https://doi.org/10.51408/issi2025_038

The Trends of Open Access Academic Books and Discipline Dynamics: A Cross-database Comparison Based on OpenAlex and Web of Science

Li Jiangbo[1], Niu Shihang[2], Ouyang Wenhao[3], Li Jian[4], Zhang Mingyue[5]

[1]*jiangbosky@sina.com* ,
Business College, Qingdao University, Qingdao 266100
Research Institute for Science and Technology System and Institutional Innovation of Qingdao, Qingdao 266100 (China)

[2] *15648652050@163.com,* [3] *2111903049@qq.com,*[4] *lj13181433265@163.com,*
[5] *zhangmingyue1@qdu.edu.cn*
Business College, Qingdao University, Qingdao 266100 (China)

Abstract

This study conducts a quantitative analysis of the inclusion of open access (OA) academic books in the OpenAlex and Web of Science (WoS) databases from 2004 to 2023. It explores the differences and trends between the two databases in terms of the number of OA books, annual variations, and disciplinary distribution. The study finds that OpenAlex shows a significant advantage over WoS in the scale of OA book inclusion and growth rate, with a particularly strong growth potential in recent years. Through an analysis of various academic disciplines, this study reveals the differences in disciplinary preferences and resource distribution between the two databases, further discussing the advantages and challenges of OpenAlex in promoting the inclusion of OA books. The results indicate that OpenAlex demonstrates significant potential in advancing the inclusion of OA books and global knowledge sharing. OpenAlex has gradually become an important academic resource platform, providing more convenient access to the academic community. Additionally, the study innovatively uses large language models to resolve inconsistencies in cross-database disciplinary classification, enhancing the efficiency and accuracy of data matching. In conclusion, OpenAlex's rapid growth and its advantages across multiple disciplines provide crucial support for the global inclusion and sharing of OA books, while also offering important empirical data for the optimization of academic publishing policies and resource distribution.

Introduction

Academic books are primarily focused on scholarly research, using systematic and in-depth methods to explore significant academic achievements in specific fields or topics. Compared to academic papers, academic books provide deeper theoretical analysis and a more systematic framework. They offer an overall understanding and

thorough exploration of a research area, allowing for a more comprehensive and in-depth discussion of core issues within a discipline. As such, they often exert lasting influence in the academic community (Engels, 2018). Despite the increasing speed and impact of academic journals in recent years, academic books remain central in many disciplines, particularly in the humanities and social sciences. As the core medium for knowledge accumulation and academic communication, academic books carry substantial scholarly contributions and offer broader, more detailed perspectives. They serve as the foundation for building personal academic reputation and status (Zuccala et al., 2018; Kousha et al., 2018).

Open access (OA) is central to open science (OS), serving as a key framework for enhancing research transparency, reproducibility, and collaboration through practices like open data and open communication (Harnad, 2012). OA plays a key role by removing paywalls and increasing research accessibility, but its publishing model faces challenges such as high costs and selective accessibility, which may hinder the broader goals of OS (Pulverer, 2018). Despite these limitations, OA has significantly reshaped academic publishing by broadening knowledge dissemination and creating new opportunities (Zhang, 2024). Research also highlights its impact on citation diversity, as OA publications receive a wider range of citations than closed-access works (Huang et al., 2024) and are cited more frequently, particularly in recent publications (Yang et al., 2024). Additionally, technological advancements further optimize OA's role in research dissemination—artificial intelligence and machine learning enhance data processing and scientific discovery (Barbier et al., 2022), while OA articles tend to perform better in citation and alternative metrics, such as social media mentions, with green OA showing similar advantages (Clayson et al., 2021). In sum, OA is central to OS, driving more accessible and impactful research, but realizing its full potential requires addressing cost, accessibility, and quality concerns.

In the context of open science, OpenAlex, as an open academic platform, has become an important tool for academic research and data analysis due to its support for open access data, free API, and various query methods (Velez-Estevez et al., 2023; Delgado-Quirós et al., 2024; Hazarika et al., 2024; Harder et al., 2024). Through comparative research across multiple databases, Akbaritabar et al. (2023) found that OpenAlex excels in the scope and update frequency of journal inclusion, becoming an important resource for researching the latest academic achievements. Scheidsteger et al. (2023), by comparing the metadata of OpenAlex and the Microsoft Academic Graph (MAG), pointed out that both databases show a high degree of consistency in data, with OpenAlex making improvements in specifying document types, thus enhancing its value for bibliometric analysis. Aria et al. (2024) introduced OpenAlex's R package (openalexR), providing researchers with more

efficient and convenient analytical tools. However, OpenAlex also faces some challenges. Zhang et al. (2024) noted that over 60% of journal articles lack institutional information, particularly in early literature in the social sciences and humanities. The study recommends improving data quality and reducing research bias by supplementing missing data and strengthening collaboration among platforms, publishers, and users. Overall, OpenAlex has become an important resource in academic research due to its openness and diverse data support.

At present, research on academic databases mainly focuses on academic papers, particularly in areas such as literature indexing, classification, and bibliometric analysis. In contrast, research on academic books remains scarce, especially concerning the issue of OA books. Existing studies primarily concentrate on various characteristics of academic papers, such as the accuracy of literature indexing between databases (Jiao et al., 2023), the coverage of retracted literature (Ortega, 2024), differences in academic paper classification across disciplines (Singh, 2020), and the proportion of open access literature (Basson, 2022). Therefore, despite the importance of academic books as a form of scholarly communication, their distribution and influence in databases have not received sufficient attention. Web of Science, as one of the world's leading academic literature databases, is widely used in academic research and covers a vast amount of journal and conference literature. However, research on its inclusion and analysis of academic books, particularly open access books, remains relatively limited. This study aims to fill this gap by comparing the inclusion of OA academic books in WoS and OpenAlex across different disciplines, analyzing the differences and trends in OA book inclusion and disciplinary distribution between the two databases. The goal is to provide the academic community with a new perspective on the dissemination and development of OA academic books across disciplines and offer data support and theoretical foundations for future policy development and the optimization of academic resources.

