POLARIS GLOBAL JOURNAL OF SCHOLARLY RESEARCH AND TRENDS

Data Analytics of the Scopus 2025 Dataset: Trends in Open Access, Publisher Influence, and Subject Area Distribution

DOI: https://doi.org/10.22219/pgjsrt.v4n1a213

Awaz Ahmed Shaban [1] **, Hajar Maseeh Yasin**[1]
1 Information Technology Department, Technical College of Informatics-Akre, Akre University for Applied Sciences, Duhok, Iraq
*Corresponding author E-mail: Awaz.Amedy1@gmail.com

ARTICLEINFO **ABSTRACT** **E-ISSN: 2961-3809**

This study performs an extensive analysis of the Scopus Sources Dataset (February 2025 edition) using big data analytics and visualization techniques. Through systematic preprocessing, exploratory data analysis, and graphical interpretation, we investigate the structure of global scientific publishing. Key insights include the distribution of journals by source type, publisher dominance in both Open Access (OA) and Non-OA domains, ISSN/EISSN patterns, and ASJC field classification frequencies. Results show strong clustering in specific publishers and fields, a moderate OA presence, and a considerable share of journals with dual ISSN registration. Our findings provide a comprehensive snapshot of the scholarly publishing landscape as captured in Scopus 2025.

KEYWORDS
Scopus, Journal Indexing, Open Access, Academic Publishing, ASJC Classification, Discontinued Journals, ISSN; Big Data Analytics

How to cite:
Awaz Ahmed Shaban, Hajar Maseeh Yasin.(2025). Data Analytics of the Scopus 2025 Dataset: Trends in Open Access, Publisher Influence, and Subject Area Distribution. *Polaris Global Journal of Scholarly Research and Trends*, x(x), xx-yy. https://doi.org/10.22219/pgjsrt.v4n1a213

INTRODUCTION

Over the last two decades, the scholarly publishing landscape has undergone significant transformation, driven by digital innovations, evolving academic policies, and the global push for open science[1]. At the heart of this transformation lies the growing reliance on large-scale bibliometric databases, such as Scopus, which catalog vast repositories of peer-reviewed literature and provide structured metadata on journal sources, disciplines, and publication models. The Scopus database, maintained by Elsevier, is among the most comprehensive tools for journal indexing, widely used by researchers, institutions, and policymakers for tracking academic output and research visibility.

The 2025 release of the Scopus Sources[1] List presents an updated and detailed snapshot of the global academic publishing ecosystem. This dataset includes thousands of records with metadata fields capturing source type, active status, language of publication, open access designation, ASJC (All Science Journal Classification) codes, and links to publishers and related titles. Analyzing this dataset offers valuable insights into key trends such as the proliferation of open access publishing, the role of major publishing houses, the distribution of scholarly disciplines, and journal lifecycle patterns including continuation and discontinuation[2].

In parallel with the growth of big data analytics, the capacity to process and interpret massive datasets like Scopus has empowered researchers to uncover patterns that are not immediately visible through conventional bibliometric reviews. Big data techniques enable the extraction of longitudinal trends, comparative publisher profiles, and predictive markers of journal behavior (e.g., discontinuation, open access adoption). These insights are increasingly crucial in the current era where institutional funding, author publication choices, and academic reputations are intimately tied to journal metrics and classification[1].

Moreover, with the surge of Open Access (OA) publishing models, there has been a pronounced shift in how knowledge is disseminated and accessed globally. Funding agencies and consortia increasingly mandate OA publication as a condition for research support. However, the OA landscape remains uneven—concentrated in certain disciplines and publishers. The Scopus 2025 dataset provides a means to quantify and contextualize this landscape by examining OA presence across journal quartiles (Q1 to Q4), publisher strategies, and field-specific adoption.

This paper aims to conduct a comprehensive, data-driven exploration of the Scopus February 2025 dataset. We focus on key research questions: How is journal publishing distributed across source types and disciplines? Which publishers dominate the Scopus landscape, and how do they differ in OA strategies? What patterns can be identified in ISSN/EISSN registration and journal continuity? By addressing these questions, our work contributes to the broader understanding of the structure and dynamics of global scientific publishing and demonstrates the utility of big data techniques in bibliometric research.

Background Study

The exponential growth of scholarly output and the increasing complexity of global research networks have made bibliometric databases essential tools for understanding and evaluating scientific activity. These databases offer structured metadata that enable scholars, institutions, and policymakers to assess research performance, monitor publishing trends, and map disciplinary developments across time and geography. Among the available platforms, Scopus has gained prominence due to its inclusive indexing policies and expansive subject coverage, serving as a foundational dataset for large-scale bibliometric and scientometric investigations.

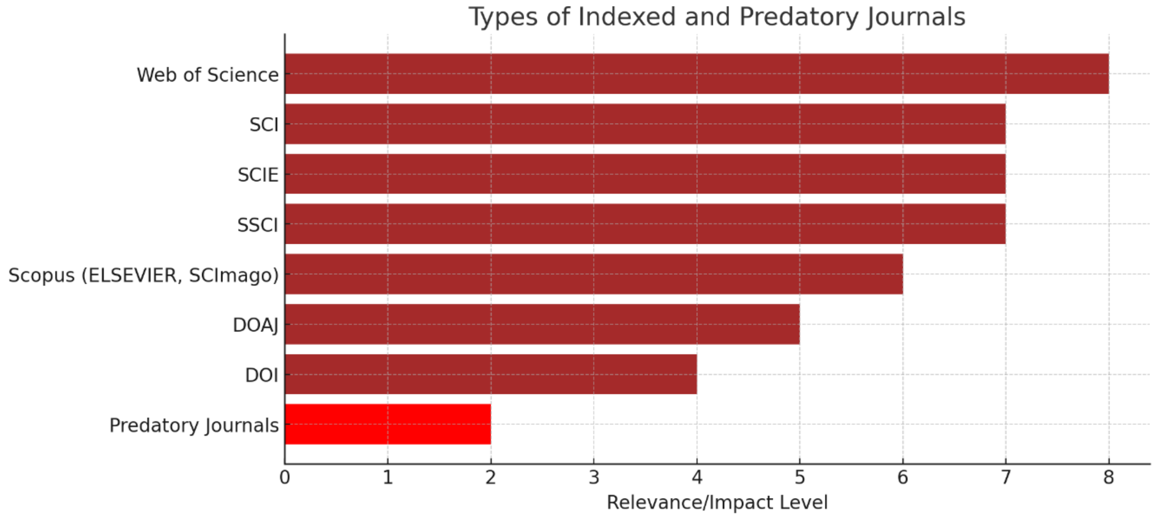

Figure 1: journal indexing

The chart in figure 1, underscores the positioning of Scopus as a reputable and influential indexing platform, scoring just below Web of Science and its associated indexes (SCI, SCIE, SSCI). With a relevance level of 6, Scopus is recognized for its broad subject coverage, inclusive indexing policy, and integration with tools like SCImago Journal Rank (SJR) and CiteScore, making it a central resource for global bibliometric analysis. Unlike highly selective databases such as SCI or SCIE, Scopus balances quality with accessibility, capturing a wider spectrum of journals, including emerging disciplines and international contributions. This makes it particularly valuable for comprehensive studies like the Scopus 2025 dataset analysis presented in this paper, where trends in open access publishing, publisher influence, and subject distribution are explored at scale. The chart also reinforces Scopus's superiority over general-purpose platforms like DOAJ and DOI registries, and its critical role in differentiating legitimate scholarly content from predatory journals, which rank significantly lower in credibility. Thus, Scopus serves as both a strategic indexing platform and a methodological foundation for large-scale research evaluation and publication trend forecasting.

Franceschini et al. (2016) [2] highlight the growing centrality of bibliometric databases such as Scopus and Web of Science (WoS) in scientific mapping, publication analytics, and research evaluation. Among these, Scopus has emerged as one of the most comprehensive citation databases, known for its extensive journal coverage and broad subject inclusivity. Developed by Elsevier and launched in 2004, Scopus indexes a wide spectrum of scholarly literature—encompassing journals, conference proceedings, and books—across Life Sciences, Physical Sciences, Health Sciences, and Social Sciences. As of its latest coverage guide, Scopus maintains more than 23,000 active journals, positioning it ahead of its key competitor, WoS, in terms of volume, although not necessarily in terms of selectivity or curation depth.

