科学信息“三巨头”Web of Science、Scopus 与 OpenAlex 的比较文献计量综述

Daniel Torres-Salinas & Wenceslao Arroyo-Machado

EC3metrics，格拉纳达大学，西班牙

摘要

本比较研究考察三大主要多学科书目数据库*Web of Science* 核心合集、*Scopus*和*OpenAlex*，旨在提供关于覆盖范围、元数据质量与功能特征的最新证据，以辅助科研评价中的战略决策。本报告由两个互补的方法学部分构成。首先，报告对近期学术文献进行系统综述，考察记录数量、开放获取覆盖率、语言多样性、参考文献覆盖率与元数据质量；随后开展针对2015—2024年期间的原创文献计量分析，探讨纵向分布、文献类型、主题概况、语言差异以及各数据库之间的重叠。全文最后给出十点执行摘要和五项建议。

关键词

文献综述、文献计量分析、科学信息、书目数据库、Web of Science、Scopus、OpenAlex、覆盖范围

作者贡献

两位作者对本工作的各个方面贡献均等，包括概念构思、方法论、分析和撰写。

如何引用本文

Torres-Salinas, D., & Arroyo-Machado, W.（2026年）。科学信息“三巨头”：Web of Science、Scopus 和 OpenAlex 的比较文献计量综述。格拉纳达：InfluScience Editions。https://doi.org/10.5281/zenodo.18411229

版本：1.2。英语 / 日期：2026年1月22日

报告编号：PLUGG01EN26

数字对象标识符: 10.5281/zenodo.18411229

许可证: 知识共享署名-非商业性使用-禁止演绎

由 InfluScience Editions 于2026年在格拉纳达出版

西班牙语版本: 10.5281/zenodo.18336509

目录

引言.2
第一部分.4
文献综述.4
1.1. 总体与商业概述.4
1.1.1. Web of Science 核心合集 - 科睿唯安.4
1.1.2. Scopus - 爱思唯尔.6
1.1.3. OpenAlex – OurResearch.7
1.2 精选与近期文献.9
1.2.1 记录总量.9
1.2.2 开放获取期刊的覆盖范围.10
1.2.3 语言覆盖率.11
1.2.4 参考文献覆盖率.11
1.2.5. 元数据质量与可靠性.12
1.3. 比较总结表.15
第2部分.17
定量分析.17
2.1 纵向分析.17
2.2. 按文献类型划分的差异.18
2.3. 主题概况.20
2.4. 语言差异.22
2.5. 研究领域详情.23
2.6. 来源的重叠与趋同.24
第3部分.26
3.1. 执行摘要.26
3.2. 最终建议.28
附录.29
参考文献 s. 30

引言

过去几年间，科学信息的格局经历了前所未有的转变，这种转变远不止于引入新工具，更代表着科学知识在生产、传播、获取和评估方式上的结构性变化。十多年前，Web of Science 和 Scopus 在书目信息的检索与分析领域几乎占据绝对主导地位；而如今，信息来源日趋多元，提供了不同视角、更广泛的覆盖范围，并最终带来了检索和分析科学活动的新方式（Torres-Salinas 等，2023年）。尤其是过去十年间，开放数据库的兴起已成为一种显著趋势，对传统的商业订阅模式构成了强有力的挑战。

这一变革进程若以某个具体里程碑为起点，可追溯至谷歌学术（2004年）。它通过大规模索引来自不同来源的学术出版物，扩大了获取科学文献的渠道，使之“民主化”[1]。此后，各类开放式机构知识库持续发展，其中 OpenDOAR[2], 等倡议促进了作者自行存档。近来，OurResearch 于 2022年推出的 OpenAlex 作为 Microsoft Academic Graph 的继任者，通过整合 Crossref、PubMed、ORCID 和机构知识库等异质来源，将自身定位为一种可行的替代方案。与此同时，Dialnet 和 SciELO 等区域数据库已发展成为科学信息基础设施，为长期处于边缘地位的人文学科和社会科学领域的伊比利亚美洲科学产出提供了可见度。此外，Altmetric.com 和 PlumX 等替代计量指标（替代计量学）平台也通过捕捉传统引文之外的科研社会影响力，进一步完善了这一生态系统。

这一变化也与科研评价转型密切相关。该转型始于2012年发布的《旧金山科研评估宣言》（DORA）[3]以及《莱顿宣言》（Hicks 等，2015年），这些文件质疑将影响因子作为科学质量替代指标的做法。由欧洲大学协会和欧盟委员会于2022年1月推动的科研评估促进联盟（CoARA）[4], ，标志着这场改革运动实现了制度化。在西班牙，大多数研究机构已承诺遵循这些原则，这些原则如今已广为人知；ANECA、FECYT 和国家研究署（AEI）等组织也作出了同样承诺。因此，新的开放科学信息基础设施与评估体系改革相互融合，形成了这样一种格局：获取渠道的民主化与对学术职业生涯更具情境针对性的评价相辅相成（Torres-Salinas, 2024年）。这正是我们所处的背景，也是本文立足的语境。

因此，问题不在于该与哪个数据库签约，或该使用哪种科学信息服务，而在于公共机构何时会迈出最后一步，在没有安全网的情况下转向开放的出版与传播体系。事实已充分证明，只要具备政治意愿，就能够推动科学共同体转向开放获取的世界（Arroyo-Machado & Torres-Salinas，2025年），尤其摆脱文章处理费（APCs）的惊人增长（Haustein 等，2024年）。这一愿景进一步得到了

[1] 我们有意以含义模糊的方式使用“民主化的”一词。谷歌学术并非科学开放获取的开创者；相反，它借助了期刊、知识库和平台此前利用公共资源开展的工作。它的贡献在于通过私人基础设施，让这项集体努力得以显现。[2] https://opendoar.ac.uk/ [3] https://sfdora.org/read/read-the-declaration-espanol/ [4] https://www.coara.org/

《巴塞罗那开放研究信息宣言》（2024年）[5]承诺，享有盛誉的学术组织和资助机构将把研究信息开放确立为新规范，并承认科学信息获取和评价不应建立在由商业基础设施管理的封闭证据之上。

尽管长远未来由我们的组织正在签署的机构政策以及新兴的科学信息开放世界所塑造，但短期和中期形势表明，我们必须继续沿着惯常的路径前行。就此而言，商业数据库仍是一种站得住脚的解决方案，因为它们对图书馆、研究群体和研发管理者仍具有不可否认的价值。如今比以往任何时候都更有必要以可靠数据和经核实的来源支撑我们的决策并维护我们的学术履历，而开放来源并不能始终保证这一点。因此，为阐明这一格局，我们撰写了本篇对当前主要多学科书目数据库——当下的“三大数据库”：Web of Science、Scopus 和 OpenAlex——所做的基础层面比较分析。

因此，我们可以指出，本研究的主要目标是对这三个数据库的覆盖范围、质量和功能特征进行比较。为此，我们将本报告分为两个明显不同的部分：

第一部分：文献综述
本节将简要回顾相关文献。首先，我们以技术概况表的形式，向读者介绍各数据库的主要特点。随后，我们将分析发表于这三个数据库中的相关近期学术研究成果。本节无意穷尽所有研究，而是旨在呈现与本报告背景相关的信息。最后，我们通过汇总表归纳上述各方面的内容。

第2部分：比较文献计量研究
为补充前文综述，我们基于 Web of Science、Scopus 和 OpenAlex 的原创未发表数据开展比较文献计量研究，使用过去十年（2015—2024年）的书目数据，并重点深入分析最近的五年期（2020—2024年）。这项定量分析主要包括纵向研究、不同文献类型的差异、主题概况以及语言差异。

本报告结构清晰：首先介绍数据库，回顾现有文献，再进行定量分析，以便统一并核实前文提供的数据，最后给出结论。对于希望直接了解主要结论的读者，我们在文末附上了执行摘要。此外，本报告还附有方法附录，详细说明了所进行的定量分析。我们希望能为这一主题提供有益的见解。



作者，格拉纳达，2026年1月22日

[5] https://barcelona-declaration.org/signatories/

第一部分

文献综述

1.1. 总体与商业概述

在第一部分，我们将分析本报告所考察的三个数据库的主要特征。我们使用的信息来自专业网站和商业网站，旨在借此提供初步概览，尤其面向对数据库了解不多或不太熟悉的读者。对于这三个数据库，我们都将分析以下基本信息：(1) 商业模式，(2) 数量覆盖范围，(3) 编辑标准，以及 (4) 分析功能。

1.1.1. Web of Science 核心合集 - 科睿唯安

商业模式

Web of Science 核心合集采用基于机构订阅和访问许可的商业模式，由科睿唯安[6]管理。其策略着重提供不同的订阅层级，并根据[7] 学术机构的需求量身定制，主要面向需要可靠平台来评价科学活动的大学、研究中心和公共组织。科睿唯安通过长期订阅协议维持经常性收入模式，并辅以分析服务收入。

数量覆盖范围

Web of Science 核心合集收录 22,704 种期刊[8]。科学引文索引扩展版（SCIE）涵盖 182 个科学学科的 9,450 多种期刊，共计 6,900 万条记录[9]；社会科学引文索引（SSCI）涵盖 58 个学科的 3,541 种期刊，共有 1,180 万条记录[10]；艺术与人文引文索引（AHCI）收录 25 个类别的 1,808 种期刊，共计 560 万条记录[11]。部分学科的历史数据可追溯至 1900 年，其他学科则可追溯至 1945 年。另一方面，新兴来源引文索引（ESCI）包含 9,504 种期刊、590 万条记录，涵盖 252 个类别，其索引始于 2005[12]。核心合集还包括图书和会议论文集等其他内容，记录总数达 9,700 万条[13]。

最近，在2025年9月，Research Commons[14] 产品已被纳入该平台，并独立于核心合集，新增了过去10年在期刊上发表的逾3,200万篇文献。然而，这实际上不过是对OpenAlex的整合——我们稍后将对此展开讨论——也体现了该公司的双重性：一方面，

[6] https://clarivate.com/academia-government [7] https://developer.clarivate.com/apis/ [8] https://clarivate.libguides.com/librarianresources/coverage [9] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/science-citation-index-expanded/ [10] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/social-sciences-citation-index/ [11] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/arts-humanities-citation-index/ [12] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/emerging-sources-citation-index/ [13] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/ [14] https://clarivate.com/academia-government/blog/expanding-the-web-of-science-platform-with-research-commons/

它在核心合集内维持着经过筛选的均衡增长策略，同时通过纳入互补数据库来扩大覆盖范围；另一方面，我们早已可以通过其他途径免费访问这些数据库。

编辑标准

Web of Science 核心合集的内容遴选由科睿唯安编辑负责，他们独立开展工作，不隶属于任何出版机构或研究机构，从而避免潜在的利益冲突[15]。期刊评估基于 28 项明确定义的标准：24 项质量标准，用于评估编辑严谨性和良好的出版实践；以及 4 项影响力标准，旨在以引文作为主要指标，识别最具影响力的期刊[16]。这或许是该数据库的鲜明特点之一：入选索引的难度，而这一标准近年来有所放宽。该数据库纳入了新兴来源引文索引（ESCI），这是一个较为“宽松”的索引，允许收录相关性较低、尚未达到进入主要索引（科学引文索引扩展版、社会科学引文索引、艺术与人文引文索引）所要求的影响力水平的期刊。

