OpenAlex：用于检索和分析学术成果的开放数据库——特征、优势与局限性

预印本

Ángel Borrego 和 Cristóbal Urbano巴塞罗那大学
信息与视听媒体学院 信息、传播与文化研究中心（CRICC）
Melcior de Palau,140; 08014 巴塞罗那，西班牙
borrego@ub.edu; urbano@ub.edu

2025年12月19日

摘要

OpenAlex 是一个开放书目数据库，在科学评价体系旨在转变为基于开放数据、更加透明的信息来源这一背景下，它被视为商业平台的替代方案。本文分析了其特征、信息来源、实体、优势与局限性。结果显示，大量记录缺少摘要、隶属机构和参考文献；文献类型和语言的识别存在不足；规范控制和版本管理也存在问题。尽管 OpenAlex 已被应用于一些重要举措，且所得结果与商业数据库相当，但其元数据存在缺口且缺乏一致性，这表明需要进行大量数据清洗，也说明使用时应谨慎。研究最后提出了改善数据质量的三项行动方向：提高出版商完善一手来源元数据的积极性；建立协调机制，以汇集机构用户的贡献；为项目配备充足的人力资源，并制定可靠流程，以开展内部质量控制工作和处理用户支持请求。

关键词 书目数据库 · 元数据 · OpenAlex · 科研评价 · 学术传播

arXiv:2512.16434v1 [cs.DL] 18 Dec 2025

1 引言

引文索引起源于1963年，当时 Eugene Garfield 创立了科学引文索引，最初收录613种期刊和140万条引文（Urbano，2016年）。这个数据库不仅提供了一种基于引文的创新型信息检索系统，使人们能够追踪研究之间的影响与联系，还通过促进对出版物可见度的测量，彻底改变了科学评价。随着社会科学引文索引和艺术与人文引文索引的创立，其覆盖范围得以扩大，最终这两个索引被整合进科学网平台。

多年来，科学网在全球科研评价中几乎处于独占地位，是学术出版物引文分析和影响力测量的主要来源。2004年，随着爱思唯尔推出 Scopus，这种主导地位开始衰退。Scopus 是一个数据库，期刊覆盖范围更广，尤其涵盖社会科学和人文学科领域的期刊，以及非英语出版物（Mongeon 和 Paul-Hus，2016年）。

同样是在2004年，谷歌学术作为一款免费工具推出，采用了不同的方法：对网络上可获取的各种学术文献进行自动全文索引，包括文章、书籍、报告和学位论文。谷歌学术的覆盖范围远广于传统书目数据库，但对其来源质量的控制较弱（Orduña-Malea 等，2016年）。

2009年，微软学术推出，旨在与谷歌学术竞争。尽管其早期发展受限，且曾一度暂停，该项目于2016年重新启动，覆盖范围有所扩大，语义分析能力也更加先进。尽管有所改进，微软还是宣布[1] 于2021年5月正式停用微软学术（Chawla，2021年）。

微软学术于2022年初关闭后不久，OurResearch便宣布推出OpenAlex。[2]OurResearch是一家非营利组织，以创建Unpaywall而闻名。Unpaywall是一款浏览器扩展程序，可索引科学文章的免费合法开放获取版本。OpenAlex以亚历山大图书馆命名，是一个开放书目数据库，旨在提供一种可持续、易于访问且开源的替代方案，用于分析全球科学产出；其模型和数据部分源自微软学术图谱（Chawla，2022年）。2025年9月，OurResearch宣布将品牌名称变更为OpenAlex，[3] 涵盖Unpaywall和Unsub两项服务；后者是一款用于评估期刊订阅套餐及取消选项的工具。

OpenAlex是一个开源数据库，这意味着任何希望查阅、分析或重复利用其数据的人都可以免费访问这些数据。与科学网或Scopus不同，OpenAlex无需订阅即可查阅和下载其信息。此外，其技术基础设施和数据模型均有公开文档，促进了透明度和新工具的开发（Priem等，2022年）。

OpenAlex通过赠款、捐赠和高级客户获得资金支持。2024年2月，法国高等教育与研究部宣布向OpenAlex提供资金支持，[4] 认可其为开放科学的关键基础设施，并承诺改进其数据，尤其是与法国研究相关的数据。一个月后，即2024年3月，OpenAlex宣布获得Arcadia基金会提供的750万美元赠款。[5] 与此同时，OpenAlex还通过高级订阅用户获得收入，这些用户可享受更频繁的数据库更新、不限量的API查询和优先支持。[6] 因此，OpenAlex作为一个非营利项目，通过捐赠和高级订阅维持运营。目前尚不清楚这种模式能否长期持续，[7] 还是最终会停止运营，甚至回归传统的商业数据模式（Cao等，2025年）。

OpenAlex被用于CWTS莱顿排名开放版，[8] 法国的机构，例如索邦大学，[9] 以及法国国家科学研究中心（CNRS），[10] 已宣布采用它作为Scopus和科学网的替代方案。自2024年4月发布《开放研究信息巴塞罗那宣言》以来，[11] 人们对OpenAlex的兴趣显著增长。该宣言提出，通过推动使用更透明且符合开放科学原则的指标，变革科学评价体系。在此背景下，OpenAlex凭借其提供的免费、可审计的书目数据，被视为一种具有战略意义的替代方案。

2 OpenAlex 的数据来源

OpenAlex 的数据来自多个开放源，其中最重要的是微软学术图谱和 Crossref，[12] 此外，DataCite 等近期新增的重要来源也为其提供数据。[13]

Crossref 是一家成立于2000年的非营利组织，旨在促进在线科学内容的一致识别与链接。其主要职责是担任学术出版物的DOI注册机构。数字对象标识符可唯一标识数字对象（文章、数据集、专著、报告等），并在文档的整个生命周期内保持不变，