Data Acquisition and Data Processing

Data Sources and Initial Processing

This study selected the metadata of all open access academic books published between 2004 and 2023 from the OpenAlex and WoS databases as the foundational data (retrieval date: December 15, 2024). In the data processing phase, manual filtering was first applied to exclude records that lacked disciplinary classification fields or publication dates. After filtering, the valid data records in the OpenAlex database amounted to 255,810, while the WoS database contained 8,713 valid records. Given the large volume of data in OpenAlex, this study utilized OpenAlex's

cursor mechanism via its API to continuously request the OA book metadata that met the criteria, ensuring the completeness of the records. For the WoS database, due to its relatively smaller dataset, the relevant metadata was directly exported through the WoS official website.

This study extracts information on OA books based on metadata from the OpenAlex and WoS databases. OpenAlex identifies OA status using the "open_access_is_oa" field, a Boolean variable ("true" or "false") indicating whether a book is OA. In contrast, WoS assigns OA status through the "Open Access Designations" field. While both databases provide classification methods for OA books, the relatively limited number of OA books indexed in WoS may result in an insufficient sample size when further distinguishing OA types, such as gold and green OA. This limitation could affect the robustness of statistical analyses. Therefore, this study adopts a binary classification (OA vs. non-OA), focusing on the overall inclusion of OA books without differentiating specific OA models. Additionally, this approach ensures comparability across databases, thereby enhancing the reliability of the findings.

Consideration and Evaluation of DOAB as a Benchmark

At the initial stage of this study, we explored the possibility of using the Directory of Open Access Books (DOAB) as a benchmark database to evaluate the coverage of OA books in WoS and OpenAlex. DOAB, established by the OAPEN Foundation in 2012, is a non-profit platform dedicated to indexing peer-reviewed academic books published under open access models (Maginiot et al., 2019). It has played an important role in the global open access publishing ecosystem and is often used in research related to OA policy and scholarly communication.

Our original intention was to treat DOAB as a comprehensive reference collection, enabling a comparative assessment of the OA book coverage between WoS and OpenAlex. However, after data retrieval and preprocessing, we found that DOAB's coverage in certain years was significantly lower than that of OpenAlex. For instance, in 2004, DOAB recorded approximately 700 OA books, whereas OpenAlex contained over 5,000 records for the same year. This considerable gap suggests that DOAB cannot serve as a stable benchmark for cross-database comparison. While WoS had an even lower number of OA books (around 50) in the same year, its indexing scope and selection criteria differ markedly from those of DOAB, further complicating the establishment of a unified standard.

Moreover, we encountered practical limitations related to metadata structure during the matching process. Metadata exported from DOAB—either through its API or web interface—generally lacks standardized identifiers such as ISBNs, which makes precise, record-level matching unfeasible.

Metadata exported from DOAB—whether via its API or web interface—lacks ISBN information, making precise one-to-one book-level matching infeasible. In large-scale data processing scenarios, fuzzy matching based on book titles alone introduces considerable uncertainty in accuracy and requires substantial computational resources and complex algorithmic support. We experimented with several text-matching techniques on a subset of records, but the results exhibited significant inconsistencies due to variations in naming conventions, language differences, and the handling of subtitles. These issues further undermined the stability of using DOAB as a reference dataset.

In summary, while DOAB remains a valuable initiative in promoting open access books and continues to be a key player in the OA ecosystem, its current limitations in data coverage, metadata completeness, and technical interoperability prevent it from serving as a reliable benchmark for evaluating WoS and OpenAlex in this study. Nevertheless, as DOAB continues to develop and enhance its data infrastructure, it holds promising potential for future OA-related bibliometric analyses.

Disciplinary Classification Normalization Using ChatGPT

When comparing the disciplinary fields of OA academic books between WoS and OpenAlex databases, the issue of disciplinary classification consistency emerged as a critical challenge. Different databases employ distinct disciplinary classification systems, which may lead to discrepancies during cross-database comparisons (Singh et al., 2021). Therefore, ensuring data consistency and comparability, particularly in standardizing disciplinary classifications, is crucial for this study. OpenAlex is an open academic platform based on the MAG, and its disciplinary classification system closely aligns with MAG (Priem et al., 2022). MAG, an academic graph created by Microsoft, utilizes a widely applied and systematic classification framework across multiple fields, with high academic data coverage (Sinha et al., 2015). In contrast to WoS's traditional classification system, OpenAlex and MAG offer a more simplified and systematic classification approach, categorizing academic research into four major domains and 26 specific fields. This classification method not only facilitates cross-disciplinary categorization but also better accommodates the demands of big data and diverse academic resources, particularly in the inclusion and classification of OA academic resources. Compared to traditional databases, OpenAlex provides a more open, flexible, and comprehensive classification framework, effectively supporting the systematic classification and analysis of OA resources. Thus, this study adopts OpenAlex's classification system as the standard to ensure the accuracy and consistency of cross-database comparisons.