值得注意的是，当前的评估流程包括持续监测期刊表现，并可能因不端行为或引文操纵而将期刊剔除。2024年6月的更新中，17 种期刊因涉嫌引文操纵和编辑“不端行为”而被移出期刊引证报告（JCR）和 Web of Science 核心合集（Attride & Orrall, 2024）。这些措施延续了2023年确立的趋势：当年有 50 多种期刊被剔除，其中包括 Hindawi、MDPI、泰勒与弗朗西斯、施普林格和爱思唯尔等出版商旗下的知名期刊（Quaderi, 2023）。最后，在 2025年打击指标造假方面，科睿唯安还宣布，撤稿文章的引文将不再计入影响因子的计算（Arévalo, 2025）。

分析功能

Web of Science 核心合集提供了丰富的集成分析工具，使研究人员、机构和科学评价人员能够从多个角度分析研究产出及其影响力。在数据库中，我们可以通过“引文报告”选项进行文献计量分析，作者也拥有各自的个人档案。JCR 是此类分析与评价的基础工具，可提供期刊影响力指标，包括期刊影响因子（JIF）、引文指标、引文半衰期等。

此外，InCites 基准分析与分析[17], 这一先进的机构分析解决方案，从研究人员、研究领域、出版来源、组织、院系和资助机构六个不同的分析视角开展复杂的基准分析，并根据其《InCites 指标速查表》[18], 88，采用适用于22种实体类型的不同指标。该平台集成了基本科学指标（ESI）等工具，用于识别相关研究领域。近期推出的改进功能包括 Web of Science 研究助手[19], ，这是一种基于生成式人工智能的工具，支持自然语言搜索和多语言搜索。

[15] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/editorial-selection-process/ [16] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/editorial-selection-process/journal-evaluation-process-selection-criteria/ [17] https://clarivate.com/academia-government/scientific-and-academic-research/research-funding-analytics/incites- benchmarking-analytics/ [18] _ 24647474424337 https://incites.zendesk.com/hc/en-gb/article attachments/ [19] https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-research-assistant/

1.1.2. Scopus - 爱思唯尔

商业模式

Scopus 采用机构订阅的商业模式，由科学领域最大的跨国企业之一爱思唯尔管理。平台通过年度或多年期许可证向大学、研究中心、公共组织和图书馆提供商业化访问。定价结构同时依据研究强度和订阅机构规模制定，并针对联盟和国家协议采用差异化定价方案；根据协议类型，折扣幅度可能相当可观。该体系包含显著的年度涨价，并需定期协商。就此而言，这一模式与 Web of Science 的模式相似；CSIC 等许多国际研究中心都鼓励对这两家公司的模式提出质疑（Ansede，2024年）。

数量覆盖范围

Scopus 的书目覆盖范围尤为突出，这得益于其与来自 105 个国家的 7,000 家出版商建立的稳固合作关系，因此收录了 28,791 种正在出版的同行评审期刊（Harling，2025年）。2025年3月，Scopus 宣布其索引记录已超过 1 亿条[20]。在学科分类方面，Scopus 使用 Science All Journal Classification（ASJC）将期刊和会议论文集划分为四个主要主题领域，并进一步细分为若干组和 333 个次级领域[21]。Scopus 的优势之一是拥有丰富的开放获取内容，目前共收录此类文章 2,510 万篇，涉及 8,137 种期刊。此外，Scopus 还扩展了其内容资源，纳入了 356 万本图书，分布于 1,283 个馆藏和 399,000 个独立书名；同时还收录了大量学术活动相关内容，包括来自 167,000 场会议活动的记录和 1,290 万份演示文稿[22]。2021 年，Scopus 开始纳入预印本索引，目前已收录来自七个专业服务器的 264 万篇文章[23]。Scopus 从出版商处收到的数据中，超过 95% 为电子格式；内容平均在正式出版后 2 至 3 周内完成索引，而在线优先文章通常会在其在线版本发布后不到 4 天内上线（爱思唯尔，2023年）。

编辑标准

Scopus 对其平台上出版物的遴选与维护采用严格的编辑标准，并由内容遴选与咨询委员会（CSAB）监督；该委员会是一个由各科学学科公认专家组成的独立委员会[24]。每年大约收到 3,500 份收录申请，但只有 34% 达到可进入评估的最低标准，而在这一部分中，约有 50% 在委员会详尽的审查流程后最终获准收录。遴选标准围绕五个基本维度构建（编辑政策、内容质量、编辑稳定性、出版规律性以及在线可用性）。此外，Scopus 要求文章接受同行评审，最好采用双盲模式，每期至少有两名审稿人参与，并须符合其他强制性标准：英语摘要、罗马字母参考文献、有效的 ISSN，以及遵守与出版伦理委员会（COPE）原则相一致的伦理标准和良好编辑实践。就因不端行为将期刊除名而言，其政策似乎不如科睿唯安那般严格，不过确实存在一些有记录的案例（Travis，2025年）。

[20] https://blog.scopus.com/100-million-reasons-to-trust-scopus/ [21] https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/ [22] https://www.recursoscientificos.fecyt.es/sites/default/files/1%20ScopusIntroducci%C3%B3n.pdf [23] https://www.elsevier.com/products/scopus/content [24] https://www.elsevier.com/products/scopus/content/content-selection-and-advisory-board

分析功能

Scopus 整合了一套广泛的分析工具。在数据库内部，其功能包括分析检索结果，以及可自动生成唯一标识符以归集个人科研产出的作者档案。它还提供机构档案，以便对整个机构的科研绩效进行评价。期刊分析器模块可通过引文总计、已发表文章和影响力趋势的图表，同时比较多达十种期刊。Scopus 还包含三项标准化文献计量指标，这些指标会结合学科特点对影响力进行情境化分析。作为补充，它整合了 PlumX 指标[25], ，可提供多维分析，将 25 项以上的指标划分为五个类别：引文、使用情况、收藏、提及和社交媒体。此外，Scopus 与 SciVal[26], 紧密关联，后者是一个云端分析平台，利用 Scopus 数据在机构和团体层面提供基准比较工具。借助其仪表板以及分布在 10 个类别中的 38 项指标[27]，SciVal 能够分析科研趋势、识别关键合作并评价科学影响力。与科睿唯安一样，他们也在努力纳入并提供人工智能功能[28]。

1.1.3. OpenAlex – OurResearch

商业模式

OpenAlex 是一个完全开放且免费的数据库[29], ，由致力于开放科学原则的非营利组织 OurResearch 开发。完整的数据库数据集采用 CC0 许可证发布，任何用户均可不受限制、免费访问、下载、修改和重复使用这些数据。其财务可持续性基于一种混合模式，主要收入来源有两种：(1) 高级订阅和 (2) 咨询服务。Arcadia 向 OpenAlex 提供了 750 万美元的资助，使其能够在五年内建立一个可持续且完全开放的全球科研生态系统索引[30]。这种模式与 Scopus 和 Web of Science 等专有数据库形成对比；后者通过限制访问来创造收入，而 OpenAlex 则让我们置身于科学信息获取的新范式之中。

数量覆盖范围

在最近一篇关于 OpenAlex 的文献综述（Forchino 和 Torres-Salinas，2025年）中，读者可以查阅一份关于它的全面报告；但概括而言，我们可以说，它在全球科学文献索引方面已达到前所未有的规模，截至2024年底收录了约2.6亿条书目记录。为实现这一目标，OpenAlex 整合了多个开放来源的数据，尤其是 Crossref[31] （1.5亿件作品）、Microsoft Academic Graph 的遗留数据、PubMed、arXiv、ORCID、DOAJ 以及机构知识库。该数据库每天新增5万条记录，并将其主要来源扩展至 DataCite 和 HAL[32], ，这已落实在其最近宣布的 Walden 项目[33]中。因此，它不仅收录期刊论文，还包括图书、章节、会议论文集、数据集以及各类格式，并包含超过1亿名作者、11万个机构名称以及与作品相关联的6.5万个 Wikidata 概念的数据[34] （Ho，2025年）。为组织科学知识，它采用一套四级层级分类系统，该体系基于

[25] https://www.elsevier.com/insights/metrics/plumx [26] https://www.elsevier.com/products/scival
[27] =1328583 =9781971 https://elsevier.libguides.com/c.php?g &p
[28] https://www.elsevier.com/products/scopus [29] https://help.openalex.org/hc/en-us/articles/24397762024087-Pricing [30] https://blog.openalex.org/category/open-science/#:~:text=%247.5M%20grant%20from%20Arcadia [31] https://help.openalex.org/hc/en-us/articles/24347019383191-Where-do-works-in-OpenAlex-come-from [32] https://blog.openalex.org/openalex-2024-in-review/ [33] https://blog.openalex.org/openalex-rewrite-walden-launch/
[34] https://docs.openalex.org/api-entities/concepts

与莱顿大学科学与技术研究中心（CWTS）合作开发的自动化模型，涵盖252个子领域[35] 以及4000多个主题

编辑标准

OpenAlex 采用最大限度包容的政策，与 Web of Science 和 Scopus 等选择性数据库形成对比。OpenAlex 不采用限制性的遴选流程，而是实施全面索引，在不设置预先筛选条件的情况下整合来自多个开放来源的数据，涵盖期刊论文、会议论文集、图书、章节、数据集、软件和学位论文。其基本理念是，传统的选择性遴选会系统性地排除评审标准不同于科学、技术、工程和数学领域的学科、非英语语言的研究以及区域科研产出，因此 OpenAlex 收录所有类型的期刊，而不对其进行分析[36] ，并将筛选责任交给用户，使他们能够根据自身需求编制列表。然而，这项政策也有其后果。

一项基于 Cabells 掠夺性期刊报告[37] 目录中 400 种期刊样本的简要研究显示，被认定为掠夺性期刊的期刊中有 36.5% 被 OpenAlex 收录；据此推算，掠夺性期刊约有 6,300 至 8,200 种（Donner，2025年）。最令人担忧的案例是 OMICS 出版集团，该出版商被美国联邦贸易委员会勒令支付 5,000 万美元，其文章在 OpenAlex 中的记录总数超过 215,000 条。尽管 OpenAlex 提供筛选条件，但这些条件并非专门用于识别欺诈性期刊，因此用户必须意识到该数据库中存在可疑内容。

分析功能

OpenAlex 提供的分析能力超越了基本的书目信息检索。其 REST API[38] 支持对文本、标题和摘要进行全文检索，并通过自动词干提取显著扩大检索结果范围，同时还支持复杂的布尔运算符。用户可以从多个维度进行筛选。此外，OpenAlex 还提供直观的网页界面、批量数据下载，以及用于互操作性的持久标识符（数字对象标识符、ORCID、ROR），因此成为一个功能多样的平台，可用于创建需要透明且可重复使用数据的产品和服务。

依托 OpenAlex 的基础设施，已有众多项目利用其开放数据，包括编程库 openalexR 和 PyAlex、可视化工具 VOSviewer、分析仪表板 CWTS Dashboard 和 COKI Dashboard，以及搜索工具 SemOpenAlex 和 oa.m[39]。

[35] https://help.openalex.org/hc/en-us/articles/24736129405719-Topics [36] https://help.openalex.org/hc/en-us/articles/27719473439511-How-does-OpenAlex-exclude-predatory-journals [37] https://cabells.com/solutions/predatory-reports [38] https://docs.openalex.org/how-to-use-the-api/api-overview [39] https://help.openalex.org/hc/en-us/articles/27086501974551-Projects-Using-OpenAlex