[1]https://www.microsoft.com/en-us/research/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/[2]
https://blog.ourresearch.org/were-building-a-replacement-for-microsoft-academic-graph/[3]
https://blog.openalex.org/were-now-openalex/[4]
https://www.ouvrirlascience.fr/french-ministry-of-higher-education-and-research-partners-with-openalex-to-develop-a-fully-open-bibliographic-tool/[5]
https://blog.openalex.org/ourresearch-receives-7-5m-grant-from-arcadia-to-establish-openalex-a-milestone-development-for-open-science/[6]https://help.openalex.org/hc/en-us/articles/24397762024087-Pricing[7]
OpenAlex首席执行官 Jason Priem 于2025年5月举办了一场网络研讨会，阐述他对项目可持续性的看法；这一事实清楚表明，用户正在就这一问题提出关键疑问。录像可在 OurResearch 的 YouTube 频道观看：https://youtu.be/CZ5Q9To1zCc?si=k7Pht-pz8Q0D8ky[8]https://open.leidenranking.com[9]
https://www.sorbonne-universite.fr/en/news/sorbonne-university-unsubscribes-web-science[10]
https://www.cnrs.fr/en/update/cnrs-has-unsubscribed-scopus-publications-database[11]
https://barcelona-declaration.org[12]https://help.openalex.org/hc/en-us/articles/24397285563671-About-the-data[13]
https://datacite.org/blog/datacite-metadata-is-now-integrated-in-openalex/

并与其元数据（包括其 URL）相关联。这确保了对文档的稳定访问，前提是出版商在 URL 发生变化时更新元数据。

Crossref 是 OpenAlex 的重要来源，因为它提供由出版商注册的学术内容元数据：文献标题、作者、隶属机构、期刊、出版日期、参考文献等。作为开放源代码，Crossref 元数据可以自由复用。不过，Crossref 元数据的准确性和可靠性取决于参与该系统的出版商所提供的信息，因为该组织不会对这些数据进行清理或整理。[14]

一些出版商可能会提供不完整的元数据，导致记录不完整，无法反映出版物的所有相关信息。在某些情况下，缺乏验证控制可能导致错误甚至虚假元数据被纳入，从而影响依赖这些数据的系统（如 OpenAlex）的可靠性。Besançon等人（2024年）的一项研究发现，Crossref 中纳入了一些“偷偷混入”的参考文献，这些参考文献被登记为科学文章的元数据，但实际上并未出现在相应文章中；这一做法利用了出版商与 Crossref 之间的信任关系。他们对同一出版商旗下三种期刊的分析发现，所分析的参考文献中至少有 9% 属于“隐匿”参考文献。尽管这些引文并未出现在已发表的文章中，却被存入 Crossref，并传播至将其用作数据来源的平台记录中。

3 OpenAlex 实体

在 OpenAlex 中，实体是用于构建和组织数据库内书目信息的基本单位。OpenAlex 使用九种实体类型：文献、作者、来源、机构、主题、关键词、出版商、资助机构 和地理信息。

以下将介绍前六种实体的主要特征，并举例说明。这些示例取自 2025年11月最后两周通过数据库网页版提交的查询。Codina（2024年）认为，OpenAlex 的搜索界面比 Scopus 或科学网更有限、功能也更少；本文不对该界面进行分析。

3.1 文献

这些是单篇学术出版物，例如文章、书籍、数据集或报告。每篇文献都包含标题、作者、来源和出版日期等信息，以及与其他文献之间的引文关系（被引参考文献和获得的引文）。

截至2025年11月，OpenAlex 收录了 2.713 亿条已索引文献；此外还可加入一个“扩展包”（*xpac*），其中包含另外 1.92 亿条元数据质量较低的记录，来源于 DataCite 和机构知识库（图 1）。

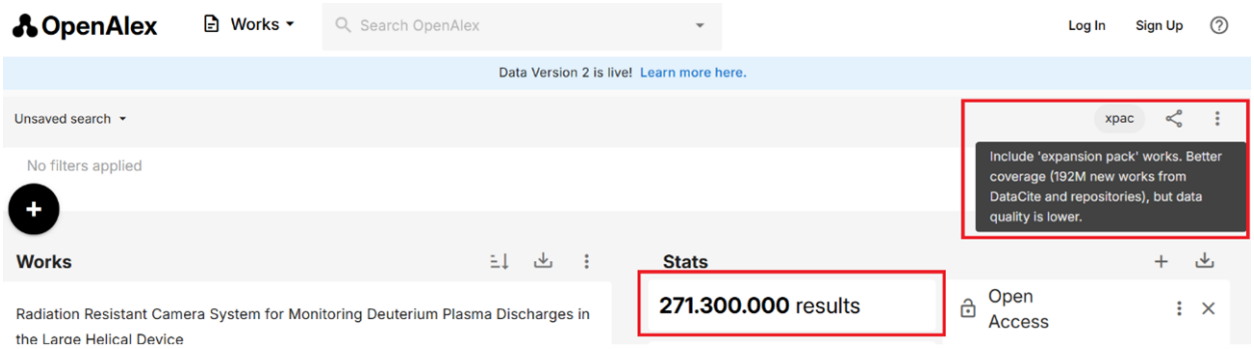


图 1： OpenAlex 中的文献

3.1.1 摘要

在标准版本中（不包含*xpac*），40%的记录（1.073亿条）缺少摘要。[15] 是否提供摘要，取决于原始来源是否提供了摘要，以及该摘要是否可以合法传播。并非所有

[14]https://www.crossref.org/documentation/principles-practices/[15]
https://openalex.org/works?page=1&filter=has_abstract:false

出版商都会在提交给 Crossref 或 OpenAlex 使用的其他来源的元数据中加入摘要；很多时候，他们只在自己的网站上发布摘要。此外，有些摘要受版权保护，这意味着 OpenAlex 无法分享这些摘要。van Eck 和 Waltman（2025）关于 2023年和2024年 Crossref 摘要存储情况的研究发现，在所研究的两年期间，爱思唯尔、泰勒-弗朗西斯、IEEE 和美国化学学会等出版商均未存储任何摘要。