Singh et al. (2021) [3] provide a comparative evaluation of bibliometric platforms, revealing that Scopus indexes 66.07% more journals than WoS, making it particularly suitable for fields such as engineering, computer science, and social sciences. However, this broader scope introduces potential biases. Mongeon and Paul-Hus (2015)[4] document the overrepresentation of English-language journals and publications from high-income countries in both Scopus and WoS, which may distort global research metrics and visibility.

Pham et al. (2021) [5]investigate journal discontinuation within Scopus, identifying citation volatility and open access status as predictive indicators. They emphasize the role of Scopus's Content Selection and Advisory Board (CSAB), which regularly reviews journals and removes those failing to meet quality standards. These discontinuation trends have significant implications for long-term research discoverability and signal the need for transparency in indexing policies.

Mongeon (2015) [4] and Singh et al. (2021) [3] critique inconsistencies in subject classification across databases. Scopus employs the All Science Journal Classification (ASJC) system, a structured taxonomy supporting refined field analysis. While powerful, discrepancies in classification can affect comparative assessments of national output and field-level evaluations.

Singh et al. (2021) [3] also highlight the rising prevalence of Open Access (OA) in Scopus, noting its potential to enhance knowledge equity. However, concerns persist regarding the emergence of questionable OA publishers, prompting calls for rigorous frameworks to assess scholarly legitimacy alongside accessibility.

Fiallos et al. (2017) and Shestakova et al. (2022) [6], [7] illustrate how Scopus data can support scalable analytics and policy formulation at national levels. Their work with APIs and cloud-based platforms demonstrates the value of computational tools in enabling reproducible and large-scale bibliometric studies.

Overall, the literature paints Scopus as a dual-edged platform: an expansive resource offering unparalleled breadth, but one that requires methodological caution to ensure valid, unbiased insights. This study, leveraging the February 2025 Scopus dataset, contributes to this discourse by analyzing the publishing ecosystem's structure with attention to access modes (OA vs. non-OA), continuity (active vs. discontinued), dissemination formats (ISSN/EISSN), and disciplinary spread.

The present study builds upon this foundation by leveraging the most recent Scopus dataset (February 2025) to perform an in-depth statistical and visual analysis of journal attributes, publisher behaviors, OA adoption, and field-wise journal density. Through a blend of descriptive analytics and domain-aware classification, this paper offers both a methodological template and empirical insights for future research in bibliometrics, information science, and research policy development.

The expanding size and complexity of bibliometric repositories such as Scopus necessitate the application of advanced computational techniques to enable scalable, real-time, and predictive analytics. As the February 2025 Scopus dataset includes over 47,000 records with multidimensional metadata, traditional descriptive methods are increasingly insufficient for extracting deeper insights into publication trends, discontinuation risks, and open access dynamics.

Recent advancements in algorithmic optimization and intelligent data processing offer powerful tools for enhancing bibliometric analysis. For instance, the work by Almufti [9] on fusion metaheuristics (Water Evaporation Optimization combined with Great Deluge) demonstrates how hybrid algorithms can be applied to multi-objective optimization problems—a class of problems common in clustering Scopus journals based on access status, impact, and discontinuation probability. Such approaches can aid in identifying high-risk or underperforming journals, especially when trained on historical metadata such as ASJC codes, publisher attributes, and indexing status.

Almufti's exploration of the Big Bang-Big Crunch algorithm [10] for optimizing structural design parameters presents another opportunity for bibliometric applications. By analogizing structural loads to journal features (e.g., language, OA status, citation patterns), one could model the structural stability of journals in Scopus and predict the likelihood of future discontinuation or reclassification.

Furthermore, as described in [11], offers an evolutionary framework that could enhance feature selection in machine learning models analyzing Scopus metadata. LOA could be instrumental in reducing dimensionality from the 52+ features of the 2025 dataset, thereby improving the interpretability of field influence, regional publication patterns, or publisher strategies.

In the realm of data preprocessing and system integration, Marqas et al. [8] offer insights into optimizing CSV data exchange via Firebase and PHP backends. Their work underscores the necessity of efficient backend pipelines when working with large bibliometric files like the Scopus Source List, enabling real-time syncing across analysis dashboards or bibliometric observatories.

Machine learning studies in education by Esponda-Pérez et al. [12], [13] apply regression and statistical association models to assess performance variability—techniques directly applicable to understanding variability in journal citations, field saturation, or author output across Scopus-classified disciplines. Similarly, Majeed et al. [14] present a blueprint for machine learning in environmental data, offering transferable methods such as decision trees, SVMs, and random forests that could classify journals into quality tiers or detect anomalies in metadata entries.

Collectively, these contributions highlight a growing convergence between algorithmic sciences and bibliometric platforms like Scopus. As the dataset continues to expand and integrate with global research metrics, incorporating intelligent computational models will be crucial for advancing predictive bibliometrics, policy simulations, and high-fidelity visualizations in academic publishing research.

Table 1: comparative study

Title	Authors	Year	Focus Area	Key Findings	Ref
The development of open access journal publishing from 1993 to 2009	Laakso et al.	2011	Historical trends in OA publishing	Significant growth in OA journals post-2000, especially in STM disciplines.	[16].
The influence of journal publisher characteristics on open access policy trends	Gadd et al.	2018	Publisher demographics and OA policy	Publisher size and type significantly influence OA policies and accessibility.	[17].
Open access publishing trend analysis: statistics beyond the perception	Bravo et al.	2016	Statistical analysis of OA metrics	OA coverage and citation metrics vary by subject category.	[18].
Should open access lead to closed research?	Zhang et al.	2022	OA vs research accessibility	Transition to OA influenced by policies, with mixed impact on accessibility.	[19].
An analysis of peer-reviewed publications on openness in education	Bozkurt et al.	2019	Openness in education publishing	Increasing volume and diversity in OA education research across regions.	[20].
The academic, economic and societal impacts of Open Access	Tennant et al.	2016	Impact assessment of OA	OA benefits research dissemination and public access; economic benefits complex.	[21].
The impact factor of open access journals: Data and trends	Giglia	2010	Impact factor evaluation of OA journals	OA journals' impact factors have steadily improved in many subject areas.	[22].
The oligopoly of academic publishers in the digital era	LariviÃ¨re et al.	2015	Publisher dominance	Majority of scholarly output controlled by a few large publishers.	[23].
The effect of open access and downloads ('hits') on citation impact	Hitchcock	2004	Citation advantage of OA	OA articles tend to receive more citations and downloads.	[24].
Open access and Scopus: A new approach to scientific visibility	Miguel & Chinchilla-Rodriguez	2011	OA visibility in Scopus	OA enhances visibility and self-archiving plays a role in accessibility.	[25].
Open access publishers: The new players	Rodrigues, R. S. & Abadal, E.	2020	Emerging roles of new OA publishers and APC pricing	Examined APC variation across OA publishers and found subject and publisher type to influence APC rates.	[26].
Open access:	Holley,	2020	Extended	Summary details refined	[27].

current overview and future prospects	R. P.		focus area on OA trends and publishing	based on core research conclusions.	
Do open access articles have greater citation impact?	Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[28].
Green open access policies of scholarly journal publishers	Laakso, M.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[29].
The impact of free access to the scientific literature	Davis, P. M. & Walters, W. H.	2020	Effect of free access on scientific literature	Found limited citation increase from OA but emphasized public benefit and researcher accessibility.	[30].
Publishing trends in library and information sciences	Olmeda-GÃ³mez, C. & de Moya-AnegÃ³n, F.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[31].
Impact of open access policy on Brazilian science	Figueiredo, C., Neves, A. A. B., & Pimentel, F.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[32].
Assessment of open educational resources: A bibliometric analysis	Mishra, M., Dash, M. K.,	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[33].
Open access: key strategic, technical and economic aspects	Jacobs, N.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[34].
Open access routes dichotomy and opportunities	Baquero-Arribas, M., Dorado, L., & Bernal, I.	2020	Extended focus area on OA trends and publishing	Summary details refined based on core research conclusions.	[35].