Currently, several solutions to the issue of disciplinary classification standardization have been proposed. In Gao et al. (2024), educational discipline classifications were

manually mapped to the MAG research discipline classification system, successfully linking education disciplines to research disciplines, which provided strong support for analyzing the relationship between AI education and research. Sīle et al. (2021) used cross-mapping tables to map categories from different classification systems to the OECD R&D field classification system, eliminating discrepancies between classification systems and ensuring consistency and comparability of cross-system data, thus improving the reliability of the results. Osmani et al. (2023) proposed an improved method combining recursive grouping, clustering, and classification techniques to enhance disciplinary classification consistency, especially when facing complex classification systems, providing more accurate and stable classification results. Furthermore, the ECHO project created by Wittenburg et al. (2004) effectively solved the issue of inconsistent cross-disciplinary metadata by establishing a unified ontology structure and mapping relationships, and ensuring the interoperability of metadata via XML, providing important support for cross-database data comparison.

To resolve the differences between the disciplinary classification systems of WoS and OpenAlex, this study introduces Large Language Model (LLM) technology. As one of the most important advancements in artificial intelligence, LLMs have demonstrated powerful capabilities in solving various tasks in natural language processing, particularly with the emergence of ChatGPT-4, which has had a significant impact on AI development (Zhao et al., 2023). These models' exceptional performance in natural language processing makes them a valuable tool for solving complex tasks. ChatGPT exhibits strong semantic understanding, enabling it to handle complex contextual information and multidimensional classification problems effectively. Research has shown that ChatGPT's scientific feedback generation closely aligns with human responses, demonstrating robust critical thinking abilities (Liang et al., 2024). De Winter (2023) experimentally demonstrated that ChatGPT not only outperformed human annotators in annotation tasks but also exhibited superior language proficiency and reasoning, surpassing traditional cognitive models. This opens up new possibilities for disciplinary classification mapping.

Specifically, this study employs a custom Python script to systematically submit the *WoS Categories* metadata of 8,713 OA academic books from the WoS database to OpenAI's API. Using the ChatGPT-4o model, these categories are mapped to the *Field* metadata in the OpenAlex database. We interact with the model using the following prompt: "*You are an academic book disciplinary field mapper. The following academic discipline from the Web of Science database needs to be mapped to one of the 26 major fields in the OpenAlex database. Please classify it into the most appropriate OpenAlex field from the following list: {', '.join(openalex_fields)}.*

Return only the field name without any explanation. WoS Category: {wos_category}". The *wos_category* variable contains disciplinary classification data extracted from an Excel spreadsheet, representing the WoS subject categories assigned to individual books. The *openalex_fields* variable stores OpenAlex's standardized taxonomy of 26 disciplinary fields as a list-type data structure. This process relies exclusively on the discipline names provided by WoS, eliminating potential interference from other information and thereby ensuring the reliability and consistency of the classification. To enhance the stability of the results, the temperature of the ChatGPT-4o model is set to 0, ensuring minimal output variability and further improving the accuracy of subject mapping (DE, 2024). Table 1 presents the mapping relationships for selected representative disciplinary fields.

Table 1. WoS-OpenAlex Disciplinary Field Mapping (Selected).

No	*OpenAlex Disciplinary Field*	*WoS Disciplinary Field*
1	Agricultural and Biological Sciences	Entomology
		Food Science & Technology
		Ecology
2	Arts and Humanities	History & Philosophy of Science
		Literary Theory & Criticism
		Humanities
		Linguistics
3	Earth and Planetary Sciences	Meteorology & Atmospheric Sciences
		Geology
4	Engineering	Transportation
		Mechanics
5	Environmental Science	Water Resources
		Environmental Studies
6	Health Professions	Health Care Sciences & Services
		Psychiatry
		Medical Informatics

		Surgery
7	Immunology and Microbiology	Parasitology
		Microbiology
8	Physics and Astronomy	Thermodynamics
		Physics
9	Social Sciences	Sociology
		Information Science & Library Science
		Education
		Political Science

Method

This study employs a quantitative analysis method, combining statistical techniques and data visualization to deeply analyze the OA books data in the WoS and OpenAlex databases. First, we analyzed the annual changes in the proportion of OA books in both databases. Specifically, we extracted the number of OA books for each year from 2004 to 2023 for each database and calculated their proportion within the total number of acdemic books. The calculation formula is as follows:

$$P_t = \frac{OA_t}{T_t} \times 100\%$$

Where P_t represents the proportion of OA books in year *t*, OA_t is the number of OA books in year *t* in the database, and T_t is the total number of books in year *t* in the database. The variation in the proportion of OA books each year is presented in line graphs, allowing for an intuitive analysis of the OA book inclusion trends in these two databases. This will help us understand the development dynamics of OA books in both databases and assess which database is growing faster, thus projecting its future potential in OA book inclusion.

Secondly, we calculated the proportion of OA books in each disciplinary field for both the WoS and OpenAlex databases. This analysis examines the preference of each database in the inclusion of OA books in various disciplines. The formula used is as follows:

$$F_i = \frac{OA_i}{OA_{\text{Total}}} \times 100\%$$

Where F_i represents the proportion of OA books in the *i*-th disciplinary field relative to the total number of OA books in that database, OA_i is the number of OA books in

the i-th field, and OA_{Total} is the total number of OA books in the database. A higher proportion indicates a stronger inclination of the database to include OA books in that field, while a lower proportion indicates more limited inclusion of books in that field. Finally, we calculated the relative proportion of each disciplinary field, as follows:

$$D_i = \frac{X_i - Y_i}{X_i + Y_i}$$

Where D_i represents the relative proportion of OA books in the i-th discipline across the two databases, X_i is the proportion of OA books in the i-th field in the OpenAlex database, and Y_i represents the proportion of OA books in the same field in the WoS database. To visualize the results more intuitively, a heatmap was used. The color intensity in the heatmap reflects the relative proportion differences, allowing us to clearly identify the dominant fields in each database. By calculating the proportion of OA books in each disciplinary field and the differences in the relative proportions between fields, this study reveals the disparities in the inclusion of OA books across different disciplines between WoS and OpenAlex. This analysis not only provides data support for database optimization and resource allocation but also helps researchers in various fields select the most appropriate database for academic retrieval to improve research efficiency.