3.1.2 文献类型

识别 OpenAlex 中文献的文献类型往往存在问题。一项针对 660 万条记录样本的研究（Mongeon 等，2025年）发现，超过 30 万条出版物记录在 OpenAlex 中被归类为“文章”或“综述”，但在科学网中却被赋予了不同的文献类型。几乎所有经过人工核验的案例都表明，OpenAlex 中的分类有误。本研究发现的差异不包括科学网将某篇文本归为“文章”、而 OpenAlex 将其归为“综述”，或反之的情况。若将这些差异也纳入考量，两个来源之间的分歧程度会显著更高。
在 OpenAlex 中搜索标题含有短语“new editor in chief”的文献，可得到 1,751 条结果。[16] 虽然这些文本大多可能是社论，但其中仅有 20%（350 条记录）被归为社论，而大多数（74%，1,301 条记录）被归类为文章。

此外，是否应将图书馆指南（图书馆指南）纳入其中也值得商榷，这类记录超过170万条，[17] 究竟能否支持对科学活动的有效分析；或者是否应将超过380万条记录归入“副文本”类别。[18] 另一方面，专利并未纳入其中。

3.1.3 语言

确定文献的语言同样存在问题。Mongeon 等人（2025年）通过比较英语与其他语言，分析了语言识别的准确性。尽管发现了一些差异，但其出现频率低于文献类型方面的差异，而且未发现科学网与 OpenAlex 之间存在明显差别。

对 OpenAlex 中 6,836 篇文章样本的语言元数据完整性与准确性进行人工分析（Céspedes 等，2025年）表明，与科学网相比，OpenAlex 的语言覆盖更为均衡；不过，分析也发现了文献语言识别方面的错误，导致数据库中的英语文献数量被高估。

上述两项研究均得出结论：文献语言识别错误主要有两个原因：OpenAlex 的语言检测算法出现错误，以及部分期刊同时以多种语言发表文章和摘要。

3.1.4 参考文献

参考文献是另一个带来困难的元数据要素。在 OpenAlex 的标准版本中（不包括 *xpac*），搜索不含参考文献的文献（参考文献数量 $= 0$）[19] 会检索到数据库中 64% 的记录（1.735 亿条）。

Culbert 等人（2025年）分析了由三个数据库收录的 1680 万篇近期出版物，得出结论：OpenAlex 的每个来源平均参考文献数量和内部覆盖率与科学网和 Scopus 相当。然而，OpenAlex 收录了这些记录中的 5.86 亿条参考文献，而科学网收录了 7.25 亿条，Scopus 则收录了 7.27 亿条。同样，Thelwall 和 Jiang（2025年）得出结论，OpenAlex 适用于大多数字段的引文分析。Scheidsteger 等人（2025年）的另一项研究将 OpenAlex 得出的领域标准化引文指标与三个商业数据库——科学网、Scopus 和 Dimensions 计算的指标进行了比较。该研究重点考察了 48 所德国大学在 2013年 至 2017年间于四个主要学科领域发表的约 33.5 万篇文章。尽管在文章层面观察到总体一致，但按大学和学科领域细分数据后，发现了显著差异。

创建作品记录时，OpenAlex 不会收录未编入该数据库索引的文献的参考文献。一篇文献引用的参考文献可能来自数据库中已编入索引的其他文献，也可能来自未编入索引的文献。在科学网

[16] https://openalex.org/works?page=1&filter=display_name.search:新+编辑+总+编 [17] https://openalex.org/works?filter=type:types/图书馆指南 [18] https://openalex.org/works?filter=type:types/副文本 [19] https://openalex.org/works?page=1&filter=referenced_文献_count:0

或 Scopus 中，第二类参考文献会作为次级条目添加到数据库中：这些文献不会在查询中检索出来，但会被识别为引文。在 OpenAlex 中，这些参考文献则会丢失。
在检索不含参考文献的文献时，所得结果中被引用次数最多的作品是《R：统计计算语言与环境》，[20] 这是一篇结构严谨的灰色文献，其中包含一个“参考文献”部分，列出了九份在 OpenAlex 中仍不可见的文献。该例说明，与商业数据库相比，OpenAlex 对范围广得多的文献类型持开放态度，同时也凸显了其无法对所有文献的参考文献进行一致处理的局限性。

图2展示了参考文献丢失的另一个例子。左上部分（A）显示了一份文献的记录，OpenAlex 认定该文献有五条参考文献。[21] 右上部分（B）列出了这五条参考文献。[22]最后，下方部分（C）展示了原始文章中包含的九条参考文献。[23] 来源文献中的参考文献1、2、6、7和8被纳入了 OpenAlex 记录。美国教育部的三条参考文献以及一条专著参考文献（原文档中的参考文献9）尽管已被 OpenAlex 编入索引，却遗失了。[24]

3.2 作者

在 OpenAlex 中，经常会发现作者档案重复。例如，OpenAlex 在其主页上以“Claudia Goldin”为例推荐作者搜索。她是一位美国经济学家，于2023年获得诺贝尔奖。然而，搜索这位作者会发现，她的出版物分散在三个独立的档案中（图3）。

OpenAlex 提供了一个表单，供用户申请更正作者档案，[25] 例如修改姓名格式、合并同一作者的多个档案、删除属于其他作者的出版物，或补充遗漏的出版物。不过，OpenAlex 目前对这类申请的响应速度非常缓慢，因此许多申请数月来一直处于“待处理”状态。[26]

3.3 机构

OpenAlex 中一个常见的问题是缺少机构隶属关系。根据 Zhang 等人（2024年）的一项研究，OpenAlex 中超过60%的记录完全或部分缺少机构信息。这个问题在较早期的出版物以及社会科学和人文学科领域尤为常见。