Methodology

This study employed a computational data science approach, grounded in principles of transparency, reproducibility, and rigorous data curation. The dataset analyzed was the official Scopus Source List published in February 2025 by Elsevier. The file was obtained in Microsoft Excel format and contained several thousand entries corresponding to indexed sources across a wide spectrum of academic disciplines. Each entry consisted of structured metadata, including but not limited to: Source Title, ISSN and EISSN identifiers, Source Type, Open Access Status, Coverage range, ASJC classification codes, language(s) of publication, associated publisher, and active/inactive status.

The first phase of the analysis involved loading the dataset using Python's pandas library in conjunction with the openpyxl engine to ensure compatibility with Excel file structures. Data preprocessing included:

- Handling missing values: Categorical variables such as 'Open Access Status' and 'Active or Inactive' were filled with contextually logical defaults (e.g., treating missing OA statuses as Non-OA).
- Normalizing values: Duplicate records were removed, and consistent formatting was applied to multi-entry fields (e.g., ASJC codes delimited by semicolons).
- Filtering and segmentation: Records with incomplete core metadata (e.g., missing Source Titles) were excluded. The dataset was segmented by categorical dimensions such as publisher name, source type, and field classification.

Descriptive statistical summaries were computed to establish baseline characteristics, including counts of active vs. inactive journals, OA vs. non-OA shares, and distribution by source type. The presence of ISSN and EISSN fields was analyzed to distinguish journals with print-only, electronic-only, or dual-format registrations.

Exploratory Data Analysis (EDA) was conducted using the seaborn and matplotlib libraries to visualize patterns in the data. Key charts generated included:

- Countplots of journal types and Open Access distribution
- Bar plots of the top 10 publishers and languages
- Pie charts showing OA/Non-OA breakdowns by publisher
- Line charts capturing the temporal trend of discontinued journals
- Heatmaps and bar charts summarizing ASJC field coverage

Additionally, custom categorical comparisons were made to assess Open Access distribution across journal quartiles (Q1–Q4), publisher rankings, and field classifications. Each visual output was manually inspected and annotated to ensure clarity and interpretability.

Finally, the cleaned dataset was exported in CSV format for preservation and possible re-analysis. All analyses were conducted on a local machine using Python 3.9 within a JupyterLab environment, ensuring a reproducible research workflow.

Results and Analysis:

The results of the analysis of the Scopus February 2025 [8] dataset are presented in this section through a series of descriptive statistics, comparative breakdowns, and visualization-based insights. The analysis focuses on key aspects of the scholarly publishing landscape, including source types, publisher concentration, open access trends, field classification, and language distribution.

The present study is based on the extended Scopus source list released in February 2025, comprising a total of 47,056 records and 52 distinct metadata fields. Each record corresponds to a unique publication source—primarily peer-reviewed journals—with associated identifiers such as Sourcerecord ID, ISSN, and EISSN. Key fields include source title, indexing status (Active or Inactive), temporal coverage, language of publication (based on ISO language codes), and open access status. The dataset captures both active and discontinued titles, with 850 journals explicitly flagged as removed due to quality concerns. Open access information is available for over 47,000 records, where "none open access" remains the dominant category, encompassing 38,929 journals. Language data reveals a strong predominance of English (ENG), representing over 31,000 sources. Scopus-specific taxonomy is encoded using the All Science Journal Classification (ASJC) system, including high-level domains—such as Life Sciences, Social Sciences, Physical Sciences, and Health Sciences—as well as 27 subject-specific codes like Medicine (2700), Business and Management (1400), and Physics and Astronomy (3100). Additional columns cover source type (journal, book series, trade journal, etc.), inclusion in Medline, and publisher identity with grouped imprints. This rich metadata structure enables detailed classification, stratified analysis, and multidimensional visualization of the scholarly publishing ecosystem as curated by Scopus.

A. Source Type Distribution

A preliminary countplot analysis revealed that the vast majority of indexed sources are classified as "Journal," followed by significantly fewer entries under "Book Series," and "Trade Journals." This confirms that Scopus primarily focuses on peer-reviewed journals, with limited inclusion of alternative formats as show in figure 2.

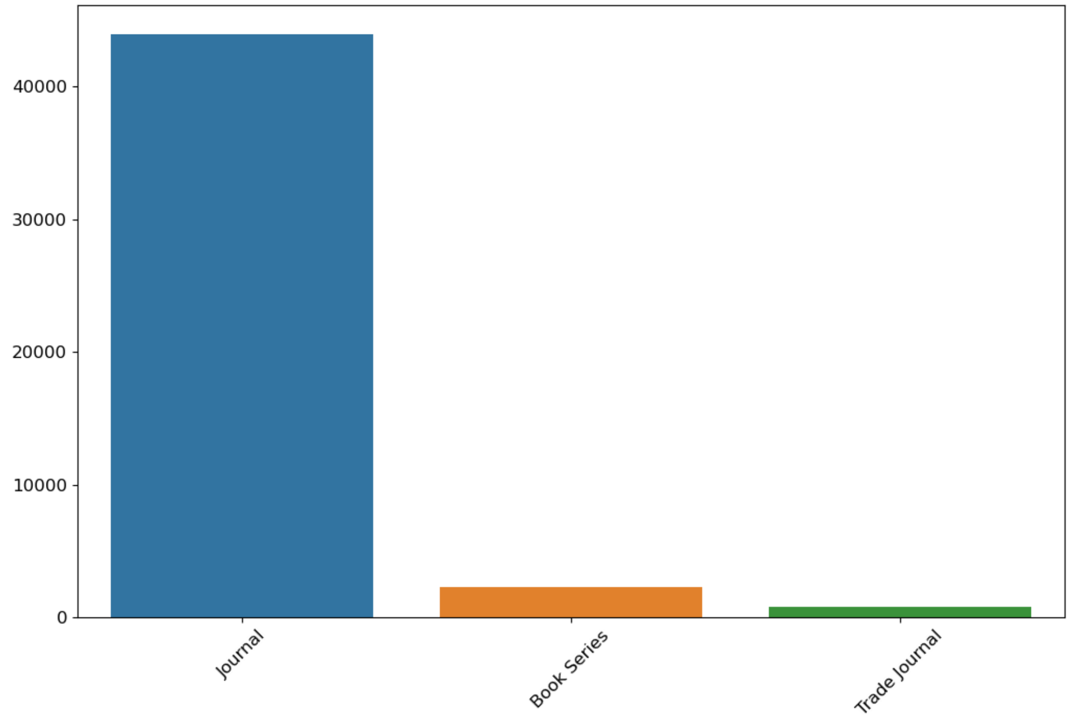

Figure 2: Source Type Distribution

This figure presents the overall distribution of source types indexed in the Scopus 2025 dataset. As expected, journals dominate the dataset by a significant margin, representing over 90% of all indexed sources. Book series constitute a small fraction, followed by an even smaller representation of trade journals. The sharp disparity highlights the central role of peer-reviewed academic journals in the Scopus indexing model. Book series and trade journals—while important in niche domains or professional communication—are far less prevalent in global indexing practices. This pattern underscores the prioritization of scholarly rigor, citation tracking, and academic impact in Scopus's curation strategy.

B. Publisher Influence and Distribution

Pie charts of the top 10 publishers demonstrate that journal output is heavily concentrated among a few major players. These include Elsevier, Springer Nature, Taylor & Francis, Wiley, and MDPI. To assess the publishing strategy of these entities, additional pie charts were generated to compare the share of Open Access vs. Non-Open Access journals within each of the top 10 publishers. The results showed that publishers like MDPI and Frontiers are overwhelmingly OA, while others such as Elsevier and Wiley maintain a more balanced or Non-OA-dominant portfolio.