1.2 精选与近期文献

本节将通过分析近期关于书目数据库评价的科学文献，补充前文概述。我们将系统地探讨五个关键维度：（1）记录总量，分析各平台之间的共享语料库和独有内容；（2）开放获取期刊覆盖范围；（3）语言覆盖率；（4）参考文献覆盖率；（5）元数据质量与可靠性，评价字段可用性以及文献分类的准确性。通过这项分析，我们将确定各产品的优势、局限性和具体问题。

1.2.1 记录总量

对这三个数据库所收录文献的分析表明，其覆盖范围存在显著差异（表 1）：OpenAlex 包含 2.43 亿条记录，而 Web of Science 为 7120 万条，Scopus 为 6560 万条（Culbert 等，2025）。当分析范围仅限于学术文章时，OpenAlex 仍以 2.006 亿条记录保持领先，而 Web of Science 与 Scopus 的数量相近，分别为 4260 万和 4350 万条。就 2015—2022年这一特定期间而言，在基于 DOI 的去重[40], 条件下，数字为：Web of Science 2260 万条记录，Scopus 2760 万条，OpenAlex 7680 万条；与此前各期间相比增幅显著，且在同一时间区间内 Scopus 的覆盖范围大于 Web of Science。

表 1. 根据 Culbert 等，2025 的数据库覆盖范围

	Web of Science	**Scopus**	**OpenAlex**
语料库总量	71,280,830	65,642,377	243,053,925
文章	42,678,632	43,579,595	200,665,940
2015—2022年	22,609,069	27,620,472	76,836,191
2015—2022年共有	16,788,282		

对来源之间二元交集的分析揭示了不对称的重叠模式（图1）[41]。这一分布表明，没有任何单一数据库能够收录全部科学文献。值得注意的是，Web of Science 与 Scopus 的内容有 90% 重合。另一方面，对 2015—2022年三大数据库之间的“共享语料库”[42] 进行分析后发现，其中有 1,670 万条具有有效数字对象标识符的共同记录，仅占 Web of Science 语料库的 74%、Scopus 的 60%，以及 OpenAlex 的 21% 左右，这表明各个平台都存在独有内容[43]。这种差异反映了前文讨论的编辑理念：Web of Science 和 Scopus 采用基于编辑筛选的选择性模式，而 OpenAlex 则采用基于自动化采集的全面收录模式。另一种解读是，如果按唯一性反比来计算，则可以看出，OpenAlex 收录的内容中有 78.2% 为独有内容，并未与 Web of Science 和 Scopus 共享。

[40] 去重流程以 DOI 作为唯一标识符，确保每篇出版物仅计数一次，并排除没有 DOI 或 DOI 重复的记录。[41] 需要强调的是，图1低估了实际的重叠程度，因为基于数字对象标识符的匹配方法将缺少 DOI 的 2170 万条 OpenAlex 记录、420 万条 Web of Science 记录和 250 万条 Scopus 记录排除在分析之外。 [42] 通过 DOI 匹配，识别出同时收录于这三个数据库（Web of Science、Scopus 和 OpenAlex）中的一组出版物，从而能够直接进行跨平台比较。[43] 这 1670 万条共享语料库记录占 Web of Science 的 74%、Scopus 的 60%，而仅占 OpenAlex 的 21%，这意味着每个数据库都包含独有内容：Web of Science 占 26%，Scopus 占 40%，OpenAlex 占 79%。这种不对称性表明，OpenAlex 收录了大量未同时见于另外两个平台的出版物。实际上，只依赖单一数据库，就意味着无法获取其他数据库所收录的数百万种独有出版物。

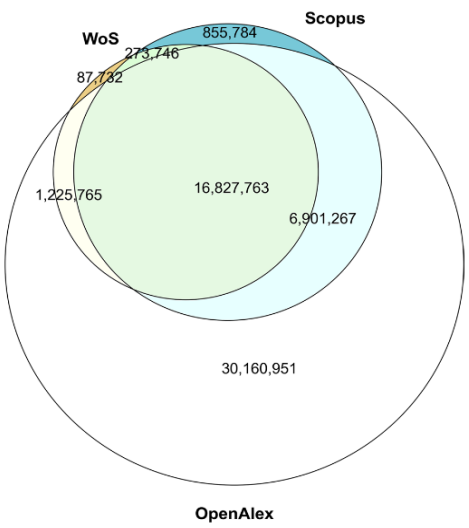


图1. 研究中不同数据库的重叠情况（Culbert 等,

‣ 30,160,951：OpenAlex 独有记录 ‣ 16,827,763：三大数据库

来源:
Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner,

1.2.2. 开放获取期刊的覆盖范围

对开放获取学术资源目录进行详细分析（以2024年5月为截止日期）后发现，Web of Science、Scopus 和 OpenAlex 在开放获取期刊索引方面存在显著差异（Maddi 等人，2025年）。在 ROAD 列出的 62,701 个有效开放获取资源中，Web of Science 收录了 6,157 种期刊（占总数的 9.8%），Scopus 收录了 7,351 种（占 11.7%），而 OpenAlex 收录了 34,217 种（占 54.6%）。这意味着，OpenAlex 覆盖的开放获取期刊数量约为 Web of Science 的 5.6 倍、Scopus 的 4.7 倍。这种数量差异也体现在独有覆盖范围上（图2）：仅由 OpenAlex 收录的开放获取期刊有 24,976 种，而仅由 Web of Science 收录的仅有 182 种，仅由 Scopus 收录的有 373 种。这表明，与传统商业数据库相比，OpenAlex 索引的开放获取期刊范围更广、种类更多。最后，ROAD 中只有 4,094 种开放获取期刊同时被三个数据库收录，约占总数的 6.5%，这说明数据库之间的实际覆盖重叠极少，而且各来源收录的全球开放获取期刊生态系统子集存在显著差异。

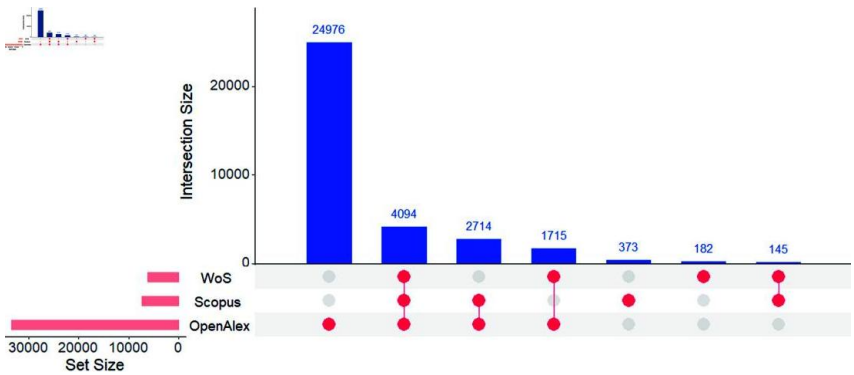


图2. Web of Science、

来源:
Maddi, A.、Maisonobe, M. 和

[44] ROAD（开放获取学术资源目录）是由国际ISSN中心提供、联合国教科文组织支持的一项免费服务，提供对 ISSN 注册库中部分内容的访问。这部分内容由描述开放获取学术资源（期刊、专著丛书、会议论文集、学术知识库及其他连续出版物资源）的书目记录组成，并补充了来自索引数据库、目录和期刊指标的元数据

结合这些结果与其他研究，我们可以推断，专有数据库对某些地理区域的代表性不足。近期一项研究（Khanna 等，2022年）分析了 25,000 多种使用开放期刊系统（OJS）平台的期刊，并估计这些期刊已发表约 580 万篇作品，分布于 136 个国家；其中 79.9% 位于全球南方，84.2% 采用钻石开放获取模式。尽管如此，Web of Science 仅收录了 1.2% 的 OJS 期刊，而 Scopus 和 OpenAlex 分别收录了 5.7% 和 63.8%。这些数据与其他有关开放获取文献覆盖范围的研究结果（Alonso-Álvarez & Eck，2025年；Simard 等，2024年）共同帮助我们理解全球学术交流中的文献多样性。所有这些研究都表明，主导性的科学评价基础设施延续了一种系统性边缘化全球南方研究成果的结构，而钻石期刊的最大规模正集中于全球南方。这种排斥不仅使整个地区的科学研究成果隐而不见，也进一步强化了英语霸权和全球北方的商业模式。

1.2.3. 语言覆盖率

主要学术数据库中的语言覆盖情况（Céspedes 等，2024）显示，英语占据明显主导地位。在 Web of Science 和 Scopus 中，超过 95% 的收录文章以英语发表，而西班牙语、中文、法语、德语、俄语或日语文章的占比各自仅为 0.3% 至 1%。相比之下，OpenAlex 呈现出不同的情况：其文章中 68% 为英语文章，同时其他语言的占比显著提高，其中中文（4.2%）、西班牙语（2.6%）、俄语（2.0%）和日语（1.3%）尤为突出，远高于 Web of Science 和 Scopus 报告的比例。

Tabl表 2. 根据 Céspedes 等人的研究，各数据库的语言覆盖情况，20 24

	Web of Science（%）	**Scopus（%）**	**OpenAlex（%）**
英语	96.4	95.0	68.0
西班牙语	0.3	0.4	2.6
语言多样性	非常低	非常低	中等

1.2.4. 参考文献覆盖率

对 2015年至2022年间发表的 1,680 万条文章记录组成的共享语料库进行深入分析（见第 1.2.1 节和脚注 43）后发现，Web of Science、Scopus 和 OpenAlex 的书目参考文献覆盖率呈现出不同的模式（Cubert 等，2025年）。将比较范围限定于这些重叠文献时，三个数据库每篇文章的来源参考文献平均数量相近，这表明对于同时被收录的近期文献，它们对来源引文的覆盖水平相似。然而，研究也揭示了各数据库在捕获参考文献总数方面存在显著差异。与 OpenAlex 相比，Web of Science 和 Scopus 每篇文章平均分别多收录 6.4 条和 6.2 条参考文献。这一差异主要集中在非来源类型参考文献（灰色文献、社论、学位论文及其他缺少标准化持久标识符的材料），以及无法规范化的书目链接上[45]。

[45] OpenAlex 在这方面的局限，源于其系统只统计被归类为来源的参考文献，而这类参考文献在很大程度上取决于 Crossref 等知识库中开放元数据的可用性。相比之下，Web of Science 和 Scopus 整合了更广泛的索引流程，因此能够纳入不同文献类型和未完全规范化的元数据中的参考文献。

表3. 根据 Culbert 等，2025 年的研究得出的参考文献覆盖率

	Web of Science	Scopus	OpenAlex
完整数据库			
参考文献数量	1,765,281,799	2,033,522,623	1,845,379,285
每条记录的平均来源参考文献数量	16.867	18.692	7.572
共享语料库（2015–2022）			
语料库中的参考文献数量	725,008,043	727,056,725	585,616,069
每条记录的平均来源参考文献数量	33.416	33.363	34.863

一些基于回溯参考文献覆盖指标的小样本案例研究[46]（Gusenbauer，2024年）显示，Web of Science 和 Scopus 对参考文献的呈现更为准确。在前瞻性计数中，Web of Science（+2.9%）和 Scopus（+2.8%）的平均百分比误差非常低，这意味着两者往往会轻微高估参考文献数量；而 OpenAlex 则呈现明显的负向偏差（−23.9%），也就是说，它系统性地漏计了文章中实际参考文献的大约四分之一。这一差异也体现在平均绝对百分比误差上，该指标衡量偏差幅度：Web of Science 为 5.8%，Scopus 为 10.1%，而 OpenAlex 为 28.9%。因此，综合 BWC 得分（结合文章覆盖率、参考文献信息和计数准确性）对 Web of Science 的评分为 94.2%，对 Scopus 的评分为 89.9%，对 OpenAlex 的评分为 68.9%。由此，Web of Science 成为回溯性参考文献检索的首选，其次是 Scopus，而 OpenAlex 则明显落后。