这一类别信息的另一个问题是，分配给某些文献的部分隶属机构有误。由于需要人工核验，这个问题在文献中受到的关注较少。Bordignon（2024）考察了巴黎桥路学校[27] 在科学网、Scopus 或 HAL 收录的书目产出中，有多少也被 OpenAlex 收录，以及在 OpenAlex 中搜索该机构时检索到的出版物是否正确归属。结果凸显了 OpenAlex 召回率之高：它检索到了巴黎桥路学校 93% 的产出；但也显示出其精确率不足：搜索该机构时检索到的文献中，有 24% 实际属于其他机构。

图4展示了一个 OpenAlex 记录示例[28]，其中算法将作者所属机构与资助部分中提及的机构[29] 以及其他看似与该文献无关的组织（例如一家韩国制药公司）合并在一起。这些错误的隶属机构信息还延伸到了作者档案：OpenAlex 中的作者档案经常列有数十个机构隶属关系，其中许多机构作者从未与之有过关联。[30]

---

[20]https://openalex.org/works/w2582743722[21]https://openalex.org/works/w4416039086[22] https://openalex.org/works?page=1&filter=cited_by:w4416039086[23] https://doi.org/10.51583/ijltemas.2025.1410000050[24]https://openalex.org/works/w2135943618[25] https://help.openalex.org/hc/en-us/articles/27283405287319-How-can-I-fix-errors-in-an-OpenAlex-author-profile [26]例如，一项于2025年8月27日提交的“请求”（ID 5410）截至2025年11月30日仍未解决。目前尚不清楚高级用户是否能享有更快的响应时间。 [27]一所隶属于巴黎综合理工学院的法国工程学院。网站：https://ecoledesponts.fr[28]https://openalex.org/works/w4402770154[29]https://doi.org/10.14198/DOXA2024.48.11[30] 例如，这个问题影响了这篇预印本的作者：对于 Ángel Borrego，有八条错误的“曾任职机构”（https://openalex.org/authors/a5079282617）；对于 Cristóbal Urbano，则有 11 条（https://openalex.org/authors/a 5065832938）。