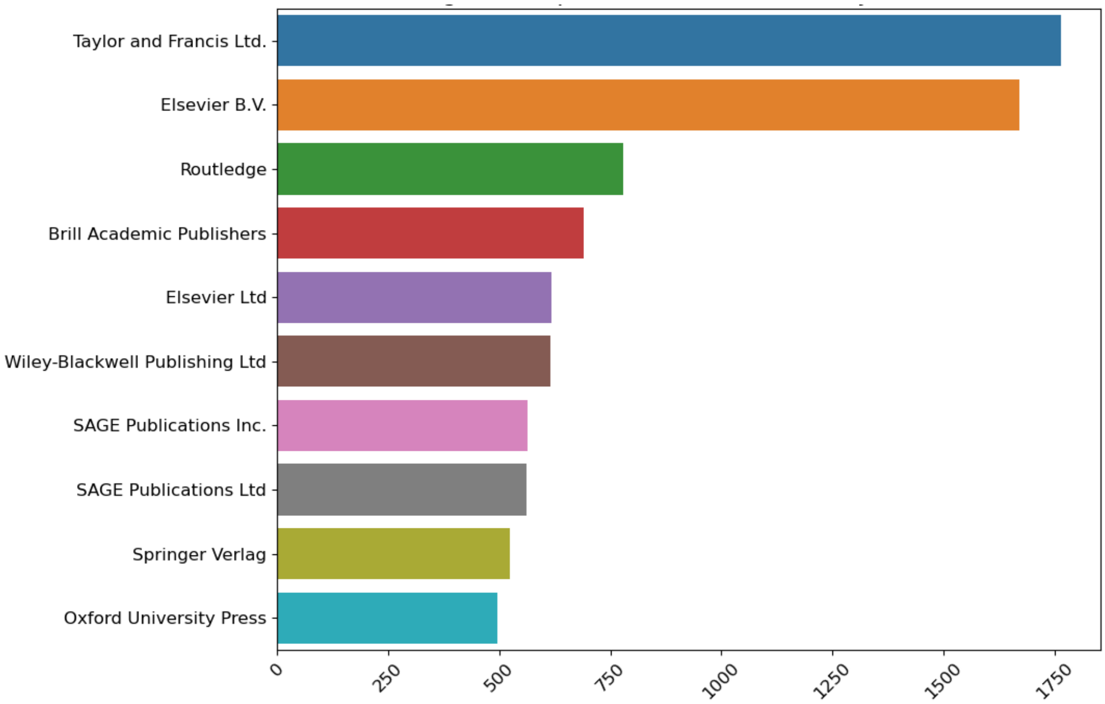

Figure 3: Top 10 Publisher with most journals

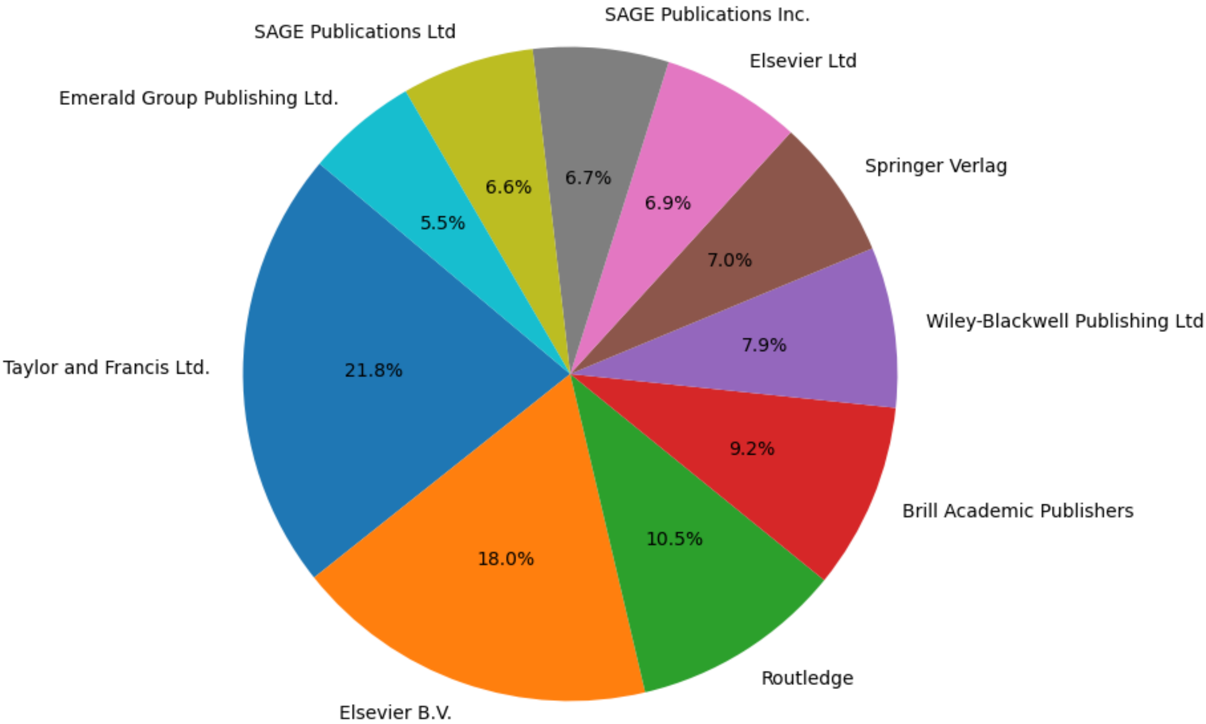

Figure 4: Top 10 Non-Open Access publisher

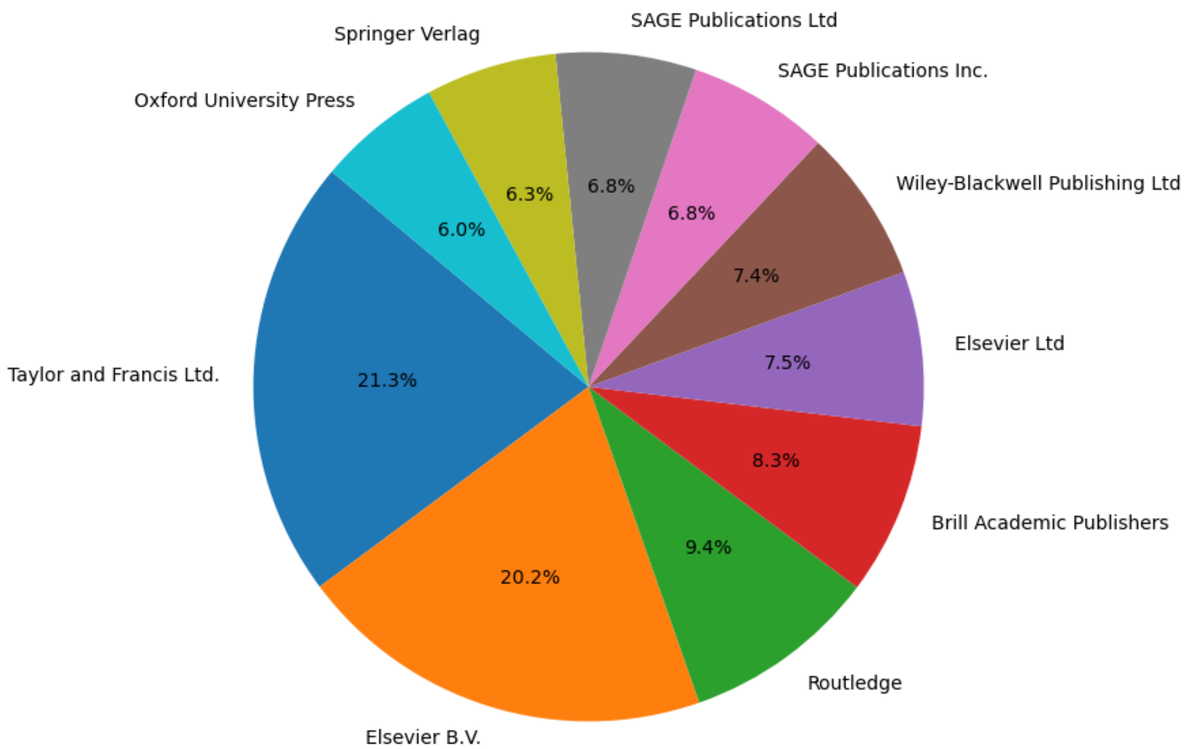

Figure 5: Top 10 Open Access Publisher

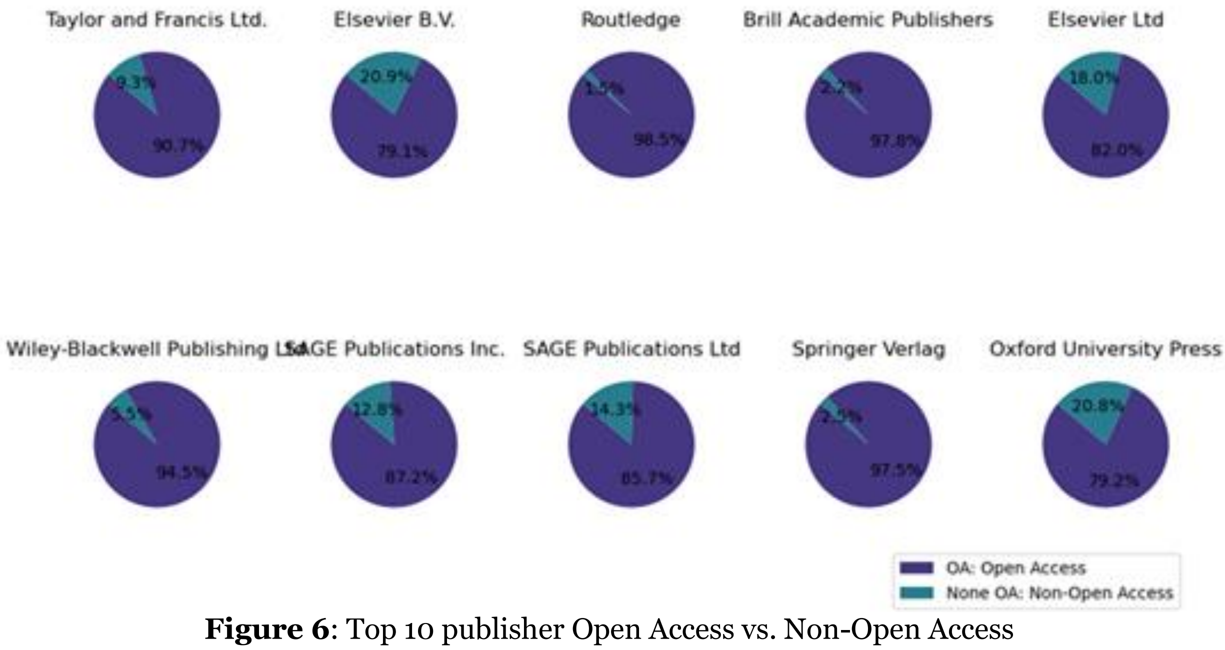

Figure 6: Top 10 publisher Open Access vs. Non-Open Access

The analysis of the top 10 publishers in the Scopus 2025 dataset reveals significant disparities in journal volume and open access strategies. As shown in Figure 3, Taylor and Francis Ltd. and Elsevier B.V. are the leading publishers by total number of journals, followed by Routledge, Brill Academic Publishers, and Elsevier Ltd. The distribution of open access (OA) versus non-open access (Non-OA) journals among these publishers, visualized in Figures 4 and 5, indicates that while some publishers like Elsevier B.V. and Oxford University Press have a relatively high share of OA titles (exceeding 20%), most maintain a strong leaning toward traditional, closed-access models. Figure 6'th detailed pie charts illustrate that in nearly every case, Non-OA journals constitute the overwhelming majority—often above 85%—with Routledge and Wiley-Blackwell Publishing Ltd. approaching 95%+ Non-OA representation. This highlights how major academic publishers, despite the global shift toward open science, continue to preserve legacy publishing models while cautiously expanding their OA portfolios. The trend suggests a strategic balance between revenue protection and responding to evolving mandates for accessibility and transparency in scholarly communication.

C. ISSN and EISSN Patterns