表4. Gusenbauer（2024年）的参考文献覆盖指标

指标	Web of Science	Scopus	OpenAlex
平均百分比误差（FWC）	+2.9%	+2.8%	−23.9%
平均绝对百分比误差（f）	5.8%	10.1%	28.9%
BWC 分数 = d·(1−f)	94.2%	89.9%	68.9%

1.2.5. 元数据质量与可靠性

正如《文献计量学来源：选择实用指南》（Arroyo-Machado，2024年）所述，元数据的质量和完整性是选择书目数据库的关键因素。当前格局清楚地呈现出成熟商业平台与新兴开放获取平台之间截然不同的做法。一方面，Web of Science 和 Scopus 一直采用选择性模式，以便进行彻底的内容筛选并实施受控索引程序，从而提高书目数据的一致性、规范化程度和可靠性。另一方面，OpenAlex 等平台优先考虑覆盖范围的广度，而非细致的质量控制，更重视可扩展性和包容性，而非元数据的最高精确度。这种方法上的分歧直接影响所提供元数据的完整性和准确性，也影响文献计量分析所需的后续数据清理与准备工作，最终决定文献计量结果的有效性和可重复性。因此，必须批判性地考察各个平台在元数据质量、规范化和可靠性方面的具体特征。

[46] Gusenbauer（2024年）的研究基于从社会科学、人文学科、健康科学和自然科学九个学科中选取的259篇期刊论文目的性样本，对59种引文索引进行了评估。样本取自近期的系统综述和荟萃分析，以确保文献具有足够完善的引文网络，从而比较不同数据库中引文链接的完整性和计数误差。研究采用了两项指标：前向引文覆盖率（FWC）和后向引文覆盖率（BWC）。

文献分类显示，各平台之间存在显著差异（Haupka等人，2024年）。Web of Science 采用包含87个文献类型类别的分类法，Scopus 使用18个类别；OpenAlex 则从非常有限的方案（3至4个类别）发展为扩展后的类型体系，此变化源于2024年5月至7月间实施的修订。可以看出，Web of Science 的分类法提供了最高的粒度，Scopus 力求在特异性与可管理性之间取得平衡，而 OpenAlex 在2024年之前优先考虑简化。在2024年5月之前的版本中，OpenAlex 存在大规模过度分类，将分析项目中的99.42%标记为“文章”；更新后，其分布更为合理，文章约占89.59%。另一项研究（Mongeon 等人，2025年）（表5）在分析2021年至2023年间发表且数字对象标识符相匹配的660多万条记录时，发现 OpenAlex 与 Web of Science 的元数据存在类似差异。最显著的差异出现在文献分类方面：有超过30万例出版物在 OpenAlex 中被归为文章或综述，而在 Web of Science 中并非如此；其中93.5%属于 OpenAlex 的错误分类。

表5. Mongeon 等人（2025年）发现的元数据差异汇总

元数据	**差异的主要特征**
文献类型	与 Web of Science 中更具体的类别相比，OpenAlex 将非原创材料过度分类为文章/综述。
语言	英语的判定结果不一致，且算法 OpenAlex 的语言检测经常出错。
出版年份	Web of 中“在线优先”年份与期出版年使用不一致。科学与 OpenAlex。
作者数量	Web of Science 对团体作者和联盟的处理方式与 Web of先锋影音科学与 OpenAlex。

差异并不仅限于文献分类。作者（Mongeon 等人，2025年）还报告了出版年份方面的差异，发现有480,884条差异记录（占总计的8.1%），这主要是因为 OpenAlex 倾向于记录首次在线发表的年份，而 Web of Science 记录的是期刊期次的年份。在作者数量方面，发现有71,133条差异记录（1.2%），这很大程度上是因为两个数据库对作者署名中列出的联盟和研究团队的处理方式不同。语言差异较少，共发现33,516例（占总计的0.6%）；不过，OpenAlex 将非英语文章错误标记为英语的错误率更高，达到93.9%，这可能是由于其基于标题和摘要进行算法语言检测。补充研究证实，与 Web of Science（95%为英语）相比，OpenAlex 具有更广泛的语言覆盖范围（75%的文章为英语），但约有7%的文章被错误分类为英语。

Culbert 等人（2025年）的研究还提供了一些关于元数据的有趣信息。例如，在2015–2022期间，他们发现 Web of Science 中有 4,186,863 条记录没有数字对象标识符，Scopus 中有 2,555,909 条，OpenAlex 中有 21,709,360 条。对于 Web of Science 和 Scopus，文献中没有标识符的比例较高，是因为纳入了数字对象标识符并非必需的文献类型和来源。对于 OpenAlex，这一问题更为突出，可能是因为它继承了来自 Microsoft Academic Graph 的语料库及其聚合政策[47]。另一方面，在共享语料库（脚注 43）中，覆盖范围

[47] 在 Web of Science 和 Scopus 中，数百万条记录没有数字对象标识符，主要是因为纳入了某些文献类型和来源，而这些文献类型和来源并未强制要求使用持久标识符，或尚未全面实施持久标识符（例如某些会议论文集、图书、图书章节和区域性期刊）；此外，数字对象标识符基础设施对不同学科、国家和语言的覆盖并不完全均衡。在 OpenAlex 中，这一现象更加突出，因为它依托从 Microsoft Academic Graph 继承而来的规模大得多的语料库，并且还从多个提供方大规模汇集期刊、知识库及其他开放来源，其中包括许多地方期刊或编辑规范化程度较低的期刊；这些来源尚未分配数字对象标识符，或其开放元数据中未显示该标识符。这在很大程度上解释了为何 OpenAlex 中没有数字对象标识符的记录超过 2100 万条，远高于 Web of Science 和 Scopus 的数字。

摘要显示，三个数据库之间存在差异。Web of Science 和 Scopus 的文章中，摘要占比均超过92%，而 OpenAlex 约为 87%，即摘要可用性相差约五个百分点。尽管从相对角度看，这一数量差距不大，但若以数千万条记录的语料库为基数进行估算，就意味着 OpenAlex 缺少数十万篇在 Web of Science 或 Scopus 中确实有摘要的文章[48]，这限制了某些搜索和内容分析应用。

尽管本研究的重点并非 OpenAlex，但元数据问题在该数据库中尤为敏感。关于这一点，我们请读者参阅我们的论文《OpenAlex 数据库回顾》（Forchino 和 Torres-Salinas，2025年），并建议通过阅读该文深入了解这一主题。除上文已讨论的问题（语言、出版年份、文献类型等）外，我们还报告了文献中指出的其他问题。例如，机构隶属信息缺失：OpenAlex 的记录中多达 61.5% 未提供任何机构隶属信息，而且这种偏差对某些出版商和地区的影响并不均衡。我们还记录了相关证据，表明 OpenAlex 仅识别出 Retraction Watch 所列撤稿中的一部分，同时还存在重复记录、引文归属错误、书目元数据缺失，以及高估参考文献数量的倾向。因此，在将 OpenAlex 用作主要数据源时，必须实施额外的验证和清理程序。

最后必须指出，在某些情境下、特定领域中，或基于规模较小且受控的样本时，上述论述可能会有所不同（Cebrian 等，2025年）。例如，西班牙作者指出，在健康科学领域，对1,404篇文章的样本进行分析后发现[49]，，“在所分析的大多数学科中，Web of Science、Scopus 和 OpenAlex 之间未观察到显著差异”；也就是说，这三个数据库为其收录的条目提供了较为完整的元数据。然而，当讨论范围涉及这些数据库的全部内容时，不能将这些观察结果直接外推。

[48] Gusenbauer（2024年）的研究基于一个目的性样本，对59个引文索引进行了评估。该样本由从社会科学、人文学科、健康科学和自然科学的九个学科中选取的259篇期刊论文组成。样本取自近期的系统综述和荟萃分析，以确保文献的引文网络足够完善，从而能够比较不同数据库中引文链接的完整性和计数误差。研究采用了两项指标：前向引文覆盖率（FWC）和后向引文覆盖率（BWC）。[49] 该研究定义了一组广泛的核心必要字段，用于评估索引质量：文章标题、作者、出版年份、来源标题（期刊）、卷、期、起止页码、被引次数、数字对象标识符、机构隶属信息、摘要、关键词、经费资助、参考文献、出版商、ISSN、PubMed ID、语言、文献类型、开放获取状态以及作者的 ORCID。研究从 Web of Science、Scopus 和 OpenAlex 中提取 I3PT（2020–2022年）同一批1,404篇文章的这些字段，并将其与各数据库实际收录的文章总数进行百分比比较，从而衡量“元数据质量”，而不仅仅是覆盖水平。

1.3. 比较总结表

表6. 引文数据库比较总结：Web of Science、Scopus 和 OpenAlex

特征	Web of Science	Scopus	OpenAlex
商业模式与访问			
访问类型	机构订阅	机构订阅	开放免费访问
所有者	科睿唯安	爱思唯尔	OurResearch（非营利）
许可证	专有	专有	CC0
数量覆盖范围			
记录总数	9700万	100+ 百万（2025年3月）	~260 百万（2024年底）
活跃收录期刊	22,704 种期刊	28,791 种期刊	来自 1.5 亿+ 部作品的数据（Crossref）
文章（完整语料库）	4260 万	4350 万	2.006亿
文章（2015–2022）	2260万	2760万	7680万
独有内容	25.7% 独有	39.2% 独有	78.2% 独有
共享语料库（2015–2022）	74.3% 重叠	60.8% 重叠	重叠部分占 21.8%
图书	收录于核心合集	356 万（全语料库范围）	收录于通用语料库
预印本	未直接纳入	264万（自2021年起）	已纳入（多个来源）
开放获取覆盖范围			
ROAD中的开放获取期刊	6,157（9.8%）	7,351（11.7%）	34,217（54.6%）
独有开放获取期刊	182种独有期刊	373种独有期刊	24,976种独有期刊
共享开放获取期刊	4,094种期刊（占ROAD总计的6.5%）	4,094种期刊（占ROAD总计的6.5%）	4,094种期刊（占ROAD总计的6.5%）
学科覆盖范围			
主要索引/学科	SCIE（9,450 种期刊）、SSCI（3,541 种）、艺术与人文引文索引（1,808 种）、新兴来源引文索引（9,504 种）	4 个学科领域，333 个 ASJC 学科领域	252 个子领域（4 级层级结构）
分类体系	182 个科学引文索引扩展版学科、58 个社会科学引文索引学科、25 个艺术	科学ASJC（333 个领域）	CWTS-莱顿模型（252 个子领域）
学科偏向	科学、技术、工程和数学领域较强（>70%），人文学科较弱（<5%），社会科学 ~15%	科学、技术、工程和数学领域较强（~60%），人文学科中等（~8%），社会科学 ~25%	对人文学科和社会科学
语言覆盖率			
英语文章	96.4%	95%	68%
西班牙语文章	0.3%	0.4%	0.026%
语言多样性	非常低	非常低	中等
地理覆盖范围			
地域偏差	明显偏向西方（北美35–40%，30–35%欧洲）	比Web of Science更偏向西方，但偏差较小（40% 欧洲，北美30%）	更公平，更能代表全球南方
已收录的 OJS 期刊	1.2%	5.7%	63.8%
编辑标准			
遴选流程	极为严格，共有 28 项标准（24 项涉及质量 + 4 影响力）	严格，5 个基本维度	包容性最高，不设预先筛选条件
评审委员会	科睿唯安学科编辑	内容遴选与咨询委员会（CSAB）	不适用（自动编目）
年度评估申请	未说明	~3,500 篇投稿（34% 符合最低标准，其中 50% 获得接收）	不适用