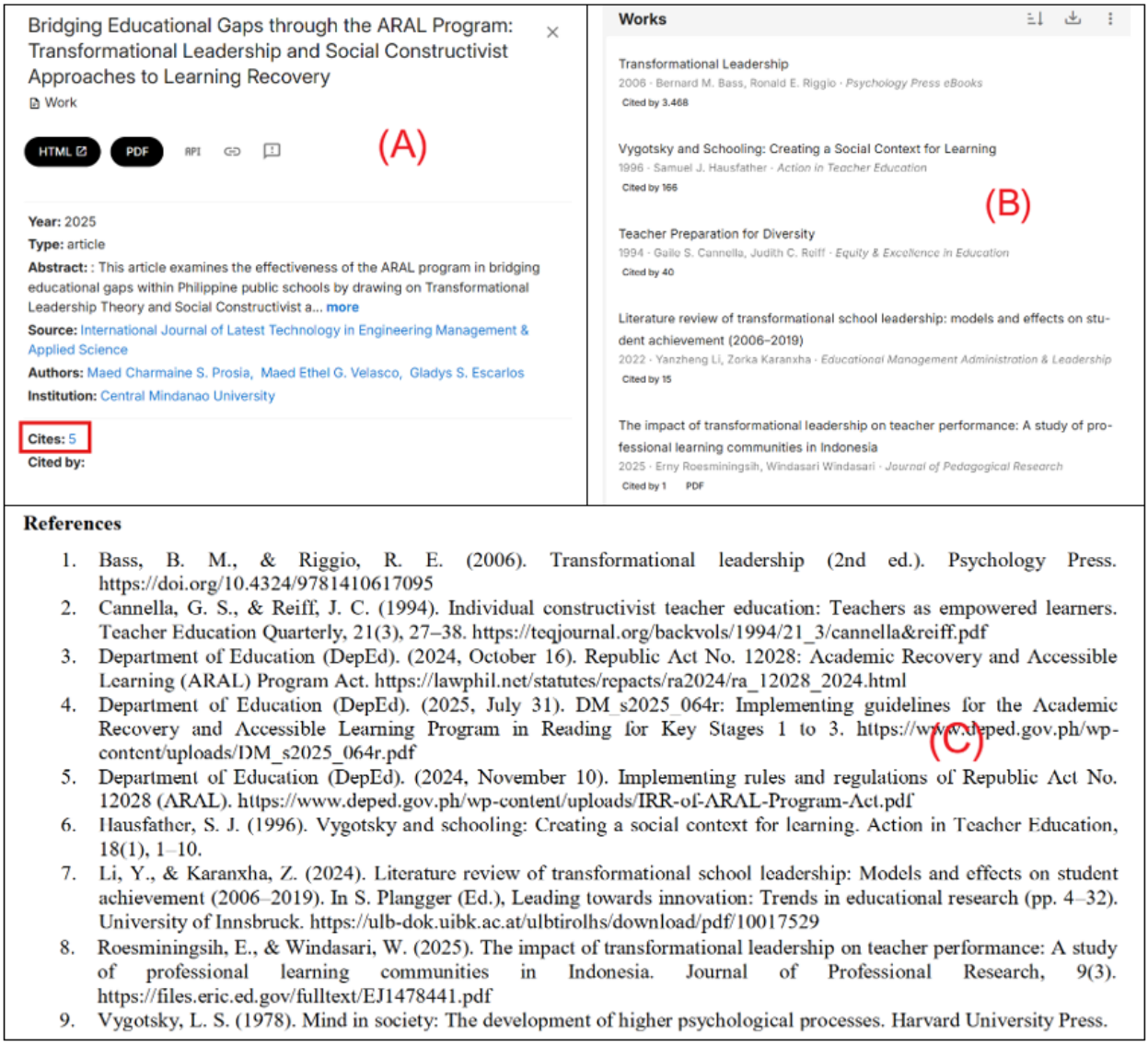


图2： (A) OpenAlex 中的记录；(B) OpenAlex 记录中的参考文献；(C) 原始文档中的参考文献

在法国，高等教育与研究部发起了 Works-magnet[31] 计划，旨在通过图书馆员和科研管理人员的审核与验证，“加速”（Jeangirard，2024年）法国出版物书目元数据的改进。这项计划由法国用户群体推动；政府（高等教育与研究部）和机构（法国国家科学研究中心与索邦大学）对该工具的支持，促使图书馆工作人员和科研管理人员分析其与各自机构相关的可靠性（Bordignon，2024年；Bach 等，2025年）。这类由社群推动的计划有望为 OpenAlex 发展成为协作生态系统铺平道路，因此 OpenAlex 会引导用户（包括来自其他国家、需要识别问题并提出解决方案的用户）使用 Works-magnet。然而，尽管 Works-magnet 扮演着“加速器”的角色，但其目前的响应能力受限于待纠正错误数量庞大，以及 OpenAlex团队可用资源相对有限；最终，所需的更改仍须由该团队落实。[32]

[31]https://works-magnet.esr.gouv.fr[32]通过 Works-magnet 请求的更改会记录在 GitHub 个人资料“openalex-affiliations”的“Issues”部分（https://github.com/dataesr/openalex-a ffiliations/issues）中，该个人资料“展示了 OpenAlex 隶属机构信息可以得到改进的案例”。截至 2025年11月30日，“Issues”部分列出了 106,087 个未解决案例，其中只有 51,420 个似乎已关闭。待处理案例的积压问题正在迅速增加，因为截至 2025年8月底，共收到 73,995 个问题，已关闭 49,903 个。

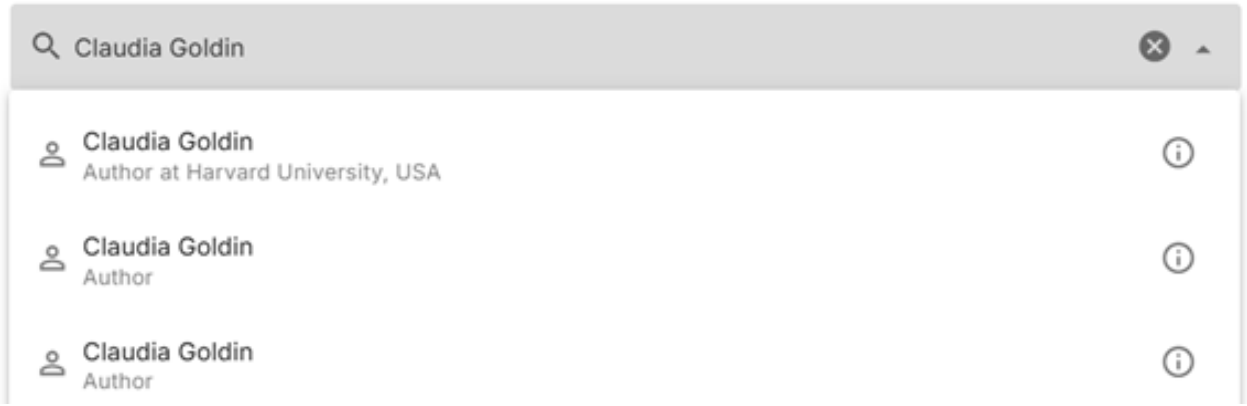


图3： Claudia Goldin 在 OpenAlex 中的个人档案

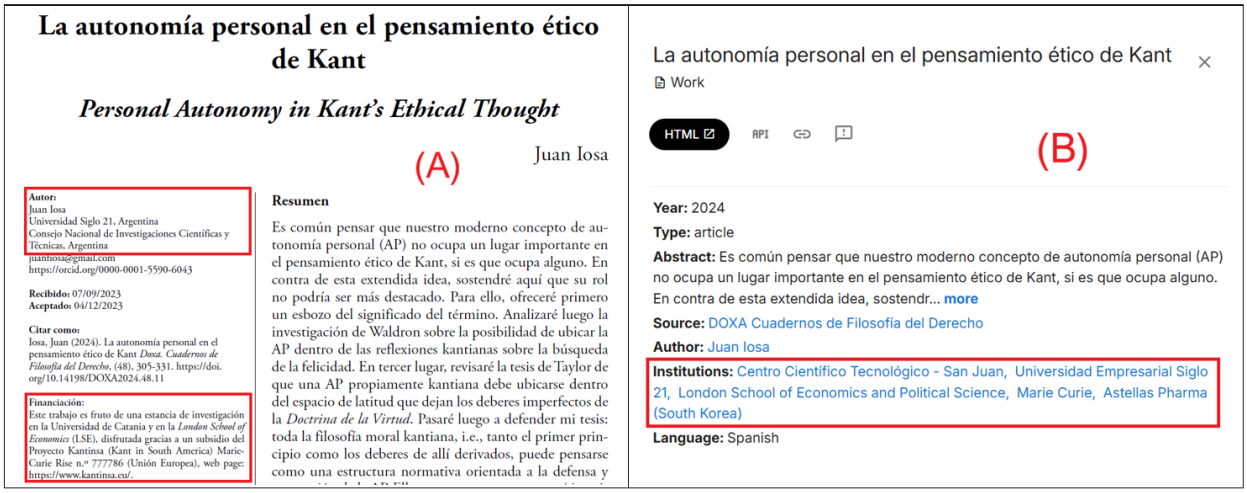


图4： （A）原始文献；（B）OpenAlex 记录，其中显示了资助部分中的隶属机构，以及其他与该文献无关的隶属机构

3.4 来源

OpenAlex 收录的文献所发表的来源包括期刊、会议、知识库和数据库。来源的多样性给文献版本控制带来了挑战，因为发布文献的一手来源（如期刊）与仅发布一手来源中所载文献参考信息的次要来源（如索引与文摘数据库）被合并在一起。[33] 其他来源（如知识库）则混合了一手文献与次要文献，例如一手文章的预印本或后印本。简言之，来源的多样性要求对文献进行出色的版本控制，而 OpenAlex 并不总能做到这一点。