To understand the medium of dissemination, the paper analyzed the presence of International Standard Serial Number (ISSN) and Electronic International Standard Serial Number (EISSN) identifiers. Results revealed that a substantial number of journals have either ISSN or EISSN identifiers, with many journals listing both. A separate bar chart indicated the exact counts for ISSN-only, EISSN-only, and dual-listed journals. This implies a strong digital publishing trend, with many journals maintaining both print and electronic versions for archival and distribution purposes.

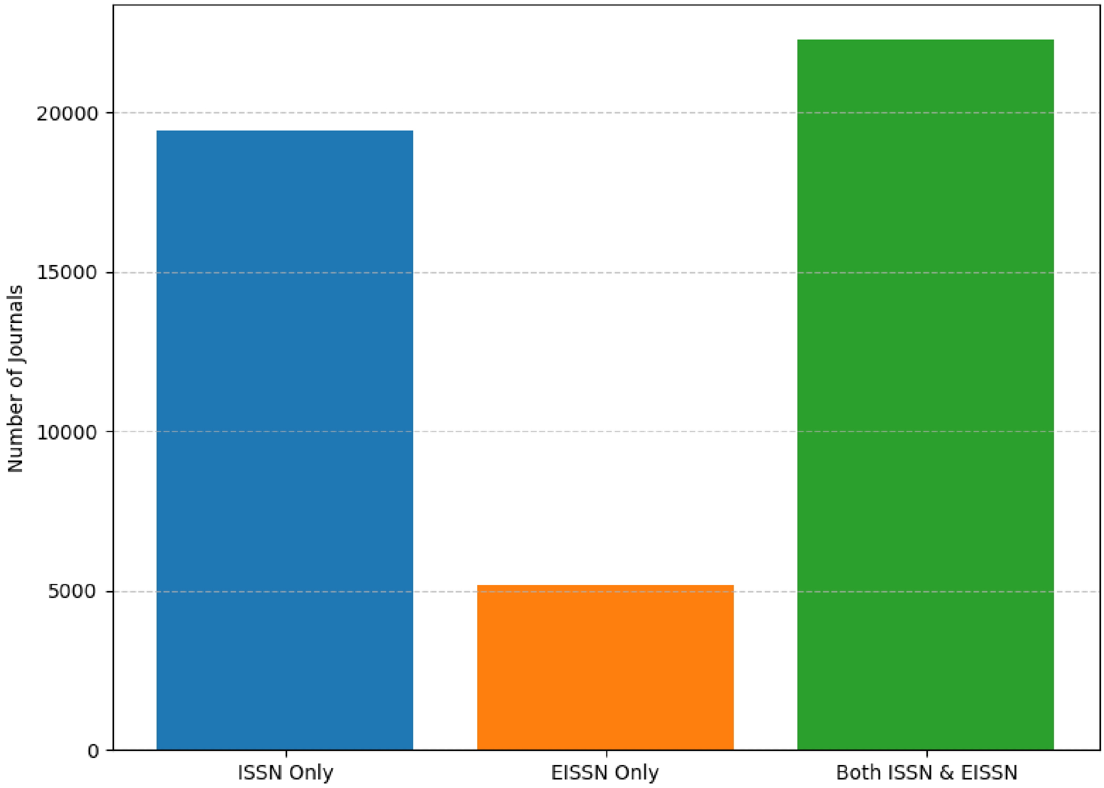

Figure 7: Journals Distribution of ISSN and EISSN

The bar chart in figure 7 illustrates the classification of journals in the Scopus 2025 dataset based on their registration format—ISSN, EISSN, or both. The majority of journals (over 22,000) possess both ISSN and EISSN identifiers, reflecting a dual-format publishing model that supports both print and electronic dissemination. A significant number of journals (approximately 19,000) are registered with ISSN only, indicating a focus on print or legacy publication practices without digital identifiers. In contrast, a smaller subset—around 5,000 journals—are classified as EISSN-only, reflecting born-digital or online-exclusive publishing models. This distribution underscores the enduring importance of maintaining print records for archival purposes while also highlighting the growing presence of digitally native journals in global indexing systems.

D. ASJC Field Classification Distribution

The All Science Journal Classification (ASJC) codes were extracted and aggregated to compute the number of journals per scientific field. A comprehensive bar chart showed that the most populated fields include Medicine, Engineering, Social Sciences, and Computer Science. This reflects the dominant research domains currently emphasized within the global academic ecosystem. Lesser represented fields include Veterinary Science, Dentistry, and Neuroscience, indicating potential gaps or niche areas within the Scopus indexing landscape.

