act as key bridges between the various research communities.

5.2.2 *Geographic Qualitative Analysis*

Furthermore, we sought to investigate the possibility of establishing a relationship between female involvement and geographic locations, aiming to understand tendencies in such contexts. However, this analysis was found to be impractical, as there were considerable limitations in the metadata. The databases, as analyzed, often failed to offer sufficient information regarding affiliation for all co-authors. In most cases, only the publisher's country or the editorial head-

Figure 15. Number of Female Authors by Country (Sample)



quarters of the journal was provided. Additionally, since the scope of this study was quite large, with almost 10,000 authors to consider, verification was found to be impossible, and the use of automated tools for name inference was dismissed to prevent any potential ethical biases and inaccuracies.

However, in order to verify and deepen the analysis constructed with the described tools, we used Cochran's method to calculate, in large populations such as the number of authors, the minimum sample size that would guarantee a desired reliability of the performances and related metrics [Cochran, 1977]. In this case, we used the population of 2408 authors identified as female, with a Z-score of 1.96 for 95% and a precision of 85%, and calculated a minimum of 42 samples, totaling a value that could be easily researched and verified manually.

$$n_0 = \frac{Z^2 \cdot p(1-p)}{e^2} = \frac{1.96^2 \cdot 0.5 \cdot (1-0.5)}{0.15^2} \approx 42.68 \tag{1}$$

$$n = \frac{n_0 \cdot N}{N + (n_0 - 1)} = \frac{42.68 \cdot 2408}{2408 + (42.68 - 1)} = 41.96 \approx 42 \tag{2}$$

The choice of precision was discussed among the team, since the desired results of 95% or more would require 332 samples, which would complicate the process within our small team and make verification more laborious and, consequently, more prone to errors due to fatigue. The 15% margin of error was considered acceptable based on the exploratory nature of the research as defined by [Babbie, 2020], as it seeks to identify previously undocumented patterns of geographic dispersion.

After that, the 42 authors' names were chosen randomly, without using any stratification by productivity or weighting. Since there are many authors with only one publication and very few highly productive authors, following Lotka's Law, more complex approaches could end up distorting representativeness if the strata are too disproportionate [Lotka, 1926]. Given each name, a brief internet search looked for geographical details about each of the authors, with an emphasis on official author pages in journals or publications, or long-term or permanent affiliation with a single educational institution or more than one within the same country.

As a result, we obtained 42 authors from 25 different countries. Given the relatively small manually validated subset of 42 authors, the following geographic and collaboration results should be interpreted strictly as exploratory observations rather than broad generalizations of the entire community. Regarding distribution, the United States, Germany, and Spain lead the sample with 4 authors each, followed by the United Kingdom and South Korea with 3 authors each, and other countries such as India, Indonesia, Tunisia, and Italy contributed 2 authors each, as we can see in Figure 15. As for publications, Spain stands out in total volume with 16 articles, with Israel also noteworthy, with 5 publications from only 1 author in the sample. Figure 16 shows the total number of articles published by the authors in the sample in each country.

The results found are consistent with previous bibliometric studies such as those by [Truong *et al.*, 2023] which cite Germany (4 universities among the institutions that publish the most) and the USA as the countries that contribute most to the total citations in BPM. In general, European

Figure 16. Total Publications by Female Authors by Country (Sample)