图 5 展示了查询来源为 PubMed 的文献时检索到的首批记录，并按引文数排序。[34]事实上，PubMed 不发布原始文献，而只索引发表于其他来源（主要是期刊）的文章。

根据查询结果中显示的第一条记录，PubMed 中的文档版本是主要版本（图 6A）。[35] 实际上，这篇文章最初发表于*Nature*，并且在 OpenAlex 中有自己的记录；该记录应与 PubMed 记录关联，并将*Nature* 中的版本设为主要版本（图 6B）。[36]

[33]作为其组成异质性的一个例子，值得注意的是，按记录数量排名前十的来源分别是：PubMed、医学昆虫学与动物学、arXiv、HAL、Dialnet、SSRN、Zenodo、全球生物多样性信息机构、开放获取期刊目录和爱思唯尔电子书：
https://api.openalex.org/works?group_by=primary_location.source.id&per_page=10[34]
https://openalex.org/works?page=1&filter=primary_location.source.id:s4306525036[35]
https://openalex.org/works/w1964184380[36]https://openalex.org/works/w2114360920

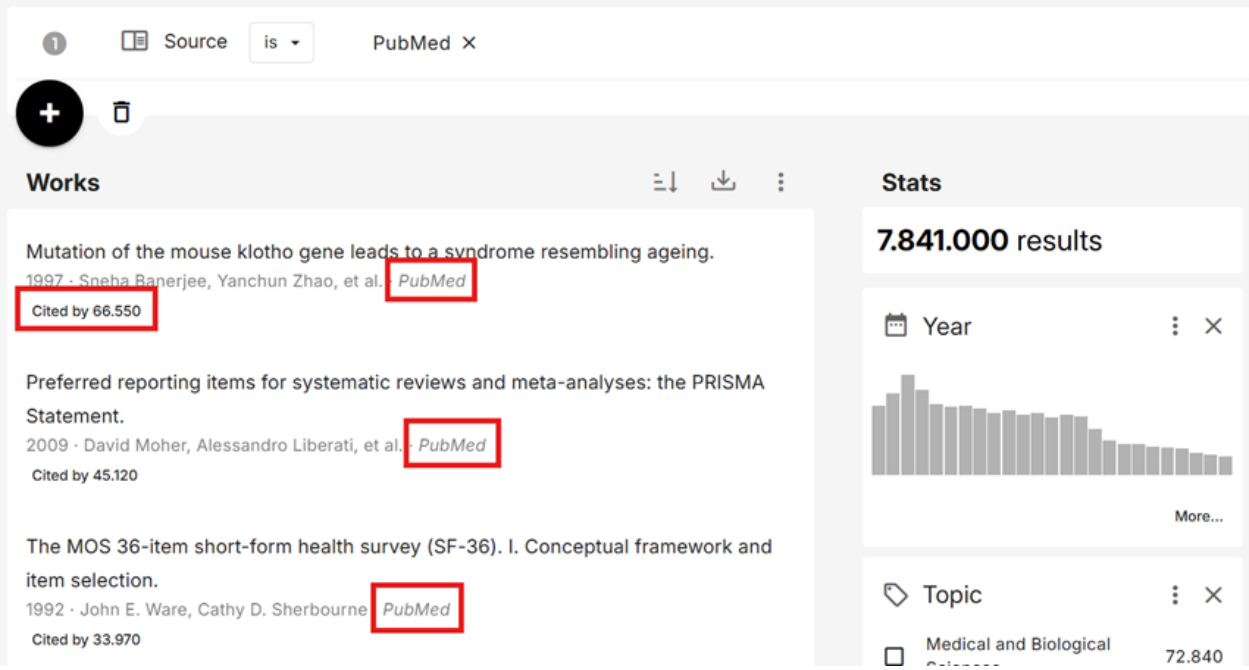


图 5：以 PubMed 为来源的文献搜索结果

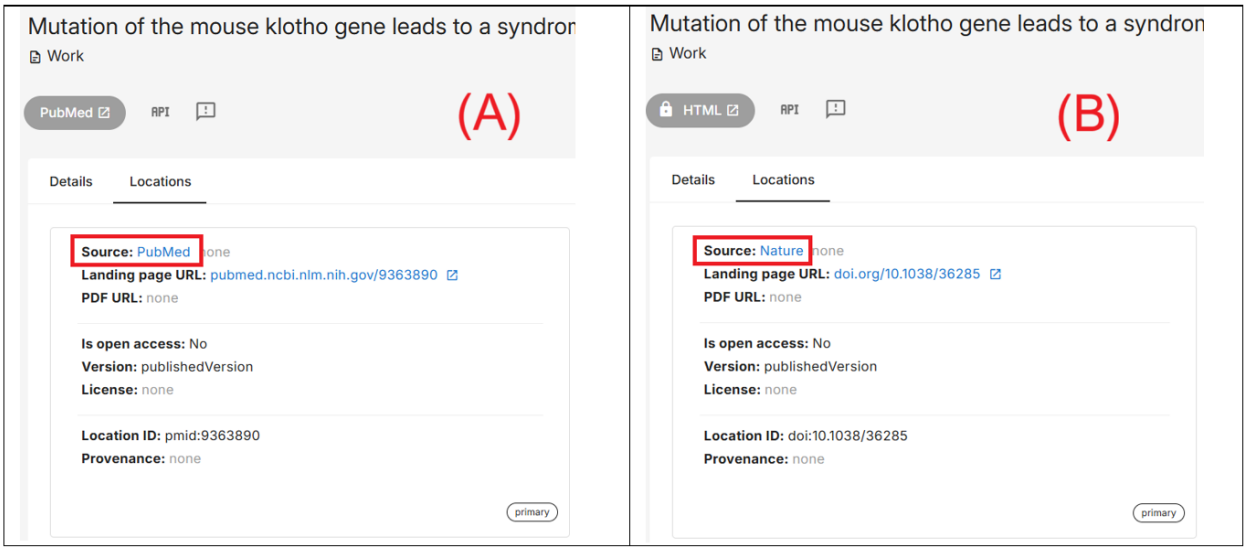


图 6：同一文献的不同版本被识别为不同记录

PubMed 记录的另一个显著之处在于，OpenAlex 将其标记为拥有 66,550 条引文，而*Nature* 记录仅显示 3,729 条，这一数字与该期刊网站上标注的 3,473 条引文相近。[37]