Results

Over the past two decades, the total proportion of OA books in the WoS database has been 6.01%, while the proportion in the OpenAlex database is 9.46%. Although this difference might not seem substantial in terms of overall proportion, considering the actual number of OA academic books in each database (OpenAlex includes 255,810 books, and WoS includes 8,713 books), it is clear that OpenAlex has significantly outpaced WoS in the volume of OA books. This indicates that OpenAlex has a broader scale and coverage in OA academic book inclusion. In other words, while the overall proportion difference is not highly significant, OpenAlex exhibits a stronger growth potential in the inclusion of OA books.

Next, this study further analyzes the annual changes in the proportion of OA books in both databases from 2004 to 2023. Figure 1 presents the line graph that visually illustrates the trend based on the proportion values for each year. It is evident that the proportion of OA books in the OpenAlex database has experienced significant growth over the past two decades. In 2004, the proportion of OA books in OpenAlex was 4.93%, and by 2023, it had dramatically increased to 25.10%, with a particularly sharp acceleration in growth after 2020. In contrast, while the proportion of OA books in the WoS database also shows an upward trend, the increase is relatively

slow. In 2004, the proportion of OA books in WoS was 6.00%, and by 2023, it had only increased to 11.32%. Notably, between 2004 and 2010, the proportion of OA books in WoS fluctuated significantly, and it consistently remained lower than OpenAlex. Although there has been some recovery in recent years, the growth rate remains considerably slower than OpenAlex.

Moreover, from the chart, it is evident that OpenAlex saw a significant increase in OA book proportion, from a low of 3.72% in 2011 to 25.10% in 2023, representing a growth of 576.34%. In contrast, WoS's proportion of OA books reached its lowest point of 3.74% in 2008 and only increased to 11.32% by 2023, a growth of 202.68%. This disparity in growth rates highlights the significant breakthrough OpenAlex has made in the inclusion of OA academic books.

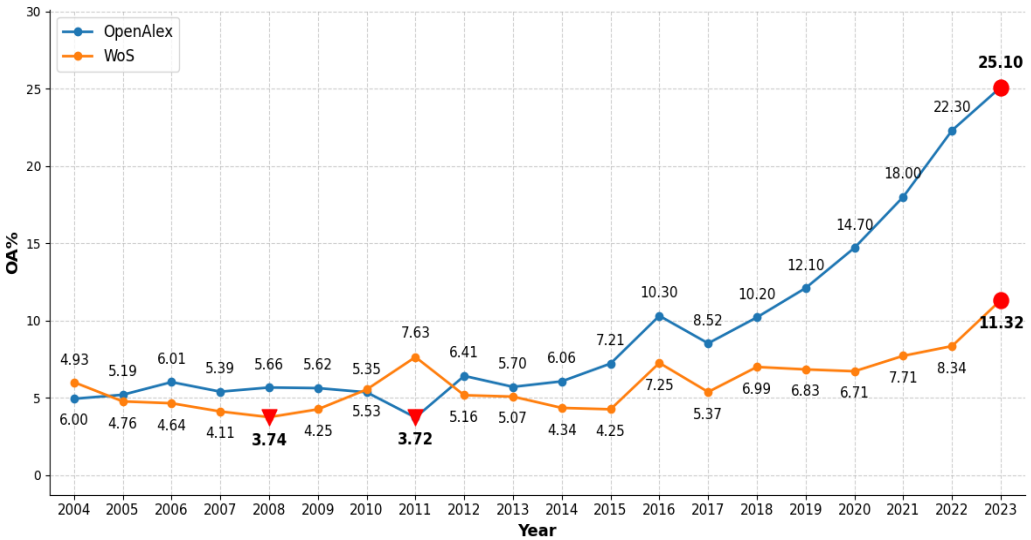

Figure 1. Trend of OA Book Proportions in WoS and OpenAlex (2004-2023).

Overall, OpenAlex's rate of growth in OA book inclusion is notably higher than WoS, demonstrating a stronger expansion potential. Particularly after 2018, OpenAlex's proportion of OA books quickly surpassed WoS and maintained a significant lead by 2023. This trend suggests that OpenAlex is likely to continue expanding its market share in the field of open access books in the coming years, while WoS may face the need to optimize its OA book inclusion strategy in response to the rapidly accelerating global open access trend.

Next, this study provides a detailed analysis of the 26 disciplinary fields, further subdivided based on the four major domains in the OpenAlex database (Life Sciences, Physical Sciences, Health Sciences, and Social Sciences). For each field, we calculated the proportion of OA books within that specific discipline to more accurately assess the inclusion of OA books in different fields. By independently

analyzing each field's sub-disciplines, this approach allows us to provide a more nuanced understanding of the inclusion of OA books within each respective domain.

In the Social Sciences domain, there are significant differences in the proportion of OA books between WoS and OpenAlex in several disciplines (Fig. 2A). OpenAlex shows higher proportions of OA books than WoS in the fields of *Social Sciences*, *Economics, Econometrics and Finance*, *Decision Sciences* and *Psychology*, with the largest difference in the *Social Sciences* field, where OpenAlex holds 53.61%, significantly higher than WoS's 35.20%. However, in the *Arts and Humanities* field, OpenAlex's proportion is 19.12%, much lower than WoS's 43.43%, indicating that WoS places greater emphasis on supporting and developing the humanities within the Social Sciences domain.