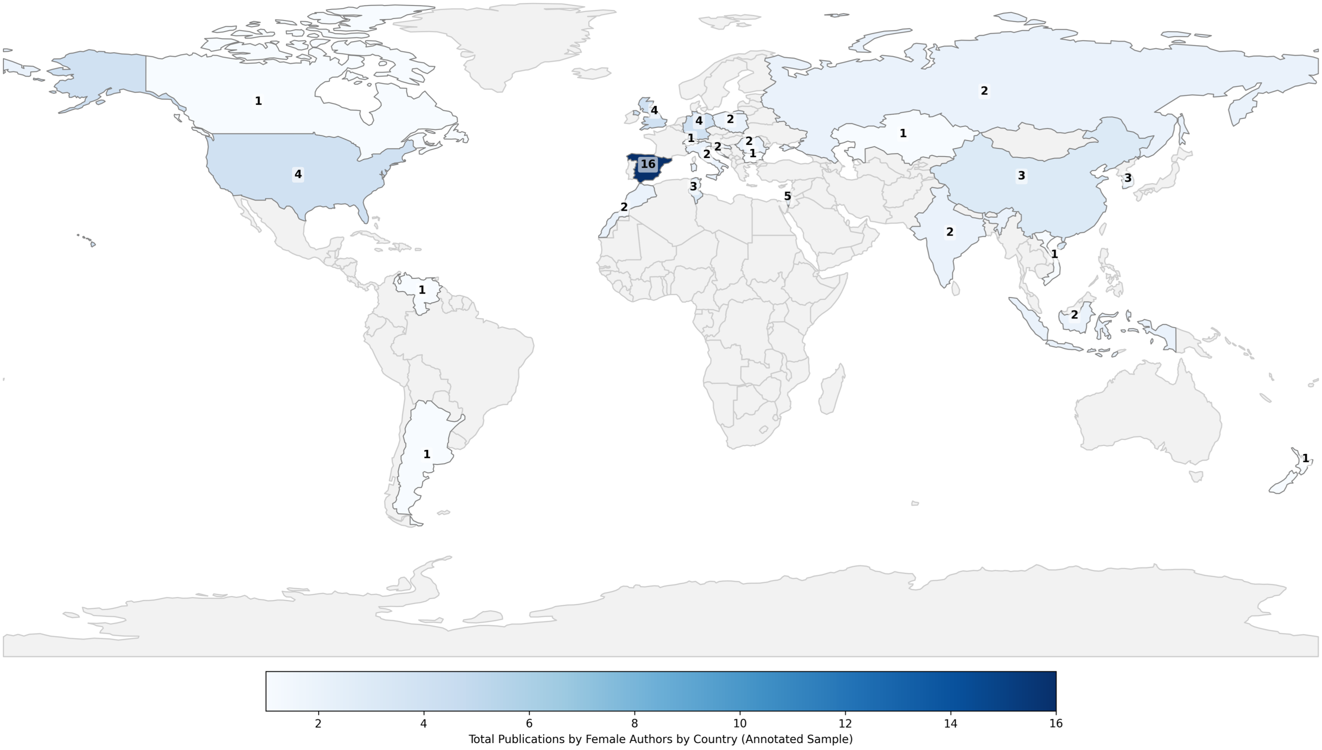


Table 4. Geographic Distribution and Productivity of Female Authors in BPM with more than 1 Publishing (Sample $n = 42$)

Country	Authors	Total Pubs
Spain	4	16
Israel	1	5
USA	4	4
Germany	4	4
UK	3	4
South Korea	3	3
Tunisia	2	3
China	1	3

countries, where there are well-defined academic centers of gravity. However, the emergence of countries such as India, Indonesia, and Tunisia fuels market suspicions regarding the growth of BPM solutions, particularly in the Asia-Pacific region.

These results are in line with what is expected in general, regardless of gender. Europe continues to be the intellectual center of BPM for both men and women, with few other regions around the globe showing similar prominence. However, when analyzing more thoroughly the date of publication of the sample of authors, it becomes evident that the process of geographical decentralization is actually characterized by temporal characteristics. As can be seen from Figure 17, the authors coming from traditional BPM centers such as Spain and Germany have their productivity concentrated during an earlier period, especially in 2012-2015, in contrast with authors from emerging economies like India

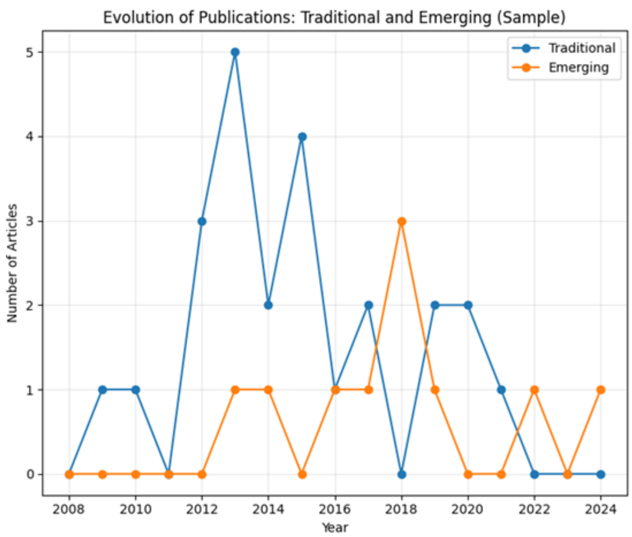


Figure 17. Number of annual publications from traditional and emerging BPM centers according to the sample.

and Tunisia, whose productivity peaked in 2018.