3.5 主题和关键词

OpenAlex 使用自动化系统，依据标题、摘要、来源和参考文献为文献标注多个主题。得分最高的主题即为文献的主要主题。该主要主题又归属于一个子领域、领域和域，可据此按从最具体到最一般的粒度对文献进行分类。

[37]https://doi.org/10.1038/36285

这种分类体系以及为每篇文献指定单一主题的做法，限制了主题检索。例如，在法律这样的领域中，研究范围涵盖众多活动的监管方面，通常会将活动本身识别为主题，而忽略文献的法律性质。图7所示的文章便是如此：尽管文章讨论的是中国对虐待动物行为的监管，其主题分类却未提及法律。[38]

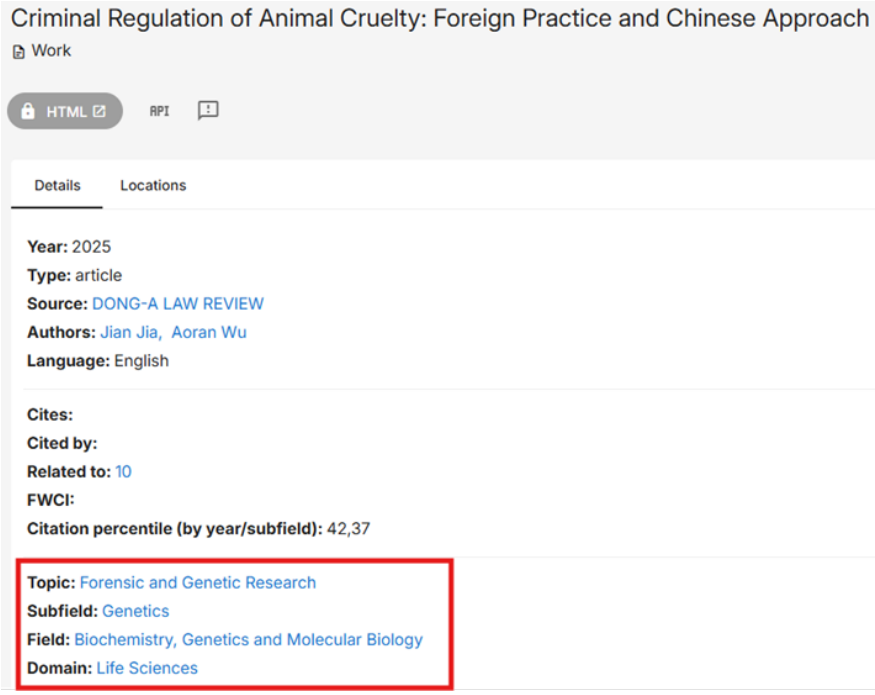


图7： 一篇关于虐待动物监管的文章的主题分类

OpenAlex 根据主题为文献分配关键词。虽然这些关键词可作为可搜索字段使用，但它们不会显示在记录中。

4 结论

OpenAlex 是一个免费开放的数据库，被视为科学网和 Scopus 等商业平台的替代选择。在科学评价体系因推广有助于提高评价过程透明度的开放源而发生转变的背景下，人们对其使用的兴趣日益增长。

尽管 OpenAlex 纳入了与商业数据库类似的功能，但其元数据质量存在显著局限性。最突出的问题包括大量记录缺少摘要、隶属机构和参考文献，以及文献类型和语言识别方面的不足。此外，作者和机构的规范控制以及文献版本管理也存在问题，导致出现重复记录。

这些结果表明，使用 OpenAlex 时需要谨慎。该数据库已用于莱顿排名的开放版等产品，其结果与基于科学网的版本所得结果相对一致。然而，编制这些排名时会进行特定的清洗流程，例如排除未以英语发表或未刊于国内期刊的文献（van Eck 和 Waltman，2024年）；这会将分析范围限制在一组于科学网中收录情况较好、同时通常向 Crossref 及其他为 OpenAlex 提供数据的来源提供更完整元数据的来源之内。

与商业数据库所提供的文献相比，OpenAlex 能够提供覆盖范围广得多的文献的引文信息，这是它在巴塞罗那宣言所倡导的科研评价开放来源讨论中发挥重要作用的原因之一。OpenAlex 对来源文献的覆盖范围更广、可免费访问，且提供可不受限制地重复使用的开放数据，这些特点将促使未来几年基于该数据库的文献计量研究和评价报告不断增加

[38] https://openalex.org/works/w4414088320

。然而，我们仍需更清楚地了解其引文计数的可靠性；由于目前缺乏系统性研究，有理由认为此处指出的案例可能反映了一种普遍现象。例如，图 6 显示，根据 OpenAlex 所采用的来源不同，同一文献的引文数存在显著差异。

提升数据质量最有效的策略，是加强一手来源的完整性，尤其是 Crossref。不过，这需要出版商积极投入。对此，Kramer 等人（2025年）的一项研究指出，出版商在 Crossref 注册元数据的能力，在很大程度上取决于其在自身生产工作流中捕获和维护这些元数据的能力。稿件管理系统似乎在决定哪些信息能够有效提供给 Crossref 方面发挥着重要作用。不过，即使出版商使用相同的系统，其间仍可观察到显著差异，这表明技术并非唯一的决定因素，商业考量也发挥着作用。