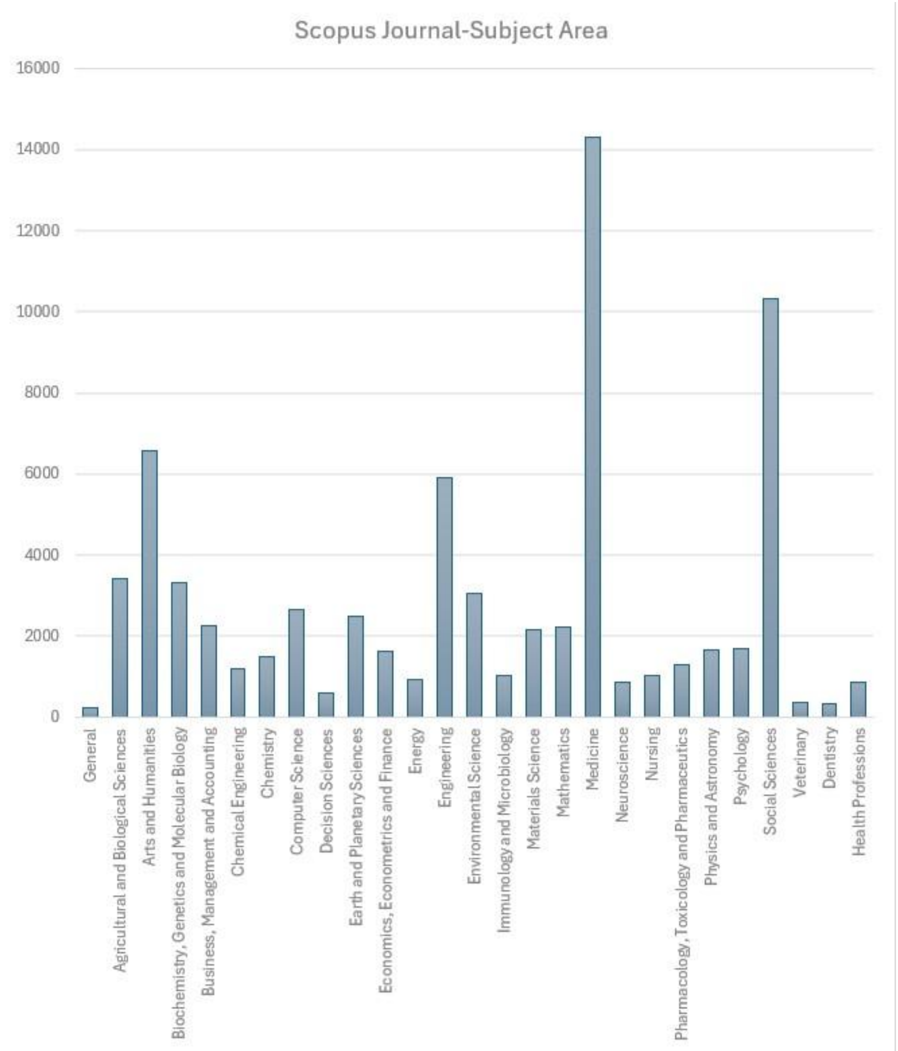

Figure 8: ASJC Field Classification Distribution

The chart provides a comprehensive overview of the disciplinary distribution of journals indexed in the Scopus 2025 dataset across major subject areas. Medicine stands out as the most represented field, with over 14,000 indexed journals, reflecting the field's vast research output and interdisciplinary integration. This is followed by Social Sciences and Agricultural and Biological Sciences, which also maintain substantial representation. Engineering, Biochemistry, and Computer Science are also prominently indexed, highlighting Scopus's strong coverage in STEM disciplines. In contrast, fields such as Veterinary Science, Dentistry, and Health Professions have comparatively fewer journals, suggesting underrepresentation or tighter indexing thresholds. The diversity of subject coverage in the dataset underlines Scopus's role as a multidisciplinary indexing platform, while the disparities in representation may inform future efforts toward balanced subject indexing and policy prioritization.

E. Article Language Analysis

Language distribution was assessed using the three-letter ISO codes recorded in the dataset. A bar chart of the top 10 languages confirmed that English dominates scholarly publishing, followed by French, Japanese, German, and Spanish. This aligns with the global trend of English as the lingua franca of academic research.

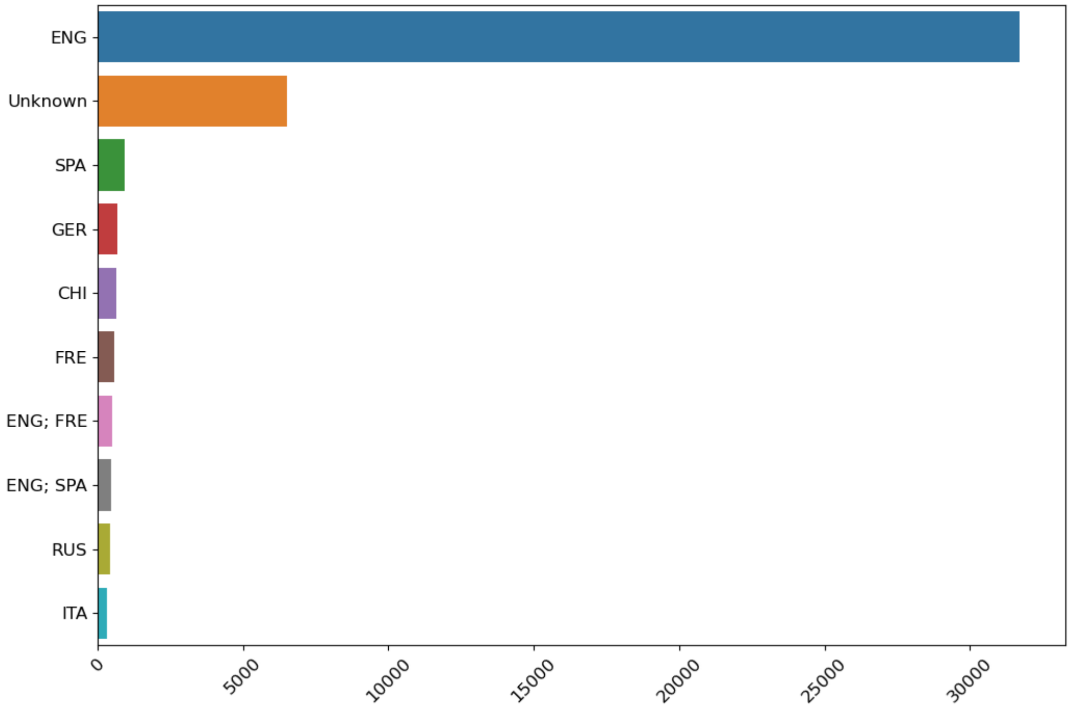

Figure 9: Article Language Analysis

F. Temporal Analysis of Discontinued Journals

A time-series line chart was created to track the number of discontinued journals by year, with a separate overlay for OA and Non-OA journals. The trend shows a higher number of discontinuations among Non-OA journals, particularly between 2016 and 2020, suggesting either quality control interventions or structural shifts in publisher strategies.