需要同行评审	是，强制要求	是，最好采用双盲评审	非强制要求
剔除政策	非常严格（2024年有17本期刊，2023年有 50+ 本）	较不严格，有记录的案例	无筛选／移除收录
掠夺性期刊的存在	极低（严格控制）	低（中等程度的控制）	Cabells 样本的 36.5%（估计为 ~6,300–8,200）
参考文献覆盖率			
参考文献总数	1,7.65亿	2,0.33亿	1,8.45亿
每篇文章平均参考文献数	16.8	18.7	7.6
共享语料库中的参考文献	7.25亿	7.27亿	5.85亿
每篇文章平均参考文献数（	33.4	33.4	34.9
平均百分比误差（FWC）	+2.9%（轻微高估）	+2.8%（轻微高估）	−23.9%（显著低估）
平均绝对百分比误差	5.8%	10.1%	28.9%
元数据质量与可靠性			
摘要可用性	>92%	>92%	~87%
无数字对象标识符的记录（2015–2022）	420万	260万	2170万
文献类型类别	87 个细分类别	18 个类别	2024年5月之前为3–4种，2024年之后有所扩展
文献类型分类错误	极低（~2%）	低（~1%）	高（占错误总数的 ~97%，其中研究论文/综述中的错误分类占 93.5%错误分类）
缺少机构隶属关系的记录机构隶属关系	Low	Low	High
撤稿覆盖率（Retraction Watch）	High	High	低（仅识别出有限比例）
重复问题分析功能	极低	Low	存在
集成工具	引文报告、JCR、InCites（88 项指标）、ESI、研究助手（人工智能）	分析检索结果、期刊分析器、PlumX、SciVal（38 项指标）	REST API、全文检索、多重筛选、4500主题分类
作者档案	Yes	是，提供唯一标识符	Yes
机构概况	是（InCites）	Yes	是（~110,000 个机构实体）
AI工具	Web of Science 研究助手	正在开发中	未说明
外部工具生态系统	InCites	SciVal	广泛（openalexR、PyAlex、VOSviewer、CWTS Dashboard、COKI Dashboard 等）
更新与处理			
索引速度	因索引而异	~2–平均 3 周，已刊前 <4 天	每天 50,000 条记录
数据接收格式	混合	>95% 电子格式	多个开放来源
时间覆盖范围	始于1900–1945年（因学科而异）	主要涵盖1996年及以后	广泛的时间覆盖范围
可持续性与透明度			
资金模式	机构订阅	机构订阅	Arcadia 基金会拨款（750 万美元）、高级订阅、咨询服务
透明度	低（专有模式）	低（专有模式）	高（开源、开放数据）
经证实的可持续性	是（60+ 年）	是（30年）	新兴（始于2022年）
互操作性	有限	有限	高（数字对象标识符、ORCID、ROR、Crossref 等）

第2部分

定量分析

本部分研究呈现基于原始数据的定量分析；详见方法学附录，涵盖2015—2024年期间的*Web of Science*、*Scopus*和*OpenAlex*。本节包括按文献类型、语言和研究领域进行的比较，从而可以观察三个数据库之间的异同，以及这些差异如何影响其覆盖范围和代表性。

2.1 纵向分析

表7. 按数据库统计的期刊出版物时间分布（2015—2024年）及五年期分布

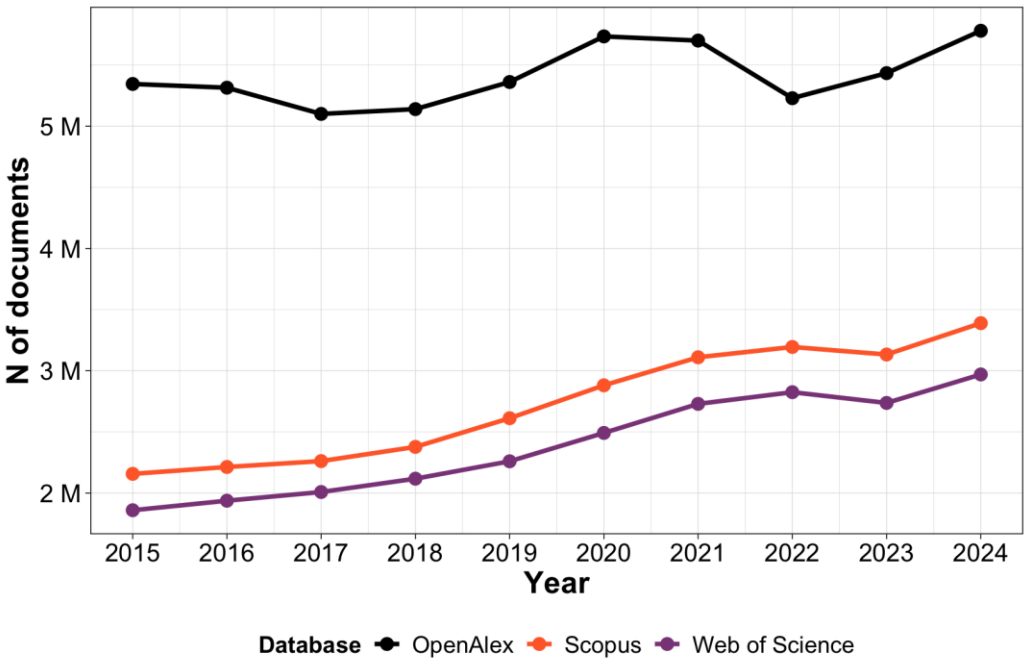
年份↓	Clarivate Web of Science™	Scopus®	OpenAlex
2015	1,859,511	2,157,076	5,345,449
2016	1,937,508	2,213,298	5,314,813
2017	2,008,616	2,261,339	5,100,536
2018	2,117,402	2,377,531	5,139,701
2019	2,259,716	2,612,050	5,361,377
2020	2,491,692	2,880,996	5,734,366
2021	2,729,213	3,110,358	5,700,252
2022	2,824,941	3,194,670	5,228,965
2023	2,736,927	3,132,908	5,433,988
2024	2,969,954	3,389,322	5,780,514
2015—2019年	10,182,753	11,621,294	26,261,876
2020—2024年	13,752,727	15,708,254	27,878,085
总计	23,935,480	27,329,548	54,139,961

2015年至2024年间期刊出版物的时间分布（表7）显示，三个数据库均呈现持续增长的模式，但增长幅度存在明显差异。*Web of Science*的记录从2015年的1,859,511条增至2024年的2,969,954条，绝对增量为111万条。*Scopus*从2,157,076条增至3,389,322条，增量为123万条，略高于*Web of Science*所记录的增量。*OpenAlex*在整个区间内始终保持显著更高的数值，介于2015年的5,345,449与2024年的5,780,514之间。其年度数值的稳定性与Web of Science和Scopus之间所观察到的较低变异性形成对比。在第一个五年期2015—2019年，*Web of Science*累计10,182,753条记录，而*Scopus*为11,621,294条，*OpenAlex*总计26,261,876条。在第二个五年期2020—2024年，*Web of Science*达到13,752,727篇出版物，*Scopus*达到15,708,254篇，这些数值表明两个数据库的比例增幅相近。*OpenAlex*在同一区间记录了27,878,085篇文献，该数字略

高于前一个五年期。在2015—2024年的累计总计中，Scopus收录27,329,548篇文献，超过*Web of Science*的23,935,480篇，而*OpenAlex*达到54,139,961篇。

图3以直观方式呈现了相同的时间分布，清晰地展示了*Web of Science*和*Scopus*在2015—2024年期间的变化轨迹十分接近。两组数据均呈现持续增长趋势，起始值约为180万至210万篇文献，到2024年超过290万至330万篇。两条曲线的斜率相似，期刊出版物数量也大致相当。相比之下，*OpenAlex*的数值始终明显更高，年度记录在500万至600万之间，且波动情况与其他数据库截然不同。虽然*Web of Science*和*Scopus*在几乎整个期间都保持稳定增长，但*OpenAlex*在某些年份出现下降，尤其是2022年。

图3. 按数据库划分的期刊出版物年度分布（2015—2024年）



2.2. 文献类型的差异

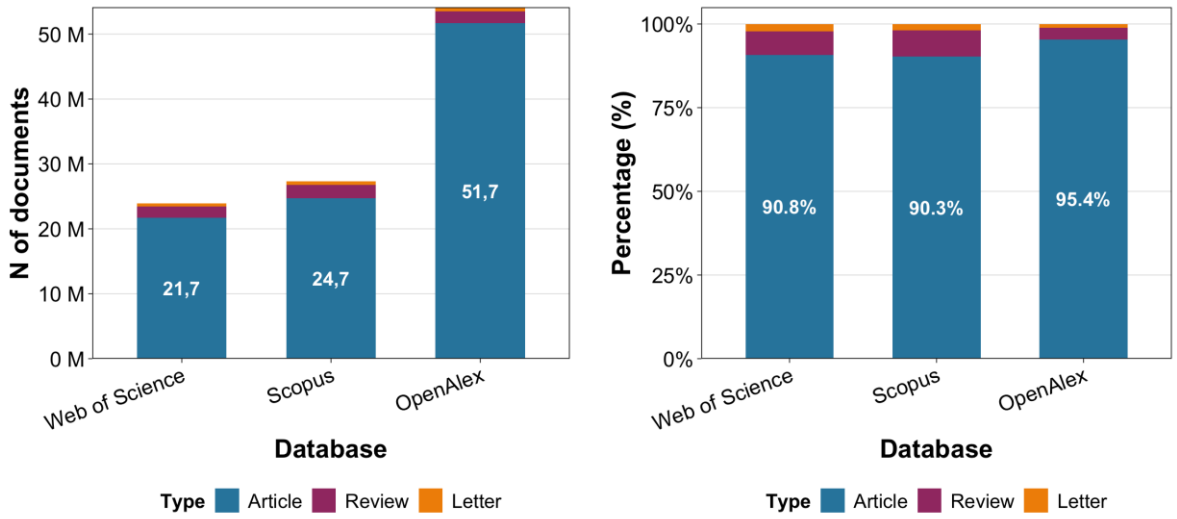
按文献类型划分的分布显示，*Web of Science*和*Scopus*（表8）的模式非常稳定：在整个分析期间，文章始终占据大量记录。2015—2024年期间，Web *of Science*收录了21,735,538篇文章，占总数的91%；而*Scopus*记录了24,683,822篇文章，占90%。两个数据库中综述的比例相近，*Web of Science*为7%，*Scopus*为8%。快报所占比例较低且相近，两者均约为2%。两个五年期也呈现出相同趋势。2015年至2019年间，*Web of Science*中的文章占92%，*Scopus*中占90%；而2020年至2024年间，两者的文章占比均为90%，综述和快报的比例也十分接近。*OpenAlex*则表现不同，整个期间文章占比更高，达到95%，而综述比例较低，为3%。快报在该数据库中所占比例微乎其微，接近1%。