In the Physical Sciences domain, both WoS and OpenAlex exhibit significant advantages in different disciplines for the inclusion of OA books (Fig. 2B). In the fields of *Computer Science* and *Environmental Science*, OpenAlex's inclusion rate is significantly higher than WoS, likely reflecting its greater support for emerging disciplines. However, in the fields of *Mathematics*, *Chemistry*, *Materials Science* and *Physics and Astronomy*, WoS has a higher proportion of OA books, indicating that WoS has a stronger coverage in these traditional and foundational scientific fields, likely due to its longstanding authority and influence in scientific research. Additionally, the inclusion in other disciplines is relatively close between the two databases. For example, in the *Engineering* field, WoS's proportion is 27.76%, slightly lower than OpenAlex's 28.70%.

In the Life Sciences domain, WoS leads in the *Biochemistry, Genetics and Molecular Biology* field, with a proportion of 56.38%, significantly higher than OpenAlex's 31.25% (Fig. 2C). In contrast, OpenAlex shows stronger inclusion in the fields of *Agricultural and Biological Sciences* and *Neuroscience*, particularly in *Agricultural and Biological Sciences*, where OpenAlex's proportion is 54.73%, much higher than WoS's 35.90%. This difference indicates that OpenAlex has a stronger capability in the inclusion of OA books in emerging fields, and may continue to expand its inclusion in the Life Sciences domain. Furthermore, both databases show relatively weak inclusion in the *Veterinary* field, suggesting a need for further support and resource integration in this area.

Finally, in the Health Sciences domain, WoS has a significant lead in the *Health Professions* field, with a proportion of 84.73%, far exceeding OpenAlex's 38.53% (Fig. 2D). This indicates that WoS has stronger inclusion capabilities for OA books in health-related disciplines. However, in the *Medicine* field, OpenAlex has a proportion of 56.54%, while WoS is only at 9.36%. This disparity suggests that OpenAlex has a stronger capability in the inclusion of OA books in the medical sciences. In certain specialized subfields, such as *Pharmacology, Toxicology and*

Pharmaceutics and *Dentistry*, WoS shows higher OA book proportions of 3.20% and 2.22%, respectively, compared to OpenAlex's 0.99% and 0.93%. This indicates that WoS has stronger growth potential in specialized fields like pharmacology and dentistry.

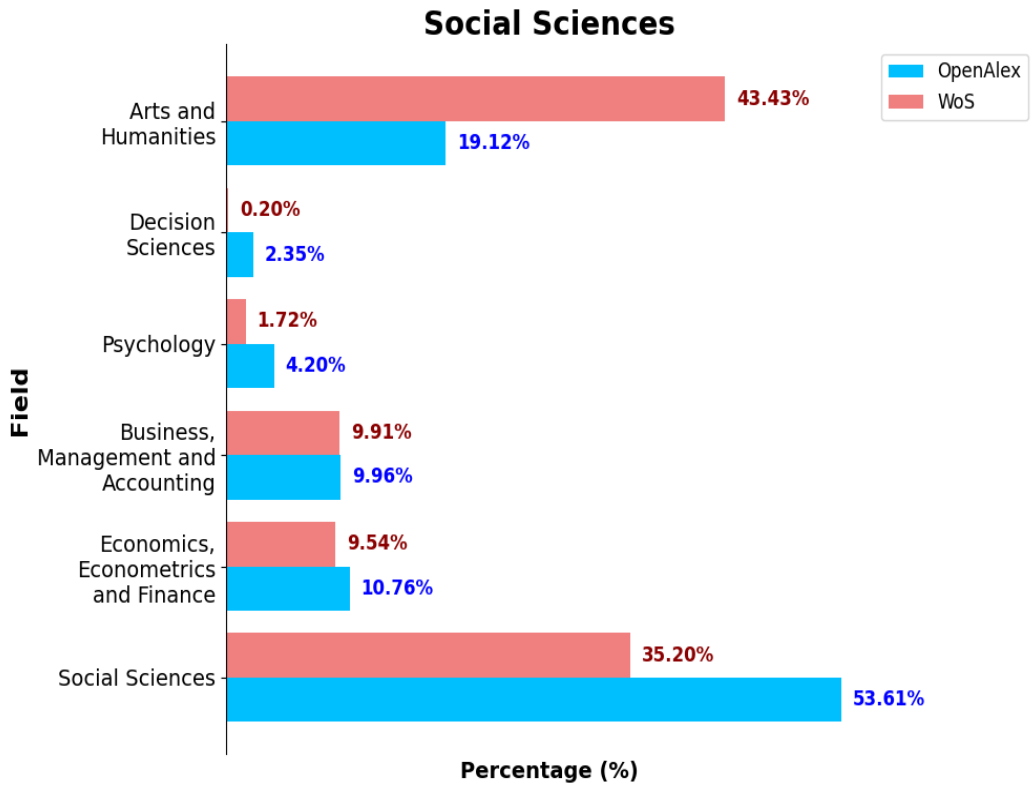

Figure 2A. Proportion of OA Books in Social Sciences by WoS and OpenAlex.