5.2.3 Analysis of Name Identification

Taking advantage of the smaller sample universe, it was possible to have a better visualization of the person's real gender in relation to what was measured by the tool. With this, we were able to observe that five of the 42 names were erroneously classified as feminine (11.9%).

Table 5. List of Gender Misclassifications (False Positives)

Author Name	Country	Given Prob. Female
Bhakti S. O.	UK[1]	99.4%
Nicola M.	Italy	88.2%
Lizhen C.	China	98.8%
In So K.	South Korea	95.0%
Yongsun C.	South Korea	86.6%

[1] Although born in Singapore, his academic residence remains in the United Kingdom.

Names like Bhakti and Yongsun are considered truly neutral and are commonly listed as names for babies of both sexes. Nikola is highly dependent on the region and culture, where it is primarily masculine in countries like Italy, Bulgaria, and Croatia, but also used as a feminine name in Germany, Poland, and the Czech Republic. In these cases, although there are both masculine and feminine cases, it is important to emphasize that it is a flaw in the analysis API, since ambiguous results should present small probabilities.

An interesting case, the name Lizhen, generally neutral, can specifically be considered masculine or feminine depending on the characters used, something that is lost in translation. Thus, it is concluded qualitatively that there is a great need to improve the tool for names of Asian origin, even inferring very high probabilities.

5.3 Threats to Validity

In spite of the thoroughness of the employed methodological framework, certain concerns arise with regard to gender prediction and its generalization.

5.3.1 Construct and Internal Validity

The main threat to validity that arises in the context of this research is associated with the core technique of gender inference from first names. Since automated algorithms are built on the basis of historical census data and social perceptions of gender, a binary classification (either male or female) becomes inevitable, and thus excludes non-binary and transgender identities from consideration. Hence, what is measured in this study is not gender but the social association between names and genders. Moreover, this methodology produces certain systematic internal biases. According to our manually validated sample, the gender detection API misclassifies around 11.9% of names, which tends to occur more often in the case of epicene names and names that follow non-Western conventions. To address potential noise and to guarantee high levels of accuracy, we chose a conservative confidence level of ≥ 0.75, as a result of which around 13.4% of names, namely those that could not be classified or had abbreviations, were removed. Although this methodology strengthens the reliability of our female sample, it may inadvertently underrepresent specific cultural groups whose naming conventions are poorly captured by the APIs' training data.

5.3.2 External and Conclusion Validity

As for generalizability, the external validity of the study is limited due to the nature of the biases in the chosen indexing databases. Despite the fact that these databases are commonly used in BPM research, there is a well-known language and geographical bias that leads to publishing predominantly English and western works. Thus, the participation of females at non-indexed BPM events and non-English journals may be underestimated. These issues affect the conclusion validity of the quantitative panorama provided above. Since a significant amount of information was excluded from the analysis due to its inapplicability to the methodology used, one can state that the statistics obtained represent a conservative estimate of the issue under discussion.

6 Conclusion

In this work, a SMS regarding women influence in the BPM area was provided, analyzing 5,029 scientific papers spanning from 1997 to 2024. Although the gender issue in IT area has been acknowledged in several studies recently, statistics regarding female participation in BPM academic or research fields remains scarce. Addressing the underrepresentation of women might be one way to mitigate the shortage of qualified BPM professionals, foster diverse teams, which existing evidence suggests produce higher-quality scientific outputs, and motivate further policy interventions aiming at organizations for more gender diversity.

Throughout the results from the mapping analysis on our first research question (RQ1) regarding demographic distribution, it was observed that female participation remains significantly lower than male participation across BPM academic and research fields. During the analyzed period, the total number of male authors was almost three times higher than that of female authors. This was, unfortunately, an expected result since this study considers BPM articles applied to the computing area, and gender underrepresentation is a known phenomenon in IT-related fields. Considering the number of BPM authors that have been published since 1997, the total number of male authors is almost three times the number of female authors. Related to the gender authorship distribution, only 7.9% of the articles were published exclusively by women, compared to 47.9% authored exclusively by men. Although the numbers found are far from ideal, an optimistic point to be observed is that 44.2% of publications include collaboration from both genders and that this is a growing trend.