改善元数据的另一种补充方式，是由机构用户补充完善。不过，鉴于目前没有明确定义的用户群体，也没有既定的协调机制，这一方案能否以去中心化方式实施仍有待观察。许多机构已签署声明，支持基于开放来源的、更透明的评价流程，但重要的是要记住，这些举措也会产生成本，并需要持续投入资源。

一个目标如 OpenAlex 所述的项目，需要配备与其所承担挑战相匹配的资源和流程。目前，OpenAlex 的人工团队似乎规模很小；其基于自动化和算法化文献处理的流程仍有很大的改进空间，上文关于作者隶属关系分配的讨论便证明了这一点。近几个月公布的响应时间和未解决问题积压情况，以及为增加记录数量而不断扩充新来源的做法，都表明所记录信息的质量控制面临重大风险。

参考文献

Bach, C., Bourguignon, L., Guélé, C., Houdry, P. 和 Kremer, A.（2025）。使用 OpenAlex 的即用型文献计量学：经验回顾。https://cnrs.hal.science/hal-05003502v1。Besançon, L., Cabanac, G., Labbé, C. 和 Magazinov, A.（2024）。暗中掺入的参考文献：伪造的参考文献元数据扭曲引文计数。《信息科学与技术协会会刊》，75(12):1368–1379。https://www.doi.org/10.1002/asi.24896。Bordignon, F.（2024）。OpenAlex 是文献计量学／文献计量学家的变革性机遇，还是一项挑战？UDICE 开放科学工作组。https://enpc.hal.science/hal-04520837。Cao, Z., Zhang, L., Huang, Y. 和 Haunschild, R.（2025）。学界如何引用开放研究信息数据源？基于 OpenAlex 和微软学术系列的综述研究。*Scientometrics*，第 1–27 页。https://www.doi.org/10.1007/s11192-025-05347-6。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. 和 Larivière, V.（2025）。评估 OpenAlex 的语言覆盖范围：元数据准确性与完整性评估。《信息科学与技术协会会刊》，76(6):884–895。https://www.doi.org/10.1002/asi.24979。Chawla, D. S.（2021）。微软学术图谱即将停用。接下来会怎样？《自然》指数。https: // www.nature.com/nature-index/news/microsoft-academic-graph-discontinued-whats-next。Chawla, D. S.（2022）。大型开放学术论文索引上线。*Nature*。https://www.doi.org/10.1038/d41586-022-00138-y。Codina, L.（2024）。OpenAlex：一个有望改进的数据库。https://www.lluiscodina.co m/openalex-scopus/。Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P. 和 Mayr, P.（2025）。与科学网和 Scopus 相比的 OpenAlex 参考文献覆盖分析。*Scientometrics*，130(4):2475–2492。https: //www.doi.org/10.1007/s11192-025-05293-3。Jeangirard, E.（2024）。使用 Works Magnet 加速元数据整理。YouTube。https://www.youtube.com/ watch?v= HivHHCfqXZw。Kramer, B., de Jonge, H. 和 Korzec, K.（2025）。期刊投稿与托管系统如何影响 Crossref 开放元数据的水平。载于*Septentrio* 会议系列，第 2 期。https://www.doi.org/10.7557/5.8317。

Mongeon, P., Hare, M., Riddle, P., Wilson, S., Krause, G., Marjoram, R. 和 Toupin, R.（2025）。调查 OpenAlex 与科学网在文献类型、语言、出版年份和作者数量方面的差异。https://www.doi.org/10.48550/arXiv.2508.18620。Mongeon, P. 和 Paul-Hus, A.（2016）。科学网和 Scopus 的期刊覆盖范围：比较分析。*Scientometrics*，106(1):213–228。https://www.doi.org/10.1007/s11192-015-1765-5。Orduña-Malea, E., Martín-Martín, A., Ayllón, J. M. 和 Delgado López-Cózar, E.（2016）。谷歌学术革命：揭开学术潘多拉魔盒。格拉纳达大学出版社；西班牙大学出版社联盟。Priem, J., Piwowar, H. 和 Orr, R.（2022）。OpenAlex：全面开放的学术成果、作者、出版场所、机构和概念索引。https://www.doi.org/10.48550/arXiv.2205.01833。Scheidsteger, T., Haunschild, R. 和 Bornmann, L.（2025）。基于德国大型大学的实证比较：OpenAlex 与三个热门商业数据库得出的领域归一化引文影响力得分有多相似？*Scientometrics*，130(7):3537–3569。https://www.doi.org/10.1007/s11192-025-05338-7。Thelwall, M. 和 Jiang, X.（2025）。OpenAlex 是否适用于科研质量评估？哪种引文指标最佳？《信息科学与技术协会会刊》，76(12):1660–1681。https://www.doi. org/10.1002/asi.70020。Urbano, C.（2016）。Eugene Garfield：书目控制的创新者和有使命感的企业家。*BiD*：图书馆学与文献学大学文本，（37）。https://www.doi.org/10.1344/BiD2016.37.4。van Eck, N. J. 和 Waltman, L.（2024）。在 OpenAlex 中识别核心来源和核心出版物的方法。https://www.doi.org/10.5281/zenodo.13879947。van Eck, N. J. 和 Waltman, L.（2025）。Crossref 作为开放书目元数据来源。https://www.doi.org/10.31222/osf.io/smxe5_v2。Zhang, L., Cao, Z., Shang, Y., Sivertsen, G. 和 Huang, Y.（2024）。OpenAlex 中缺失的机构：可能原因、影响及解决方案。*Scientometrics*，129(10):5869–5891。https://www.doi.org/10.1007/s11192-023 -04923-y。