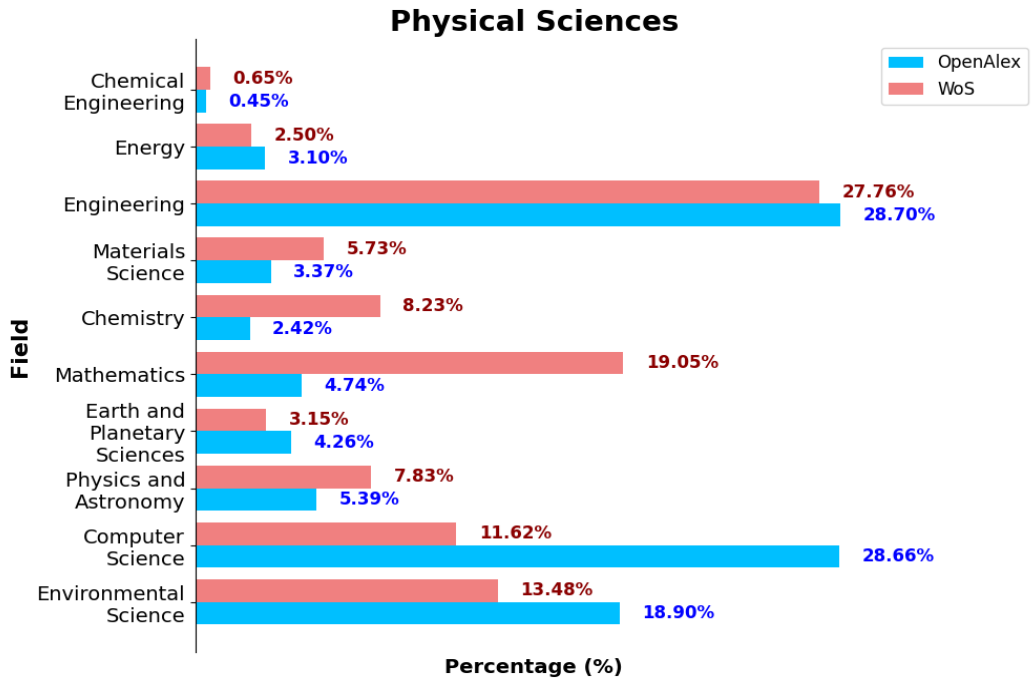

Figure 2B. Proportion of OA Books in Physical Sciences by WoS and OpenAlex.

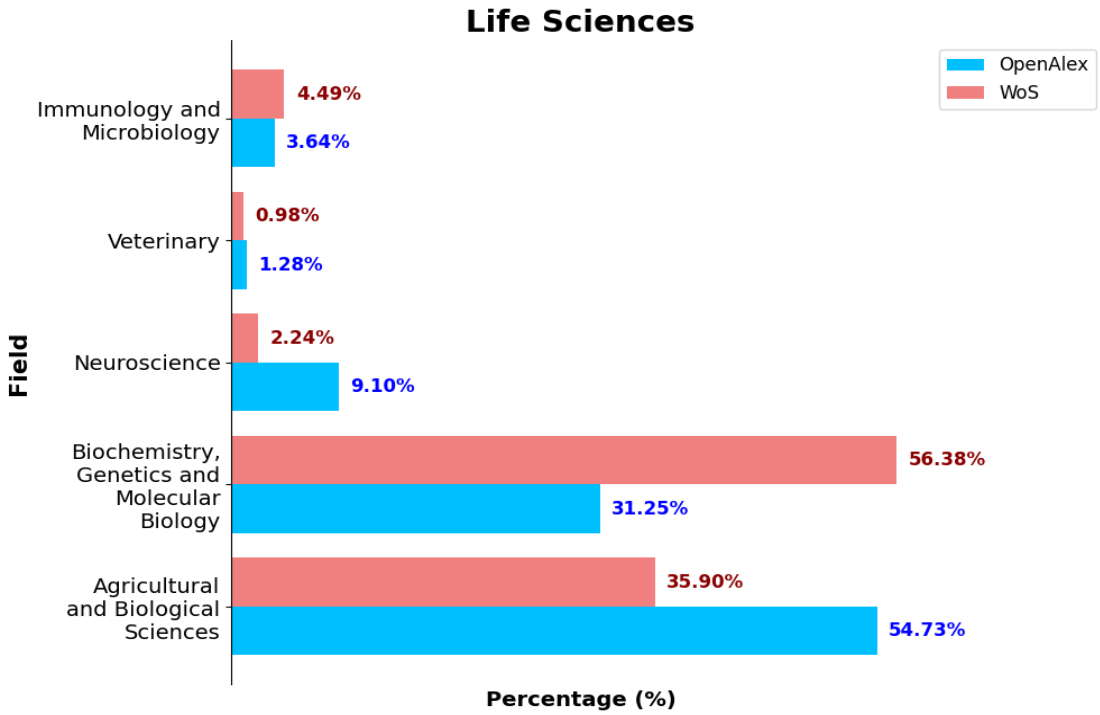

Figure 2C. Proportion of OA Books in Life Sciences by WoS and OpenAlex.

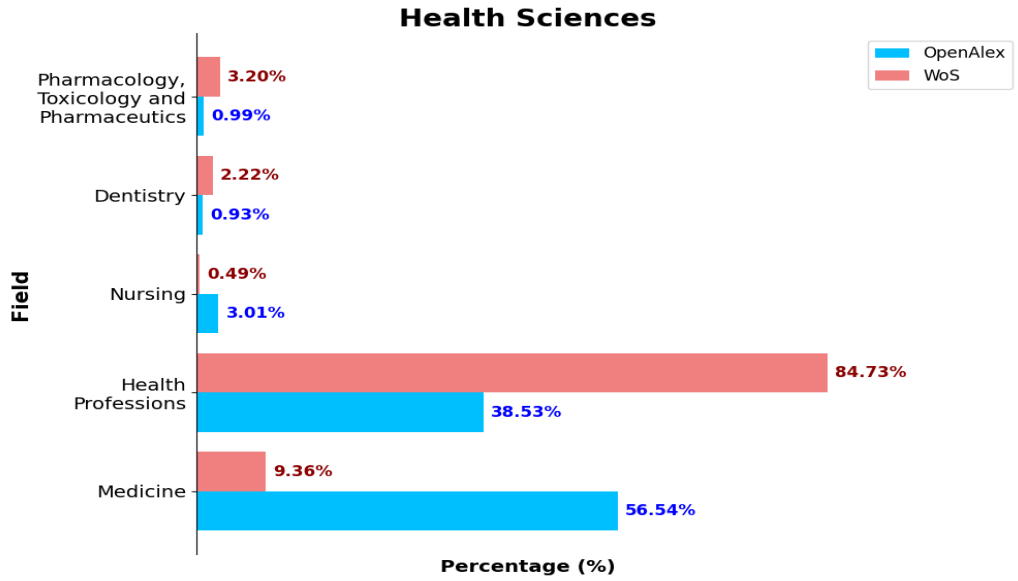

Figure 2D. Proportion of OA Books in Health Sciences by WoS and OpenAlex.