表8. 按类型和数据库划分的期刊出版物分布，按五年期及总计区分

TYPE	Clarivate Web of Science™	Scopus®	OpenAlex
完整时期			
文章	21,735,538 91% 占该期间总数的 期间	24,683,822 90%	51,654,802 95%
综述	1,679,308 7%	2,126,063 8%	1,877,982 3%
信件	520,634 2%	519,663 2%	607,177 1%
2015—2019年			
文章	9,321,194 占总数的92%, 期间	10,493,146 90%	25,207,080 96%
综述	621,343 6%	886,970 8%	742,411 3%
信件	240,216 2%	241,178 2%	312,385 1%
2020—2024年			
文章	12,414,344 90% 总计的 期间	14,190,676 90%	26,447,722 95%
综述	1,057,965 8%	1,239,093 8%	1,135,571 4%
信件	280,418 2%	278,485 2%	294,792 1%

图4直观展示了整个期间按文献类型划分的分布情况，使三个数据库在数量级上的差异更加清晰。*Web of Science*和*Scopus*无论在绝对卷数还是内部结构上，数值都十分接近。两者的文章均占总数的最大部分，数量分别接近2,200万和2,500万篇，占比均约为91%。综述数量较少，占比约为7%—8%；快报所占比例几乎可以忽略，接近2%。无论是绝对数量图还是百分比图，都能看到这种分布上的相似性；两个商业数据库中的比例几乎完全一致。*OpenAlex*对应的柱形则呈现出不同的特征，其文章数量显著更高，超过5,100万篇。百分比图显示，文章占比达到95%，凸显了这一类别在其内部结构中的主导地位。

图4. *Web of Science*、*Scopus*和*OpenAlex*中文献类型的绝对数量及百分比分布（2015—2024年）



2.3. 主题概况

整个期间的主题分布（表9）显示，Web of Science和Scopus在领域分类方面保持着非常相似的结构。在这两个数据库中，生物医学与健康科学均为主要核心领域：Web of Science有10,178,006条记录，Scopus有9,932,031条记录，分别占分类总数的43.3%和37.3%。物理科学与工程、生命与地球科学这两个领域在两个数据库中的占比相近，Web of Science中分别介于14.8%和23.1%之间，Scopus中则介于17.7%和22.5%之间。社会科学与人文学科方面，数据差异较小，Web of Science占9.0%，Scopus占12.1%。OpenAlex呈现出不同的分布，在社会科学与人文学科方面占比更高，共有7,390,532条记录，相当于20.4%；而生物医学与健康科学的占比则较低，为34.5%[50]。

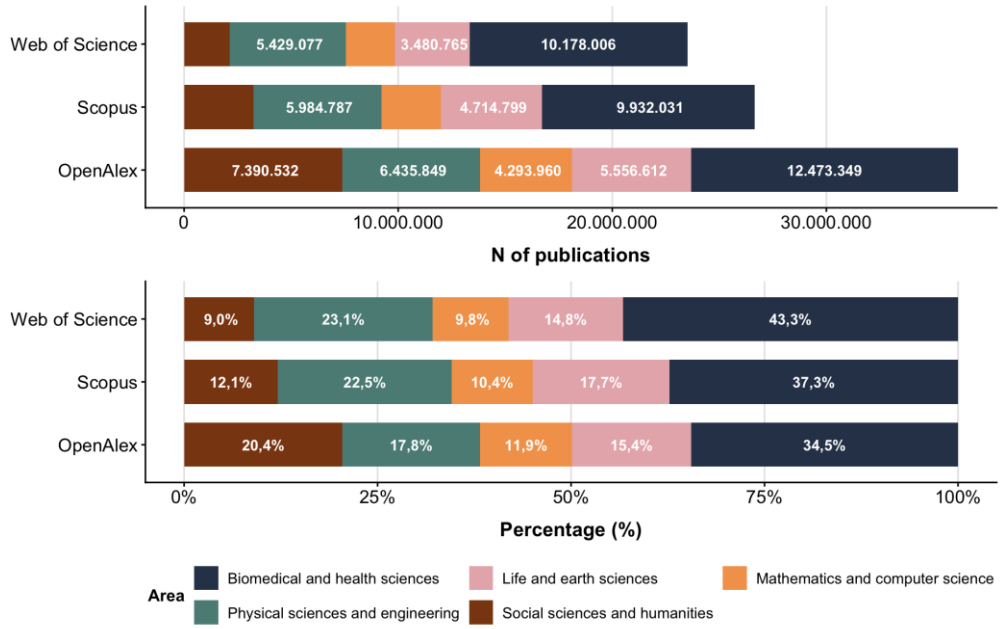
图5可以直观地展现主题构成。从图中可以看出，Web of Science 和 Scopus 的生物医学与健康科学领域的产出明显占主导地位，在 Scopus 中占比超过 37%，在 Web of Science 中则超过 43%。其余领域的分布较为均衡，物理科学与工程、生命与地球科学的占比相近。尽管两个数据库的主题模式相似，但社会科学与人文学科领域存在一定差异：Web of Science 占 9%，Scopus 则达到 12.1%。数学与计算机科学领域也存在轻微差异。OpenAlex 的可视化结果呈现出不同的特征：社会科学与人文学科的占比显著上升，在已分类内容中超过 20%；而生物医学与健康科学的相对占比低于商业数据库。

[50] 值得注意的是，由于采用了算法处理过程，主题分类未能覆盖整个语料库。*Web of Science* 对总量的分类比例为 98%，*Scopus* 为 97%，*OpenAlex* 为 67%。覆盖范围的这一差异直接影响比较，因为 OpenAlex 中未分类的部分要大得多，从而影响对主题占比的相对解读。

表9. *Web of Science*、*Scopus*和*OpenAlex*中按研究领域分类的出版物分布（2015—2024年）

AREA	Clarivate Web of Science™	Scopus®	OpenAlex
完整时期			
社会科学与人文学科	2,121,562 9.0% 占分类总量	3,229,265 12.1%	7,390,532 20.4%
生物医学与健康科学	10,178,006 43.3%	9,932,031 37.3%	12,473,349 34.5%
物理科学与工程学	5,429,077 23.1%	5,984,787 22.5%	6,435,849 17.8%
生命与地球科学	3,480,765 14.8%	4,714,799 17.7%	5,556,612 15.4%
数学与计算机科学	2,298,690 9.8%	2,784,176 10.4%	4,293,960 11.9%
总计	23,508,100 98% 的总计	26,645,058 97%	36,150,302 67%

图5. 完整时期2015—2024年*Web of Science*、*Scopus*和*OpenAlex*中分类研究领域的绝对数量及百分比分布



2.4. 语言差异

按语言分布（表10）显示，三个数据库中的文献高度集中于英语。*Web of Science*记录了23,022,125篇英语文献，占总数的96.2%。*Scopus*的比例略低，为90.8%。在次要位置上，两个数据库呈现出相似的模式：西班牙语、中文、俄语、德语和葡萄牙语的占比较低，介于0.3%和1%之间。这些数据凸显出除英语之外语言多样性不足的构成特征。所谓“其他语言”类别在*Web of Science*中涵盖60种语言，在*Scopus*中涵盖54种语言，但所占比例均很小。*OpenAlex*则呈现出不同的格局，其中77.5%的文献使用英语。西班牙语位居第二，占2.8%；其次是印度尼西亚语，占2.4%，以及葡萄牙语，占2.3%，这表明其语言多样性更高，非西方语言的存在也更为显著。在排名中段，俄语、日语、法语和德语的占比介于1.5%和2.2%之间。*OpenAlex*中的其他语言集合达到44种语言，共计2,849,423篇文献，占5.3%，这进一步证实了其分布更加分散，并且对多语言学术产出的收录范围更广。

表10. *Web of Science*、*Scopus*和*OpenAlex*主要出版语言分布（2015—2024年）

排名	Clarivate Web of Science™	Scopus®	OpenAlex
1st	英语 GB 23,022,125 – 96.2%	英语 GB 24,826,757 – 90.8%	英语 GB 41,967,957 – 77.5%
2nd	西班牙语 ES 262,823 – 1.1%	中文 CN 1,209,576 – 4.4%	西班牙语 ES 1,538,391 – 2.8%
3rd	俄语 RU 117,226 – 0.5%	西班牙语 ES 273,381 – 1.0%	印度尼西亚语 ID 1,316,112 – 2.4%
4th	德语 DE 103,343 – 0.4%	俄语 RU 254,843 – 0.9%	葡萄牙语 PT 1,241,259 – 2.3%
5th	葡萄牙语 PT 101,569 – 0.4%	德语 DE 158,368 – 0.6%	俄语 RU 1,177,828 – 2.2%
6th	中文 CN 87,605 – 0.4%	法语 FR 127,461 – 0.5%	日语 JP 1,077,453 – 2.0%
7th	法语 FR 83,265 – 0.3%	葡萄牙语 PT 91,341 – 0.3%	法语 FR 831,271 – 1.5%
8th	意大利语 IT 34,716 – 0.1%	日语 JP 76,222 – 0.3%	德语 DE 787,045 – 1.5%
9th	土耳其语 TR 25,622 – 0.1%	意大利语 IT 51,961 – 0.2%	韩语 KR 735,882 – 1.4%
10th	韩语 KR 18,370 > 0.1%	韩语 KR 39,255 – 0.1%	中文 CN 615,510 – 1.1%
其他	**60种语言** 78,816 – 3.3%	**54种语言** 220,383 – 8.1%	**44种语言** 2,849,423 – 5.3%

2.5. 研究领域详情

表 11. 按文献类型、英语出版物和带有数字对象标识符的出版物划分的研究领域详情

		Clarivate Web of Science™		Scopus®		OpenAlex	
		文献	%	文献	%	文献	%
HSS	文章 2	,052,532	96.7%	2,962,058	91.7%	7,317,486	99.0%
	综述	62,845	3.0%	249,058	7.7%	69,205	0.9%
	信件	6,185	0.3%	18,149	0.6%	3,841	0.1%
	以英语出版	1,934,582	91.2%	2,860,590	88.6%	5,226,194	70.7%
	具有 DOI 的出版物	,037,193	96.0%	2,967,147	91.9%	6,951,718	94.1%
SHB	文章 8	,598,877	84.5%	8,328,442	83.9%	10 ,895,689	87.4%
	综述 1	,107,386	10.9%	1,180,386	11.9%	1,282,681	10.3%
	信件	471,743	4.6%	423,203	4.3%	294,979	2.4%
	以英语出版	9,754,305	95.8%	9,188,553	92.5%	11,523,340	92.4%
	具有 DOI 的出版物	9 ,325,621	91.6%	9,533,514	96.0%	12 ,282,682	98.5%
ESP	文章 5	,179,302	95.4%	5,696,167	95.2%	6,280,987	97.6%
	综述	243,575	4.5%	278,955	4.7%	151,250	2.4%
	信件	6,200	0.1%	9,665	0.2%	3,612	0.1%
	以英语出版	5,354,329	98.6%	5,562,684	92.9%	6,256,346	97.2%
	具有 DOI 的出版物	,372,293	99.0%	5,840,192	97.6%	6,352,946	98.7%
SEL	文章 3	,275,520	94.1%	4,420,199	93.8%	5,323,138	95.8%
	综述	191,037	5.5%	270,084	5.7%	221,623	4.0%
	信件	14,208	0.4%	24,516	0.5%	11,851	0.2%
	以英语出版	3,407,947	97.9%	4,243,497	90.0%	5,185,030	93.3%
	具有 DOI 的出版物	,395,790	97.6%	4,517,536	95.8%	5,408,595	97.3%
SCM	文章 2	,239,998	97.4%	2,708,628	97.3%	4,231,447	98.5%
	综述	52,880	2.3%	65,975	2.4%	59,559	1.4%
	信件	5,812	0.3%	9,573	0.3%	2,954	0.1%
	以英语出版	2,269,424	98.7%	2,576,189	92.5%	4,037,105	94.0%
	具有 DOI 的出版物	2 ,253,799	98.0%	2,646,581	95.1%	4,155,320	96.8%