A deeper temporal analysis also exposed nuanced career progression barriers. While the number of female first authors has grown continuously, reaching its peak in 2024, the proportion of women as last authors (typically academic advisors) remains consistently lower, avaraging 18% less participation than as first authors. This discrepancy reflects the known "scissors effect" in STEM, indicating that structural barriers still prevent manu women from advancing to senior academic positions. Interestingly, demographic test utilizing a GLM confirmed that the COVID-19 pandemic did not hinder women's productivity disproportionately, maintaining its growth and suggesting it is driven by structural shifts and diversity policies rather than short-term fluctuations.

Addressing our second research question (RQ2), the profile of female BPM researchers revealed distinct thematic, collaborative, and geographic patterns. Thematically, female researchers demonstrate a strong concentration in highly technological domains, with "Enterprise Resourse Management", "Admnistrative Data Processing", and "Information Systems" emerging among the most frequent keywords. Collaborativelly, the network analysis highlighted a robust structural homophily, with 63.4% of female authors actively engaging in co-authoring with other women, and forming hubs.

Geographically, the study uncovered a temporal decentralization of female BPM research. While traditional academic centers in Europe (e.g., Spain, Germany) and the USA historically dominated the literature, there is a recent emergence of productivity from developing economies, with female researchers showing peak productivity in later years.

The results of this work are useful as a broad panorama of the gender gap in the BPM area. Therefore, the implementation of policies to increase the participation of women is one of the future challenges of the BPM community, along with the identification of contextual factors that explain the results presented in this and other gender-related studies.

Finally, it is imperative to acknowledge the methodological and ethical boundaries of this study. Automated gender inference relies on historical societal name-gender associations, which inevitably imposes a binary framework that cannot account for non-binary or transgender identities. By adopting a conservative confidence threshold (≥ 0.75) and explicitly reporting biases against epicene and non-Western naming conventions, this study represents a rigorous, albeit conservative, estimate of female representation.

Ultimately, the insights obtained from these results serve as a foundation to define concrete actions aimed at promoting the participation of more female researchers in the BPM field. Our bibliometric study aims to support the community by providing a robust empirical baseline for future initiatives. We hope that this exploration of geographic and relational dynamics helps institutions and conference organizers to actively foster these existing networks. By relying on this evidence, stakeholders can develop thoughtful, data-informed initiatives, such as targeted funding for female-led projects, inclusive conference representation metrics, and structured mentorship programs, that not only attract female professionals but also actively dismantle the structural barriers preventing their progression into senior academic roles.

The automated methodology applied in this SMS can be replicated in other domains with considerably low effort since the work detailed here was developed to be as automated as possible. Regarding future work, the limitations and findings encountered in this study suggest several promising directions. First, qualitative studies, such as surveys or interviews, are necessary to understand the contextual factors and structural barriers preventing women from reaching senior authorship positions. Second, future research should develop or integrate more sophisticated demographic inference tools capable of overcoming Western-centric biases, particularly for names of Asian origin, and moving beyond binary gender classifications. Finally, evaluating the citation impact of the identified gender-diverse teams could provide empirical evidence on how diversity actively enhances scientific innovation within the BPM community.

Declarations

Funding

This work was funded by the project PID2024-156158OB-I00 funded by MICIU/AEI/10.13039/501100011033/ and by the European Union FEDER, UE; the 0100_TID4AGRO_4_E project, co-financed by the European Regional Development Fund (ERDF) through the European Commission's INTERREG VI-A Spain-Portugal Program (POCTEP) 2021-2027; Grant File Number: GR24099: this activity has been co-financed 85% by the European Union, the European Regional Development Fund and the Regional Government of Extremadura. Managing Authority: Ministry of Finance.

Record GR24099 Project PID2024-156158OB-100 funded by

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The work of PVA was supported by the National Council for Scientific and Technological Development (CNPq) - PIBIC process number 162123/2025-6.

Authors' Contributions

TM conducted the investigation and data processing, developed the methodology and software, performed formal analysis, produced visualizations, and wrote the manuscript. GRR performed formal analysis of the data, contributed to the investigation, and revised the manuscript with updated conclusions. RBS updated and curated the dataset. PVA refactored and optimized analysis scripts, performed validation, and developed tools for visualization. MESS contributed conceptual input and provided critical review and suggestions. LHT contributed to conceptualization, provided supervision and research guidance, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of research artifacts

All scripts used in the analyses of this study are publicly available in the associated repository, among the files created: `https://github.com/bpm-research-lab-ufrgs/Women-in-BPM`.