To visually represent the differences in the inclusion of OA books across disciplines in WoS and OpenAlex, this study calculated the relative proportions for each field. Figure 3 presents a heatmap of the relative proportions, providing a more intuitive visual representation of the databases' inclusion tendencies across different disciplines. In the heatmap, colors closer to red indicate a relative proportion near 1, suggesting that OpenAlex has a higher proportion of OA books in that field. Conversely, colors closer to blue represent a relative proportion near -1, indicating that WoS leads in that field, while shades of grey represent fields where the databases' inclusion is more similar. From the heatmap, it is apparent that a larger number of disciplines show a tendency towards red, indicating that OpenAlex includes more OA books in more fields.

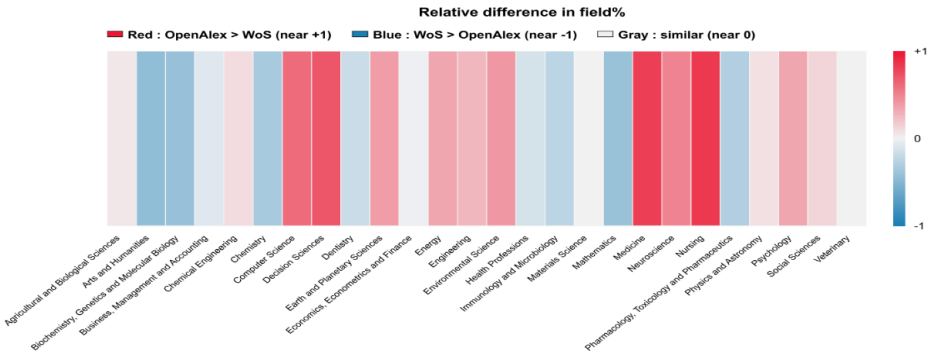

Figure 3. Heatmap of Relative Proportions of OA Book Inclusion in WoS and OpenAlex Across 26 Disciplinary Fields.

Discussion

This study employs quantitative analysis to examine the inclusion of OA academic books in the WoS and OpenAlex databases from 2004 to 2023, with a focus on comparing the number and distribution of OA books across different disciplines in both databases. The findings indicate that OpenAlex has a significantly higher number and growth rate of OA books than WoS, particularly in the past five years, during which the proportion of OA books in OpenAlex has increased substantially. This finding suggests that OpenAlex has the capacity to significantly contribute to the advancement of open-access book publishing.

The accuracy of discipline classification standardisation is a fundamental prerequisite for data analysis in this study. As WoS and OpenAlex employ different classification systems, their comparability and consistency are directly affected. To address these discrepancies, this study employs ChatGPT to match disciplinary classifications across the two databases, thereby ensuring a standardized classification system. Leveraging its advanced semantic understanding capabilities, GPT efficiently and accurately resolves inconsistencies between the classification systems of the two databases. Additionally, GPT serves as an auxiliary tool in this study, maintaining high accuracy while reducing human error and providing fast and reliable classification mapping. In comparison with conventional manual classification methodologies, GPT, as a large language model, not only enhances classification efficiency but also handles more complex and interdisciplinary

classification tasks, thereby significantly expanding the boundaries of academic data processing. The integration of GPT thus presents a novel approach to disciplinary classification and demonstrates the potential of artificial intelligence in academic research.

Moreover, this study conducts an in-depth analysis of OA books across a range of academic disciplines. The results reveal substantial differences in the proportion of OA books across disciplines in the two databases. A more pronounced advantage in the inclusion of OA books is exhibited by OpenAlex in fields such as *Computer Science* and *Environmental Science*, whereas a higher share of OA books in disciplines such as *Biochemistry, Genetics and Molecular Biology* and *Physics and Astronomy* is exhibited by WoS. This analysis provides valuable insights into the inclusion preferences of the two databases across various disciplines.

OpenAlex is a vital element of the open science ecosystem, offering a more open and sustainable model for academic resource sharing through its freely accessible API and extensive data coverage. In comparison to conventional subscription-based databases, OpenAlex boasts substantial advantages in terms of accessibility, openness, and interoperability. These qualities contribute to the reduction of inequalities in access to academic resources and the enhancement of global research collaboration. In light of the growing emphasis on open science policies, the development of OpenAlex is of paramount importance in promoting equity and transparency in academic publishing. Nevertheless, despite its strong potential in the inclusion of OA books, OpenAlex still faces certain limitations concerning metadata quality. For instance, the absence of institutional affiliation information for some books may affect the accuracy of author attribution and research impact analysis. In order to enhance its value in academic research, it is recommended that OpenAlex continue to improve metadata quality and enhance its integration with other open science tools.