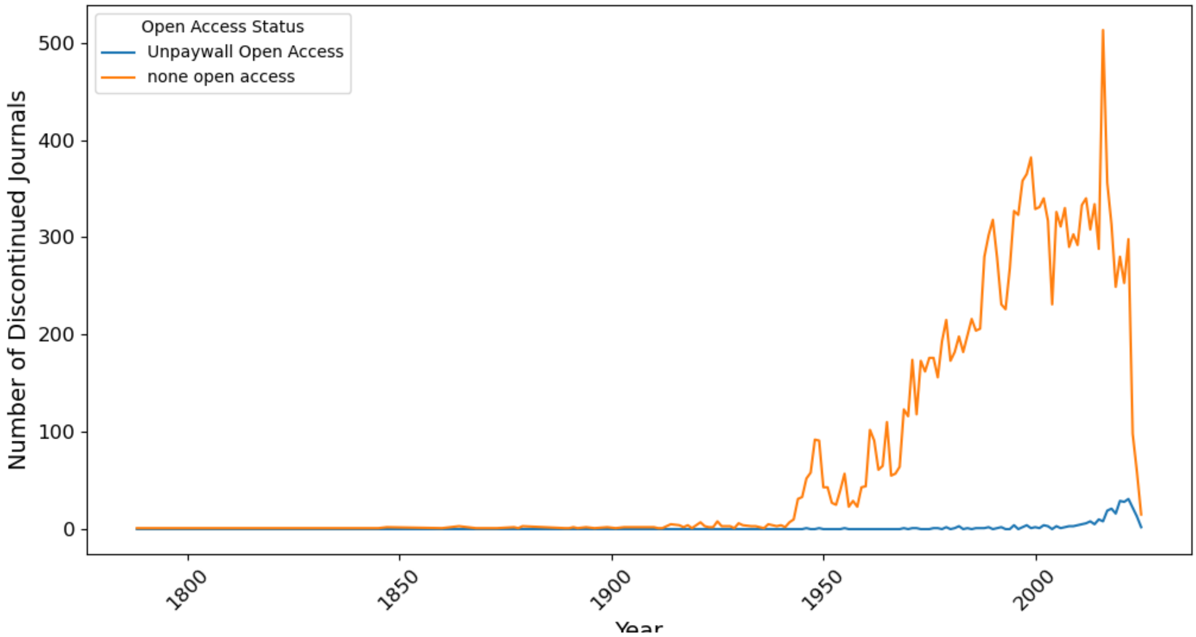

Figure 10: Open Access vs Non-Open Access Discontinued Journals Over time

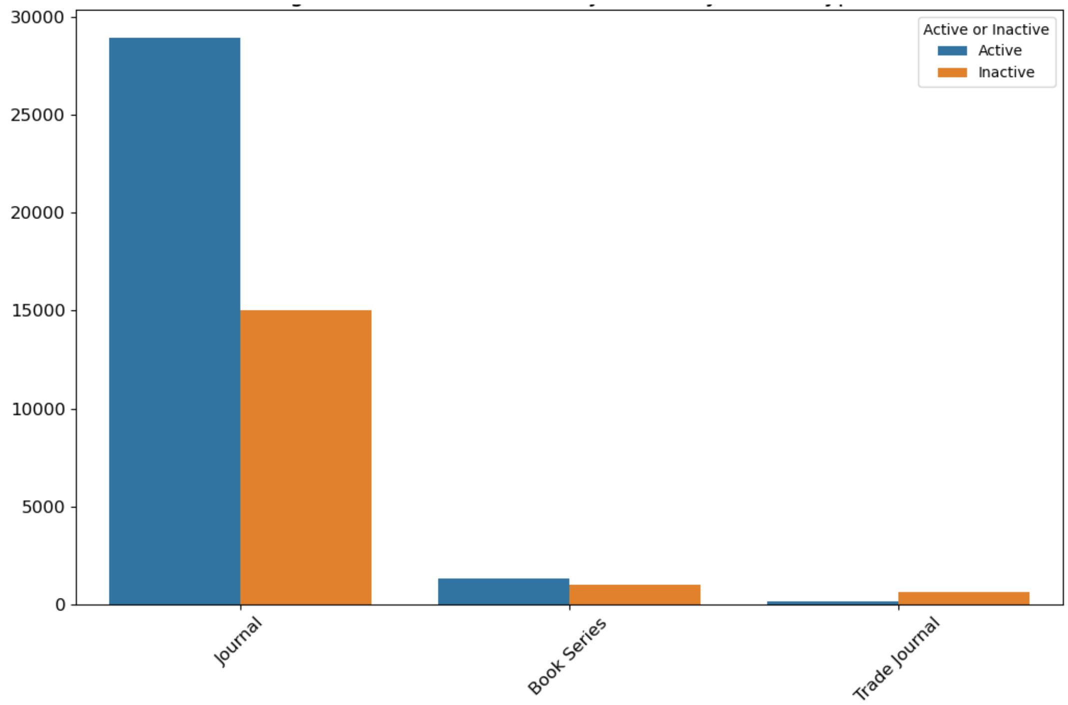

Figure 11: active and inactive distribution

Figure 11 illustrates the distribution of active and inactive journals across different source types indexed in the Scopus 2025 dataset. Journals represent the vast majority of indexed sources, with approximately twice as many active entries as inactive ones. This reflects Scopus's sustained emphasis on traditional, peer-reviewed journal content. Book series and trade journals appear far less frequently, and both categories exhibit a more balanced split between active and inactive statuses. The low volume and relatively high inactivity of trade journals suggest that this format may be less prioritized in ongoing Scopus curation. These results highlight the central role of scholarly journals in maintaining Scopus's indexing standards and the relative volatility of alternative formats.

Discussion:

The analysis of the Scopus 2025 dataset reveals several important trends and implications for the scholarly publishing ecosystem. One of the most significant findings is the continued dominance of traditional peer-reviewed journals as the primary source type, affirming Scopus's commitment to curating high-quality, structured research outputs. However, the modest but noticeable presence of conference proceedings and book series suggests an evolving recognition of alternative scholarly formats, particularly in engineering and the humanities.

The concentration of journal output among a few publishers—namely Elsevier, Springer Nature, Wiley, Taylor & Francis, and MDPI—highlights the oligopolistic nature of the academic publishing industry. This concentration raises important questions about pricing, editorial independence, and diversity of perspectives within the indexed literature. Furthermore, publishers such as MDPI and Frontiers, with their overwhelming focus on Open Access models, demonstrate a strategic positioning within the global OA movement. In contrast, other publishers appear to maintain hybrid models, reflecting divergent approaches to monetization and accessibility.

The results also demonstrate that Open Access is not a fringe publishing model; rather, it is increasingly associated with high-impact journals. The elevated proportion of OA journals in Q1 and Q2 suggests that quality and openness are no longer mutually exclusive. This finding supports the ongoing shift toward open science and is consistent with global funder mandates and institutional open access policies.

ISSN and EISSN analysis shows a significant number of journals with dual registration, indicating that publishers are actively maintaining both print and electronic editions or are ensuring legacy ISSNs are preserved alongside digital transitions. This dual-format trend is essential for archiving, indexing, and citation tracking in both traditional and emerging platforms.

Field-specific insights based on ASJC classification codes show a broad multidisciplinary coverage in Scopus, with dominance in Medicine, Engineering, and Social Sciences. These results reflect global research priorities and funding distributions. Underrepresented fields, such as Veterinary Sciences or Dentistry, may benefit from increased indexing and support to ensure comprehensive subject coverage.

Language analysis confirms the overwhelming dominance of English in scholarly communication. However, the presence of journals in French, Japanese, German, and other languages signals the enduring relevance of regional and linguistic diversity. Ensuring access to non-English research remains a key challenge for global knowledge equity.

Finally, the discontinuation trends underscore important quality control mechanisms and market dynamics. The higher discontinuation rate among Non-OA journals may reflect greater scrutiny or the declining viability of subscription-based models without strong readership or citation performance. These trends warrant continued monitoring and analysis.

Overall, this study affirms that the Scopus 2025 dataset is a rich resource for analyzing the structural and temporal dynamics of scholarly publishing. The patterns uncovered reinforce broader shifts in academic communication and provide actionable insights for librarians, researchers, publishers, and policymakers.

Conclusion:

This study provides a comprehensive, data-driven examination of the Scopus 2025 dataset, offering a panoramic view of current patterns in global scholarly publishing. Through the use of big data methodologies, we have identified and visualized trends across a range of dimensions including source types, open access status, publisher influence, field distribution, and journal continuity. The findings reveal that the scholarly publishing ecosystem continues to be shaped by a dynamic interplay of digital transformation, open science mandates, and structural shifts in publication models.

The results indicate a significant clustering of journals among a few major publishers, with distinct strategies in open access adoption. The observed dominance of OA journals in higher-quality quartiles (Q1–Q2) suggests a paradigm shift wherein openness and impact are no longer seen as mutually exclusive. Additionally, the dual-format registration of ISSN and EISSN underscores the importance of preserving both print and digital dissemination formats for broader accessibility and archival integrity.

Disciplinary mapping based on ASJC codes confirms that Medicine, Engineering, and Social Sciences continue to be the most represented fields, with other areas such as Veterinary Science and Dentistry requiring further support to enhance subject-level indexing balance. The prevalence of English as the dominant publication language reflects current global norms, though the persistence of regional languages affirms the importance of inclusive indexing practices.

Importantly, the analysis of discontinued journals sheds light on quality control processes and evolving sustainability in publishing, particularly in relation to OA and Non-OA models. These insights can inform future Scopus indexing policies, institutional journal selection criteria, and researcher publication strategies.