Use of Artificial Intelligence

The authors did not use generative AI tools in the development of the study.

Further relevant information

The authors acknowledge the inherent ethical complexities associated with automated gender inference. Inferring gender based on first names relies on historical societal name-gender associations, which inevitably imposes a binary framework and cannot account for self-identified gender, non-binary, or transgender identities. To responsibly manage these ethical implications, we adopted a highly transparent and conservative methodological protocol. First, we established a strict confidence threshold (≥ 0.75), deliberately excluding 13.4% of the dataset (unclassified or ambiguous names) rather than forcing inaccurate classifications. Second, we conducted a manual validation on a statistically representative sample to explicitly identify and report the algorithm's cultural biases, particularly regarding epicene and non-Western naming conventions. By documenting these limitations and treating the results as a measure of societal gender representation rather than absolute identity, we ensure that the methodology respects the ethical boundaries of bibliometric demographic analysis.

References

Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

Alarcón-Alarcón, J., Pérez-Arriaga, J. C., Sánchez-García, Á. J., and Hemández-González, L. A. (2024). Gender diversity in software development teams: A systematic literature review. In *2024 12th International Conference in Software Engineering Research and Innovation (CONISOFT)*, pages 304–312. IEEE. DOI: 10.1109/CONISOFT63288.2024.00047.

Aleem, S. and Ahmed, F. (2023). Practicing equity diversity inclusion (edi) in software development teams: A systematic literature survey. *IEEE Access*, 11:98977–98987. DOI: 10.1109/ACCESS.2023.3312681.

Arrazola, M., Campos, R., and de Hevia, J. (2025). Do labour market outcomes influence why women are underrepresented in engineering? *Higher Education*, pages 1–19. DOI: 10.1007/s10734-025-01445-9.

Arrow, K. (1973). The theory of discrimination. DOI: 10.1515/9781400867066-003.

Babbie, E. R. (2020). The practice of social research. DOI: 10.2307/1318620.

Becker, G. (1957). The economics of discrimination. DOI: 10.2307/1055714.

Bello, M. and Galindo-Rueda, F. (2020). "charting the digital transformation of science". *"OECD Science, Technology and Industry Working Papers, No. 2020/03, OECD Publishing"*. DOI: 10.1787/1b06c47c-en.

Cameron, A. C. and Trivedi, P. K. (2013). Regression analysis of count data. DOI: 10.1017/cbo9781139013567.

Campbell, L. G., Mehtani, S., Dozier, M. E., and Rinehart, J. (2013). Gender-heterogeneous working groups produce higher quality science. *PLOS ONE*, 8:1–6. DOI: 10.1371/journal.pone.0079147.

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why are some stem fields more gender balanced than others? *Psychological Bulletin*, 143:1–35. DOI: 10.1037/bul0000052.

Cochran, W. (1977). Sampling techniques. DOI: 10.2307/2528108.

Dumas, M., Rosa, M., Mendling, J., and Reijers, H. (2018). Fundamentals of business process management. DOI: 10.1007/978-3-642-33143-5.

Ellis, J., Fosdick, B. K., and Rasmussen, C. (2016). Women 1.5 times more likely to leave stem pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLoS ONE*, 11. DOI: 10.1371/journal.pone.0157447.

Felizardo, K. R., Ramos, A. M., de Claudia, C., Érica Ferreira de Souza, Vijaykumar, N. L., and Nakagawa, E. Y. (2021). Global and latin american female participation in evidence-based software engineering: a systematic mapping study. *Journal of the Brazilian Computer Society*, 27. DOI: 10.1186/s13173-021-00109-7.

Gorbacheva, E., Stein, A., Schmiedel, T., and Müller, O. (2015). A gender perspective on business process management competences offered on professional online social networks. *23rd European Conference on Information Systems, ECIS 2015*, 2015-May:1–17. DOI: 10.18151/7217329.

Gorbacheva, E., Stein, A., Schmiedel, T., and Müller, O. (2016). The role of gender in business process management competence supply. *Business and Information Systems Engineering*, 58:213–231. DOI: 10.1007/s12599-016-0428-2.

Hicks, M. (2018). *Programmed inequality*. Mit Press, Cambridge. book.

Holman, L., Stuart-Fox, D., and Hauser, C. (2018). The gender gap in science: How long until women are equally represented? *PLoS Biology*, 16(4). DOI: 10.1371/journal.pbio.2004956.