This study provides a systematic analysis of the differences in OA book inclusion between WoS and OpenAlex, as well as empirical evidence for the formulation of open-access policies and academic resource management strategies. As open science continues to evolve, data interoperability and accessibility will become pivotal issues in global scholarly communication. The findings of this study highlight the role of OpenAlex in promoting the dissemination of OA books and offer valuable insights for the optimisation of open science infrastructure and the evolution of academic publishing models.

References

Akbaritabar, A., Theile, T., & Zagheni, E. (2023). *Global flows and rates of international migration of scholars* (No. WP-2023-018). Max Planck Institute for Demographic Research. https://doi.org/10.4054/MPIDR-WP-2023-018

Aria, M., Le, T., Cuccurullo, C., Belfiore, A., & Choe, J. (2024). openalexR: An R-tool for collecting bibliometric data from OpenAlex. *R Journal, 15*(4), 167-180.

Barbier, L. M., Green, J. L., & Draper, D. S. (2022). The need for open access and natural language processing. *Proceedings of the National Academy of Sciences, 119*(15), e2200752119.

Basson, I., Simard, M. A., Ouangré, Z. A., Sugimoto, C. R., & Larivière, V. (2022). The effect of data sources on the measurement of open access: A comparison of Dimensions and the Web of Science. *PLOS One, 17*(3), e0265545.

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2021). The open access advantage for studies of human electrophysiology: Impact on citations and altmetrics. *International Journal of Psychophysiology, 164*, 103-111.

de Winter, J. (2024). Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts. *Scientometrics, 1-19*.

de Winter, J. C. (2023). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 1-16.

Delgado-Quirós, L., & Ortega, J. L. (2024). Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies, 5*(1), 31-49.

Engels, T. C., Istenič Starčič, A., Kulczycki, E., Pölönen, J., & Sivertsen, G. (2018). Are book publications disappearing from scholarly communication in the social sciences and humanities? *Aslib Journal of Information Management, 70*(6), 592-607.

Gao, J., & Wang, D. (2024). Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nature Human Behaviour, 1-12*.

Harder, R. (2024). Using Scopus and OpenAlex APIs to retrieve bibliographic data for evidence synthesis: A procedure based on Bash and SQL. *MethodsX, 102601*.

Harnad, S. (2012). Open access: A green light for archiving. *Nature, 487*(7407), 302-303.

Hazarika, R., Roy, A., & Sudhier, K. G. (2024). Mapping the open access publications of Indian non-profit organizations over the last 20 years based on OpenAlex insights. *Global Knowledge, Memory and Communication*. https://doi.org/10.1108/GKMC-02-2024-0106

Huang, C. K., Neylon, C., Montgomery, L., Hosking, R., Diprose, J. P., Handcock, R. N., & Wilson, K. (2024). Open access research outputs receive more diverse citations. *Scientometrics, 129*(2), 825-845.

Jiao, C., Li, K., & Fang, Z. (2023). How are exclusively data journals indexed in major scholarly databases? An examination of four databases. *Scientific Data, 10*(1), 737.

Kousha, K., & Thelwall, M. (2018). Can Microsoft Academic help to assess the citation impact of academic books? *Journal of Informetrics, 12*(3), 972-984.

Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., ... & Zou, J. (2024). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI, 1*(8), AIoa2400196.

Maginiot, F., Mounier, P., & Pellen, M. (2019). *DOAB Foundation: Toward a quality label for academic books in open access*. CNRS. https://www.cnrs.fr/en/press/doab-foundation-toward-quality-label-academic-books-open-access

Ortega, J. L., & Delgado-Quirós, L. (2024). The indexation of retracted literature in seven principal scholarly databases: A coverage comparison of Dimensions, OpenAlex, PubMed, Scilit, Scopus, The Lens, and Web of Science. *Scientometrics, 129*(7), 3769-3785.

Osmani, A., Hamidi, M., & Alizadeh, P. (2022). Clustering approach to solve hierarchical classification problem complexity. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 7, pp. 7904-7912).

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Scheidsteger, T., & Haunschild, R. (2023). Which of the metadata with relevance for bibliometrics are the same and which are different when switching from Microsoft Academic Graph to OpenAlex? *Profesional de la Información, 32*(2). https://doi.org/10.3145/epi.2023.mar.09

Sīle, L., Guns, R., Vandermoere, F., Sivertsen, G., & Engels, T. C. (2021). Tracing the context in disciplinary classifications: A bibliometric pairwise comparison of five classifications of journals in the social sciences and humanities. *Quantitative Science Studies, 2*(1), 65-88.

Singh, P., Piryani, R., Singh, V. K., & Pinto, D. (2020). Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. *Journal of Intelligent & Fuzzy Systems, 39*(2), 2471-2476.

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus, and Dimensions: A comparative analysis. *Scientometrics, 126*, 5113-5142.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B. J., & Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 243-246).

Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Muñoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management, 60*(4), 103385.

Wittenburg, P., Gulrajani, G., Broeder, D., & Uneson, M. (2004). Cross-disciplinary integration of metadata descriptions. In *LREC 2004*.

Yang, P., Shoaib, A., West, R., & Colavizza, G. (2024). Open access improves the dissemination of science: Insights from Wikipedia. *Scientometrics, 129*(11), 7083-7106.

Zhang, L., Cao, Z., Shang, Y., Sivertsen, G., & Huang, Y. (2024). Missing institutions in OpenAlex: Possible reasons, implications, and solutions. *Scientometrics, 1-23*.

Zhang, X. (2024). Is open access disrupting the journal business? A perspective from comparing full adopters, partial adopters, and non-adopters. *Journal of Scholarly Publishing, 55*(2), 145-162.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Zuccala, A. A., Giménez-Toledo, E., & Peruginelli, G. (2018). Scholarly books and their evaluation context in the social sciences and humanities. *Aslib Journal of Information Management, 70*(6), 586-591.