2.6. 来源的重叠与趋同

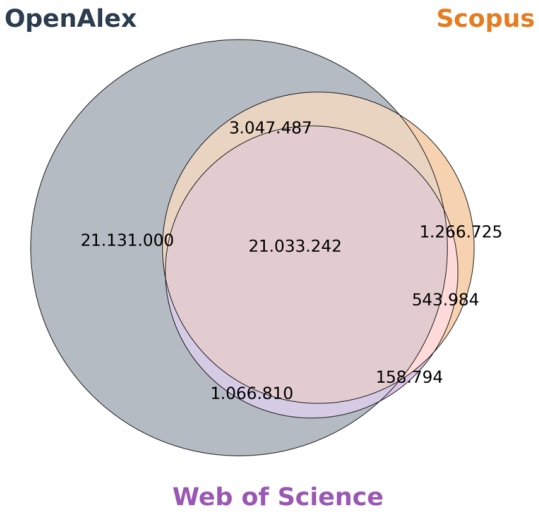


图6. 根据原始数据比较不同数

数据库之间的比较显示，存在非常广泛的共同核心。如图6和表12所示，约有2100万篇文献同时收录于*Web of Science*、*Scopus*和*OpenAlex*中，这与*Web of Science*和*Scopus*之间观察到的高重叠比例相符（*Scopus*收录了*Web of Science*中90%的文献，而*Web of Science*收录了*Scopus*中79%的文献）。尽管*OpenAlex*的覆盖范围最广，拥有超过2100万条独有记录，但它与其他数据库的整合程度较低，其文献中仅有41%收录于*Web of Science*，44%收录于*Scopus*。即便如此，它仍保持着显著的交集，尤其是与*Scopus*的交集；而*Web of Science*是独有文献数量最少的数据库（略多于100万篇）。总体而言，数据表明，*Web of Science*和*Scopus*构成了一个重合度很高且相对均衡的核心，而*OpenAlex*则是覆盖面最广的数据库，其馆藏中既包含大量自身独有的文献，也与另外两个数据库有相当程度、但明显不那么全面的重叠。

表12. 根据原始数据比较不同数据库的两两重叠情况

来自 第↓行数据库的文献 出现在第 → 列	Clarivate Web of Science™	Scopus®	OpenAlex
Clarivate Web of Science™		21,585,087 **90%** Scopus 中的 Web of	22,100,240 **92%** OpenAlex 中的 Web of Science
Scopus®	21,585,087 **79%** Web of Science 中的 Scopus		24,080,992 **88%** OpenAlex 中的 Scopus
OpenAlex	22,100,240 **41%** Web of Science 中的 OpenAlex	24,080,992 **44%** Scopus 中的 OpenAlex	

表13按五个科学领域——社会科学与人文学科（SSH）、生物医学与健康科学（BHS）、物理科学与工程（PSE）、生命与地球科学（LES）以及数学与计算机科学（MCS）——展示了*Web of Science*（WoS）、*Scopus*和*OpenAlex*之间的文献重合度。总体而言，*Web of Science*和*Scopus*在所有领域均呈现出最高且最稳定的重叠水平，重合度超过87%，在PSE和LES中接近97%，表明这两个数据库的覆盖范围高度一致。相比之下，*OpenAlex*呈现出更大的差异：尽管其收录了各领域的大量文献，但与*Web of Science*和*Scopus*的重合比例明显较低，尤其是在SSH（30%）和MCS（67%）领域。最突出的数据一方面体现了*Web of Science*和*Scopus*在所有学科上的高度一致性，另一方面也显示出*OpenAlex*的差异性更大：其覆盖范围广，但重叠程度较低，尤其是在社会科学以及数学与计算机科学领域。

表13. 根据原始数据，按领域统计的不同数据库两两重叠情况

来自 第↓行数据库的文献 出现在第 → 列	Clarivate Web of Science	Scopus	OpenAlex
Clarivate Web of Science		**SSH** 1,839,398 - **87%** **BHS** 8,889,497 - **87%** **PSE** 5,269,693 - **97%** **LES** 3,295,378 - **95%** **MCS** 2,167,372 - **94%** **Web of Science 中的**	**SSH** 1,977,081 - **93%** **BHS** 9,070,597 - **89%** **PSE** 5,251,121 - **97%** **LES** 3,295,987 - **95%** **MCS** 2,190,428 - **95%** **Web of Science 在 OpenAlex 中的**
Scopus	**SSH** 2,228,571 - **69%** **BHS** 8,026,133 - **81%** **PSE** 5,260,879 - **88%** **LES** 3,789,161 - **80%** **MCS** 2,180,309 - **78%** **Scopus 中的 Web of Science**		**SSH** 2,810,508 - **87%** **BHS** 8,971,544 - **90%** **PSE** 5,448,461 - **91%** **LES** 4,109,580 - **87%** **MCS** 2,420,616 - **87%** **Scopus 在 OpenAlex 中**
OpenAlex	**SSH** 2,183,824 - **30%** **BHS** 8,133,579 - **65%** **PSE** 5,031,123 - **78%** **LES** 3,744,308 - **67%** **MCS** 2,633,055 - **61%** **OpenAlex 在 Web of Science 中**	**SSH** 2,598,052 - **35%** **BHS** 8,809,335 - **71%** **PSE** 5,185,953 - **81%** **LES** 4,036,588 - **73%** **MCS** 2,856,309 - **67%** **Scopus 中的 OpenAlex**	

第3部分

3.1. 执行摘要

生态系统中的商业模式与定位
所考察的三个数据库采用了截然不同的模式：Web of Science 和 Scopus 沿用由来已久的订阅制策略，并通过严格的遴选流程确保稳定性和可控性。相比之下，OpenAlex 采取开放、可扩展的方式，快速扩充内容，同时依赖多元异质的来源。对于管理者而言，这带来了一个运营困境：是选择经过遴选、可靠性更高的数据库，还是选择开放、覆盖面更广的数据库。两者并非互斥，但这是一个战略性选择，必须与评价、数据审计和机构分析的要求相匹配。

覆盖范围与卷量差异
文献证实，OpenAlex 的覆盖范围显著更广，记录数量超过 Web of Science 或 Scopus 的三倍。更广的覆盖范围伴随着更高比例的独有内容，以及与商业数据库较低的重叠度。相比之下，Web of Science 和 Scopus 的规模相近，结构性重叠约为 90%，因此其实际覆盖范围也较为接近。对管理者而言，不应将卷量与经过遴选的全面性混为一谈：每个数据库收藏的出版物总体各不相同，其基础范围和用途也各有差异。

开放获取期刊与地理多样性
综述研究呈现出一个一致的模式：在全球开放获取生态系统中，OpenAlex 的代表性更强，尤其涵盖了在 Web of Science 和 Scopus 中代表性不足的地区和出版商。这直接影响到全球南方中处于边缘地位的学科和国家；这些地区的期刊往往采用钻石开放获取模式。Scopus 的表现居中，而 Web of Science 的覆盖范围则更具选择性。这一模式表明，评估决策必须明确考虑这些结构性偏差，以避免系统性地使植根于本地或非英语的研究体系处于不利地位。

语言、参考文献与引文模式
文献显示，Web of Science 和 Scopus 收录的内容中，英语占比均超过 95%；相比之下，OpenAlex 的语言多样性明显更高。在参考文献方面，Web of Science 和 Scopus 的覆盖范围和精确度更高，而 OpenAlex 则系统性地低估了大量未经规范化处理的文献，尤其是灰色文献。对于基于引文的分析，特别是回溯性分析，商业平台仍表现出更强的稳健性。不过，在广泛的描述性分析中，OpenAlex 提供的图景更加多元和包容。

元数据质量与运营风险
独立研究发现，与 Web of Science 和 Scopus 相比，OpenAlex 在文献类型、出版年份、语言和机构隶属关系方面的错误率显著更高；前两者的编辑流程更加一致，所呈现的差异也更少。这并不否定 OpenAlex 的价值，但意味着在使用前更加需要进行数据清洗和验证。对于管理者而言，相关文献的结论十分明确：使用开放数据的技术成本更高，但当目标是突破传统编辑规范、扩大分析的代表性时，这种成本是合理的。

时间动态与增长趋势
纵向分析显示，2015年到2024年期间，Web of Science 和 Scopus 呈现出平行且稳定的增长，年度增量相近。OpenAlex 的总体卷数更高，但年际波动也更为明显，这反映了来源纳入方面的变化。对于管理者而言，Web of Science 与 Scopus 之间的趋同证实了它们作为成熟且整合完善的系统所具备的稳定性，而 OpenAlex 的波动则应被理解为其更加动态、更加开放扩展的生态系统的一部分。

文献类型与结构一致性
Web of Science 和 Scopus 的分布几乎完全相同：文章约占 90%，综述占 7–8%，信件及其他类型约占 2%。相比之下，OpenAlex 将文章的比例提高到约 95%，并大幅减少其他类别。这表明，商业数据库提供了更细致的文献分类，而 OpenAlex 采用了更简化的分类方案。在评价场景中，Web of Science 和 Scopus 的这种同质性有助于开展纵向比较，并增强标准化指标的稳健性。

主题概况与学科侧重
在主题方面，Web of Science 和 Scopus 展现出非常相似的结构，生物医学与健康科学占据明显主导地位，其他领域的分布则较为均衡。OpenAlex 改变了这一分布，SSH（社会科学与人文学科）的占比相对较高，而生物医学的权重较低。这种模式可能影响机构分析、内部排名和多元化研究，因为每个数据库反映的学科格局各不相同，彼此并不能完全互换。

科学产出中的语言与文化多样性
定量结果证实，Web of Science 和 Scopus 仍由英语出版物主导，占总量的 90% 以上。OpenAlex 收录的语言种类更为广泛，拓展了地域和文化范围。对于西班牙语或非英语格式的产出占比较高的机构而言，这一差异可能具有决定性影响。开放数据库更能体现全球文献多样性，但由于语言标注可能存在错误，仍需进行额外验证。

全球重合度与战略互补性
最终分析揭示，Web of Science 与 Scopus 之间存在极为广泛的共同核心，而 OpenAlex 则提供了数量最多的独有内容。这两种情况结合起来表明，单靠任何一个数据库，都不足以全面诊断全球科学活动。对于管理者而言，结论很明确：Web of Science 和 Scopus 能提供评估上的一致性，而 OpenAlex 则显著扩大了代表性。经过周密规划，互补地使用这些数据库，可以在质量控制与覆盖广度之间取得平衡，从而减少文献计量分析中的学科、语言和地域偏差。

3.2. 最终建议

采用多源策略并进行交叉验证。
证据表明，没有任何单一数据库能够收藏全球科学产出的全部内容。对于严谨的文献计量分析，建议结合不同来源，在全面性（如 OpenAlex）与高质量元数据（如 Web of Science、Scopus）之间取得平衡，并根据拟计算的具体文献计量指标采用量身定制的验证方案。机构研究应明确报告各个案例中所使用的数据库，并为这一选择提供方法论依据。