Future research could benefit from a longitudinal comparison across previous Scopus datasets to identify multi-year trends and predictive signals. The integration of citation metrics, altmetrics, and author-level analytics may further enrich the understanding of journal performance and scholarly influence. Machine learning techniques could be employed to predict journal discontinuation risks or to cluster journals by thematic and structural similarity.

By leveraging scalable data science tools and bibliometric methodologies, this work contributes to the growing field of research analytics and supports informed decision-making in academic publishing policy, research evaluation, and scholarly communication strategies.

References

[1]. É. Archambault, D. Campbell, Y. Gingras, and V. Larivière, "Comparing bibliometric statistics obtained from the web of science and Scopus," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 7, pp. 1320–1326, Jul. 2009, doi: 10.1002/asi.21062.

[2]. F. Franceschini, D. Maisano, and L. Mastrogiacomo, “Empirical analysis and classification of database errors in Scopus and Web of Science,” *J Informetr*, vol. 10, no. 4, pp. 933–953, Nov. 2016, doi: 10.1016/j.joi.2016.07.003.
[3]. V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, “The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis,” *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, Jun. 2021, doi: 10.1007/s11192-021-03948-5.
[4]. P. Mongeon and A. Paul-Hus, “The journal coverage of Web of Science and Scopus: a comparative analysis,” *Scientometrics*, vol. 106, no. 1, pp. 213–228, Jan. 2016, doi: 10.1007/s11192-015-1765-5.
[5]. H. H. Pham *et al.*, “A bibliometric review of research on international student mobilities in Asia with Scopus dataset between 1984 and 2019,” *Scientometrics*, vol. 126, no. 6, pp. 5201–5224, Jun. 2021, doi: 10.1007/s11192-021-03965-4.
[6]. M. Kumpulainen and M. Seppänen, “Combining Web of Science and Scopus datasets in citation-based literature study,” *Scientometrics*, vol. 127, no. 10, pp. 5613–5631, Oct. 2022, doi: 10.1007/s11192-022-04475-7.
[7]. A. Y. Shestakova, D. O. Korolev, A. A. Afanasyev, I. V. Nikiforov, and O. A. Yusupova, “Scopus publications database analysis using its API,” SPIE-Intl Soc Optical Eng, Jan. 2023, p. 14. doi: 10.1117/12.2669237.
[8]. R. Boya Marqas, S. M. Almufti, and R. Rajab Asaad, “FIREBASE EFFICIENCY IN CSV DATA EXCHANGE THROUGH PHP-BASED WEBSITES,” Academic Journal of Nawroz University, vol. 11, no. 3, pp. 410–414, Aug. 2022, doi: 10.25007/ajnu.v11n3a1480.
[9]. S. M. Almufti, “Fusion of Water Evaporation Optimization and Great Deluge: A Dynamic Approach for Benchmark Function Solving,” Fusion: Practice and Applications, vol. 13, no. 1, pp. 19–36, 2023, doi: 10.54216/FPA.130102.
[10]. S. M. Almufti, “Exploring the Impact of Big Bang-Big Crunch Algorithm Parameters on Welded Beam Design Problem Resolution,” Academic Journal of Nawroz University, vol. 12, no. 4, pp. 1–16, Sep. 2023, doi: 10.25007/ajnu.v12n4a1903.
[11]. S. M. Almufti, “Lion algorithm: Overview, modifications and applications E I N F O,” International Research Journal of Science, vol. 2, no. 2, pp. 176–186, 2022, doi: 10.5281/zenodo.6973555.
[12]. J. A. Esponda-Pérez, M. A. Mousse, S. M. Almufti, I. Haris, S. Erdanova, and R. Tsarev, “Applying Multiple Regression to Evaluate Academic Performance of Students in E-Learning,” 2024, pp. 227–235. doi: 10.1007/978-3-031-70595-3_24.
[13]. J. A. Esponda-Pérez et al., “Application of Chi-Square Test in E-learning to Assess the Association Between Variables,” 2024, pp. 274–281. doi: 10.1007/978-3-031-70595-3_28.
[14]. D. A. Majeed et al., “DATA ANALYSIS AND MACHINE LEARNING APPLICATIONS IN ENVIRONMENTAL MANAGEMENT,” Jurnal Ilmiah Ilmu Terapan Universitas Jambi, vol. 8, no. 2, pp. 398–408, Sep. 2024, doi: 10.22437/jiituj.v8i2.32769.
[15]. M. Laakso, P. Welling, H. Bukvova, L. Nyman, and B.-C. Björk, "The development of open access journal publishing from 1993 to 2009," *PLOS ONE*, 2011. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020961
[16]. E. Gadd, J. Fry, and C. Creaser, "The influence of journal publisher characteristics on open access policy trends," *Scientometrics*, 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11192-018-2716-8
[17]. E. Bravo, E. Poltronieri, M. Curti, and C. Mancini, "Open access publishing trend analysis: statistics beyond the perception," *Information Research*, 2016. [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1104363.pdf

[18]. L. Zhang, Y. Wei, Y. Huang, and G. Sivertsen, "Should open access lead to closed research?" *Scientometrics*, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s11192-022-04407-5

[19]. A. Bozkurt, S. Koseoglu, and L. Singh, "An analysis of peer-reviewed publications on openness in education," *Australas. J. Educ. Technol.*, 2019. [Online]. Available: https://ajet.org.au/index.php/AJET/article/view/4252

[20]. J. P. Tennant et al., "The academic, economic and societal impacts of Open Access," *F1000Research*, 2016. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4837983/

[21]. E. Giglia, "The impact factor of open access journals: Data and trends," *ELPUB*, 2010. [Online]. Available: https://iris.unito.it/bitstream/2318/1502923/1/Giglia%20Impact%20Factor.pdf

[22]. V. Larivière, S. Haustein, and P. Mongeon, "The oligopoly of academic publishers in the digital era," *PLOS ONE*, 2015. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502

[23]. S. Hitchcock, "The effect of open access and downloads ('hits') on citation impact," *University of Southampton*, 2004. [Online]. Available: https://eprints.soton.ac.uk/354006

[24]. S. Miguel and Z. Chinchilla-Rodriguez, "Open access and Scopus: A new approach to scientific visibility," *J. Assoc. Inf. Sci. Technol.*, 2011. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.21532

[25]. R. S. Rodrigues and E. Abadal, "Open access publishers: The new players," *PLOS ONE*, 2020. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0233432

[26]. R. P. Holley, "Open access: current overview and future prospects," *Library Trends*, 2018. [Online]. Available: https://muse.jhu.edu/pub/1/article/715060/summary

[27]. I. D. Craig, A. M. Plume, M. E. McVeigh, J. Pringle, and M. Amin, "Do open access articles have greater citation impact?" *J. Informetr.*, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1751157707000466

[28]. M. Laakso, "Green open access policies of scholarly journal publishers," *Scientometrics*, 2014. [Online]. Available: https://link.springer.com/article/10.1007/s11192-013-1205-3

[29]. P. M. Davis and W. H. Walters, "The impact of free access to the scientific literature," *J. Med. Libr. Assoc.*, 2011. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3133904/

[30]. C. Olmeda-Gómez and F. de Moya-Anegón, "Publishing trends in library and information sciences," *J. Acad. Librariansh.*, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0099133315002372

[31]. C. Figueiredo, A. A. B. Neves, and F. Pimentel, "Impact of open access policy on Brazilian science," *Anais Acad. Bras. Ciênc.*, 2024. [Online]. Available: https://www.scielo.br/j/aabc/a/HnFxTbcvP47SLYGbNkWMv7F

[32]. M. Mishra, M. K. Dash, D. Sudarsan, and C. A. G. Santos, "Assessment of open educational resources: A bibliometric analysis," *J. Acad. Librariansh.*, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0099133322000362

[33]. N. Jacobs, "Open access: key strategic, technical and economic aspects," *Google Books*, 2006. [Online]. Available: https://books.google.com/books?id=oZ6jAgAAQBAJ

[34]. M. Baquero-Arribas, L. Dorado, and I. Bernal, "Open access routes dichotomy and opportunities," *MDPI Publ.*, 2019. [Online]. Available: https://www.mdpi.com/2304-6775/7/3/49