Jones, G., Chirino Chace, B., and Wright, J. (2020). Cultural diversity drives innovation: empowering teams for success. *International Journal of Innovation Science*, 12(3):323–343. DOI: 10.1108/IJIS-04-2020-0042.

Kerpedzhiev, G. D., König, U. M., Röglinger, M., and Rosemann, M. (2021). An exploration into future business process management capabilities in view of digitalization: results from a delphi study. *Business & Information Systems Engineering*, 63(2):83–96. DOI: 10.1007/s12599-020-00637-0.

Knobloch-Westerwick, S., Glynn, C., and Huge, M. (2013). The matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35(5):603–625. DOI: 10.1177/1075547012472684.

Kohl Silveira, K. and Prikladnicki, R. (2019). A systematic mapping study of diversity in software engineering: A perspective from the agile methodologies. In *2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 7–10. DOI: 10.1109/CHASE.2019.00010.

Kordaki, M. and Berdousis, I. (2020). Identifying barriers for women participation in computer science. *Pro Edu. International Journal of Educational Sciences*, 2:5–20. DOI: 10.26520/peijes.2020.2.2.5-20.

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W., Kidd, C., Marcelle, E., Marciano, D., Martin, S., Myers, N. E., Ojala, K., Perry, A., Pinheiro-Chagas, P., Riès, S. K., Saez, I., Skelin, I., Slama, K., Staveland, B., Bassett, D. S., Buffalo, E. A., Fairhall, A. L., Kopell, N. J., Kray, L. J., Lin, J. J., Nobre, A. C., Riley, D., Solbakk, A.-K., Wallis, J. D., Wang, X.-J., Yuval-Greenberg, S., Kastner, S., Knight, R. T., and Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron*, 109(13):2047–2074. DOI: 10.1016/j.neuron.2021.06.002.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, 16(12):317–323. Available at:`https://www.jstor.org/stable/24529203`.

McKinney, W. *et al*. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX. DOI: 10.25080/Majora-92bf1922-00a.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). Introduction to linear regression analysis. DOI: 10.2307/1269673.

Müller, O., Schmiedel, T., Gorbacheva, E., and vom Brocke, J. (2016). Towards a typology of business process management professionals: identifying patterns of competences through latent semantic analysis. *Enterprise Information Systems*, 10:50–80. DOI: 10.1080/17517575.2014.923514.

NIH (2020). NIH-Funded Research Workforce - Data by gender. Available at:`https://report.nih.gov/nihdatabook/category/16`. [Online; accessed 16-October-2021].

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. *12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008*, (February 2015). DOI: 10.14236/ewic/ease2008.8.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. volume 64, pages 1–18. DOI: 10.1016/j.infsof.2015.03.007.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 4(LXII):659–661. Available at:`https://www.jstor.org/stable/1806107`.

Ralph, P. and Baltes, S. (2022). Paving the way for mature secondary research: the seven types of literature review. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ESEC/FSE 2022, page 1632–1636, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3540250.3560877.

Rodríguez-Pérez, G., Nadri, R., and Nagappan, M. (2021). Perceived diversity in software engineering: a systematic literature review. *Empirical Software Engineering*, 26. DOI: 10.1007/s10664-021-09992-2.

Santamaría, L. and Mihaljević, H. (2018). Comparison and benchmark of name-to- gender inference services. *PeerJ Computer Science*, 2018. DOI: 10.7717/peerj-cs.156.

Schmiedel, T. and vom Brocke, J. (2015). *Business Process Management: Potentials and Challenges of Driving Innovation*, pages 3–15. Springer International Publishing, Cham. DOI: 10.1007/978-3-319-14430-6$_1$.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*. Austin, TX. DOI: 10.25080/Majora-92bf1922-011.

Trauth, E. M. (2011). Rethinking gender and mis for the twenty-first century. *The Oxford Handbook of Management Information Systems: Critical Perspectives and New Directions*. DOI: 10.1093/oxfordhb/9780199580583.003.0025.

Truong, B. Q., Nguyen-Duc, A., and Van, N. T. C. (2023). A quantitative review of the research on business process management in digital transformation: A bibliometric approach. *Software*, 2(3):377–399. DOI: 10.3390/software2030018.

Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F., and Uzzi, B. (2022). Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences*, 119(36):e2200841119. DOI: 10.1073/pnas.2200841119.