根据分析目的制定差异化方案。
对于会影响科研职业发展、经费资助或认证的正式评价，建议优先使用 Web of Science 或 Scopus，因为它们在关键元数据（机构隶属信息、文献类型、参考文献）方面更为可靠。对于探索性分析、学科图谱绘制或区域与语言覆盖范围研究，OpenAlex 是一种很有价值的补充资源。选择来源时，应根据评价情境，审慎权衡所需的全面程度与可接受的元数据准确性。

解决非英语研究评价中的结构性偏差。
Web of Science 和 Scopus 延续了对全球南方、非英语语言及钻石开放获取出版模式的系统性代表不足。仅依赖这些来源的评价流程，会对人文学科、社会科学以及具有本地出版传统的学科中的区域研究造成显著偏差。建议在评价设计中纳入对地理与语言偏差的评估，并以区域性专业资源（如 SciELO、Dialnet）补充核心数据库。

明确说明 OpenAlex 的局限性。
尽管 OpenAlex 大幅提升了获取科学信息的能力，但采用该数据库时，必须明确说明已知的运行局限：已发现掠夺性期刊的存在（估计为 6,300–8,200 种）、文献分类存在系统性错误、参考文献计数偏低（~24%）、机构隶属信息缺失（61.5% 的记录缺少机构隶属信息），以及对关键元数据的处理不一致。基于 OpenAlex 的报告应具体说明这些局限，并在可行的情况下，通过抽样与精选数据库进行交叉验证。

加强机构分析能力和持续监测。
书目数据库格局变化迅速。建议持续投入资源，对专业技术人员进行培训，并为文献计量分析制定稳健的质量方案，使其适用于不同的数据来源。机构还可以实施定期监测机制，随着书目数据库进行技术改进、更新编辑政策或调整覆盖范围，对其进行比较评估，从而确保战略决策以最新证据为依据。

附录方法

1. 数据
本分析基于 Web of Science 核心合集、Scopus 和 OpenAlex 的完整数据集，这些数据集分别于2025年9月、2025年3月和2025年8月下载。尽管下载日期并未完全同步，但这一差异不会影响本研究，因为分析仅涵盖截至2024年被索引的成果，且不考虑引文，而引文对时间变化更为敏感。为确保不同平台之间具有可比性，数据集限定为文章、综述和快报，并视需要应用额外筛选条件。在 Web of Science 中，仅保留期刊出版物，因此排除了图书章节和会议论文集；在 OpenAlex 中，则根据其包容性索引政策，排除了预印本、数据集及其他不可比的文献类型。

2. 数据处理
各数据库内部的计数均依据其原生唯一标识符进行：Web of Science 使用 UT（唯一文章标识符），Scopus 使用 EID，OpenAlex 使用 Work ID。这些标识符可确保在各平台内进行稳健且无重复的计数。然而，由于这些标识符无法互操作，跨数据库比较采用 DOI（数字对象标识符）作为关联键，并经过规范化处理，包括大小写转换为小写等修正与转换。使用 DOI 能够准确识别匹配项、重叠内容以及平台特有内容，从而确立分析的比较框架。

3. 规范化与主题分析
鉴于各数据库采用不同的学科分类体系，我们通过应用莱顿算法，彻底取代了原有的分类法。此方法构建出一个包含四个层级的通用层级分类，确保所有三个平台均按照相同的五大学科领域进行组织：社会科学与人文学科（SSH）、生物医学与健康科学（BHS）、物理科学与工程（PSE）、生命与地球科学（LES），以及数学与计算机科学（MCS）。这一流程避免了引入与编辑差异或平台特定分类算法相关的偏差，并形成了一个连贯、可复现且可在各数据库之间充分比较的主题框架。

4. 工具
完成数据规范化并确定共同的主题分类后，开展了技术数据处理。使用 JupyterLab 4.3.4 中的 Python 3.13 对元数据进行清洗、去重和标准化，并通过专门的程序处理数字对象标识符规范化、重复项控制、字段验证和不一致性检测。数据可视化则通过 RStudio 2025.09.0 Build 387 中的 R 4.5.1 实现，并将两个环境整合到可复现的工作流程中。

参考文献

Alonso-Alvarez, P., & Eck, N. J. van. (2025). *OpenAlex*中非洲研究出版物的覆盖范围、元数据完整性与准确性：比较分析（编号：arXiv:2409.01120）。arXiv。https://doi.org/10.48550/arXiv.2409.01120 Ansede, M. (2024). CSIC与五十多个组织反对制约全球科学发展的私有数据库。《国家报》。https://elpais.com/ciencia/2024-04-16/el-csic-y-medio-centenar-de-organizaciones-rechazan-las-bases-de-datos-privadas-que-condicionan-la-ciencia-mundial.html Arévalo, J. A. (2025). 科睿唯安发布《期刊引证报告》2025年版，在科学诚信方面有所改进，覆盖范围也前所未有。*Universo Abierto*。https://universoabierto.org/2025/06/19/clarivate-presenta-la-edicion-2025-del-journal-citation-reports-con-mejoras-en-la-integridad-cientifica-y-una-cobertura-sin-precedentes/ Arroyo-Machado, W. (2024). 文献计量学来源：选择与使用实用指南。Editorial UOC。Arroyo-Machado, W., & Torres-Salinas, D. (2025). 激励措施加快西班牙开放获取的发展。《自然》，*646*(8084), 288-288。https://doi.org/10.1038/d41586-025- 03276-1 Attride, D., & Orrall, A. (2024). 十七种期刊因涉嫌引文操纵而被取消影响因子。*Retraction Watch*。https://retractionwatch.com/2024/06/27/seventeen-journals-lose-impact-factors-for-suspected-citation-manipulation/ Cebrian, G., Borrego, Á., & Abadal, E. (2025). OpenAlex和Crossref作为替代Web of Science和Scopus的健康科学书目数据来源。《西班牙科学文献学杂志》，*48*(1), 1649。https://doi.org/10.3989/redc.2025.1.1649 Céspedes, L., Kozlowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., Poitras, C., Ninkov, A. B., Ebrahimy, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2024). 评估*OpenAlex*的语言覆盖范围：元数据准确性与完整性分析（版本2）。arXiv。https://doi.org/10.48550/ARXIV.2409.10633 Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). OpenAlex与Web of Science和Scopus的参考文献覆盖分析。《科学计量学》，*130*(4), 2475-2492。https://doi.org/10.1007/s11192-025-05293-3 Donner, P. (2025). *OpenAlex*中掠夺性期刊的普遍程度。https://www.open- bibliometrics.de/posts/20250603-QuestionableJournals/ Elsevier B.V. (2023). *Scopus*内容覆盖范围。爱思唯尔。https://assets.ctfassets.net/o78 2548 4 1 1 44 1 8 8 2 36 79 38 5972 9 5472 em y w i /EX iy VxBeQKf aN XzOp/c f db cb a ad a ec/Scopus_ContentCoverage_Guide_WEB.pdf Forchino, M. V., & Torres-Salinas, D. (2025). *OpenAlex*数据库回顾：评估其应用、能力与局限。https://doi.org/10.5281/ZENODO.17357948 Gusenbauer, M. (2024). 超越谷歌学术、Scopus和Web of Science：评估59个数据库引文索引的后向与前向引文覆盖率。《研究综合方法》，*15*(5), 802-817。https://doi.org/10.1002/jrsm.1729 Harling, L. (2025). Scopus数据量突破1亿条！爱思唯尔*Scopus*博客。https://blog.scopus.com/100-million-reasons-to-trust-scopus/ Haupka, N., Dörner, S., & Jahn, N. (2024). *OpenAlex*与*Web of Science*和*Scopus*相比，文献类型分类的最新变化。https://subugoe.github.io/scholcomm_analytics/posts/openalex_document_types/

Haustein, S., Schares, E., Alperin, J. P., Hare, M., Butler, L.-A., & Schönfelder, N.（2024年7月23日）。估算*2019*年至*2023*年间向六家出版商支付的全球开放获取文章处理费。arXiv.Org。https://arxiv.org/abs/2407.16551v1 Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I.（2015）。文献计量学：科研指标的莱顿宣言。《自然》，*520*(7548)，429-431。https://doi.org/10.1038/520429a Ho, L. H. S.（2025）。全球科研生态系统能否被开放索引？*Katina - Librarianship Elevated*。https://katinamagazine.org/content/article/resource-reviews/2024/can-the-worlds-research-ecosystem-be-openly-indexed Khanna, S., Ball, J., Alperin, J. P., & Willinsky, J.（2022）。重新校准学术出版的范围：迈向广阔去殖民化进程的一小步。《定量科学研究》，*3*(4)，912-930。https://doi.org/10.1162/qss_a_00228 Maddi, A., Maisonobe, M., & Boukacem-Zeghmouri, C.（2025）。开放获取期刊的地理与学科覆盖范围：OpenAlex、Scopus与Web of Science。*PLOS ONE*，*20*(4)，e0320347。https://doi.org/10.1371/journal.pone.0320347 Mongeon, P., Hare, M., Riddle, P., Wilson, S., Krause, G., Marjoram, R., & Toupin, R.（2025）。调查*OpenAlex*与*Web of Science*在文献类型、语言、出版年份和作者数量方面的差异（编号：arXiv:2508.18620）。arXiv。https://doi.org/10.48550/arXiv.2508.18620 Quaderi, N.（2023）。维护*Web of Science*学术记录的完整性。https://clarivate.com/academia-government/blog/supporting-integrity-of-the-scholarly-record-our-commitment-to-curation-and-selectivity-in-the-web-of-science/ Simard, M.-A., Basson, I., Hare, M., Lariviere, V., & Mongeon, P.（2024）。*OpenAlex*、*Scopus*和*Web of Science*的开放获取覆盖范围（编号：arXiv:2404.01985）。arXiv。https://doi.org/10.48550/arXiv.2404.01985 Torres-Salinas, D.（2024）。评价文献计量学原理。Editorial UOC。Torres-Salinas, D., Robinson-García, N., & Jiménez-Contreras, E.（2023）。通向技术与社会变革的文献计量学路径：当前问题与挑战综述。《信息专业人士》，e320228。https://doi.org/10.3145/epi.2023.mar.28 Travis, K.（2025）。爱思唯尔在Retraction Watch调查后将某期刊移出Scopus。*Retraction Watch*。https://retractionwatch.com/2025/05/16/elsevier-removes-journal-from-scopus-after-retraction-watch-inquiry/

本报告于2026年1月22日在纳斯里德王朝的格拉纳达城完成。这个日期令人想起西班牙图书管理员、语文学家和词典编纂者María Moliner；她是鸿篇巨著《西班牙语用法词典》的作者，于1981年的同一天逝世。她独自耗时十五年编纂此书，这项工作构成了二十世纪西班牙语领域最重要的贡献之一。第二共和国时期，她积极从事图书馆工作，推动当时的教育现代化和文化推广项目。西班牙内战结束后，她成为佛朗哥主义镇压的受害者：遭到清洗，被禁止担任公职，职业生涯还被降了十八级。尽管遭受排斥，她仍有尊严地坚持下来，并在马德里的家中倾注才华，编纂出那部知名词典。

玛丽亚·莫利纳及其遗产万岁！

InfluScience



Bibliometricians in arms