

内容提要 - STM 报告（2018 年第 5 卷）

学术交流与科学、技术和医学出版

1. 科学、技术和医学（以下简称 STM）出版是学术交流的一种。学术交流有正式的、有非正式的，在科学研究的不同阶段扮演不同的角色，并随技术驱动而演变。学术交流的模式，可按一对一、一对多或多对多分类，也可按口头或书面分类，以便分析技术对学术交流的可能影响。
2. 期刊是学术交流的核心环节，也是科研的有机组成部分。期刊不仅传播信息，也登记作者首发、通过同行评议来维护质量、建立固定归档版本以供将来引用。期刊还为科研人员提供在日益增长的出版物中的导航手段。

STM 市场

3. 2017 年全球英语 STM 期刊的销售总额大约为 100 亿美元，更宽泛的 STM 信息出版市场规模约 257 亿美元。STM 销售额（含非期刊）的全球分布大致为：美国 41%，欧洲和中东 27%，亚太 26%，世界其他地方 6%。
4. 全球全行业约 11 万雇员，40%在欧盟地区。除行业供应链上的直接雇员，STM 出版业还间接支持了大约 2 万到 3 万人的就业。
5. 本报告主要针对期刊，也关注正在向数字出版转型的 STM 图书市场（年价值约 33 亿美元）。2016 年电子书占到市场的三分之一，近年的增长大大快于整个 STM 出版市场。
6. 全球期刊出版机构的总数约 1 万家，其中约 5000 家为 Scopus 数据库收录。期刊出版机构的主要的英语的行业和专业组织，涵盖了约 650 家出版机构及其出版的 11,550 种期刊，占期刊种数总量约一半。其中大约 480 家出版机构（73%）和 2,300 种期刊（20%）是非盈利性的。
7. 在 2018 年年中，全球大约出版 33,100 种同行评议的英文期刊（此外有 9,400 种非英文期刊），年出版约 300 万篇论文。两个多世纪以来，年出版的论文数和期刊种数分别稳定年增长约 3%和 3.5%。而近年的论文数增长加速到 4%，期刊种数增长加速到 5%。其原因是研发支出和人员的增长。科研人员总数目前估计在 700 万到 800 万之间（按不同的统计口径），其中约 20% 属于多次发表论文的作者。
8. 中国已超过美国成为论文产出最多的国家，论文篇数占到约 19%，而按目前的趋势，其研发投入也将在 2020 年代早期超过美国。美国论文篇数占 18%，而印度近年也快速增长，占到 5%，超过各占 4% 的德国、英国和日本。

商业模式和出版成本

9. 聚集在供给和需求侧都是常态，期刊打包售卖，而图书馆组团采购。电子书领域也出现类似的模式。
10. 虽然大交易（Big Deal）模式及打包折扣价的做法有助于让研究人员访问更多的期刊和降低获取论文的篇均费用，打包模式还是面临图书馆员的压力，他们希望方案更灵活可控，定价更合理和确实更低。近年欧洲有一些引人注目的采购谈判僵局，少量机构恢复到逐刊采购的模式。尽管如此，大交易模式仍有足够的优点以持续存在，虽然可能会在范围（如捆绑或抵销开放获取的费用）和定价模式上变化。
11. 科研人员可访问的学术内容处在历史高点，内容打包和联合采购模式继续带来前所未有的获取水平，每年的全文下载估计超过 25 亿次，而平摊到每次下载的费用则达到历史低点（对很多大客户来说，每篇文章不到 1 美元）。各种调研显示，科研人员就获取期刊的评分为“好”或“很好”，并承认在改善。而同样是这些科研人员，当被问到最需要改进的方面时，还仍然把期刊论文获取放在首位。看来相对过去来说的“很好”，却仍然可能不够满足当下的需要，当内容更容易检索到（比如通过搜索引擎），而科研人员无法立刻获得检索结果的全文时，他们会有受挫感。
12. 关于获取内容的障碍，除了常被提及的成本和价格，还有：不了解有些什么内容资源、繁琐的购买流程、对数字出版物加征增值税、文件格式和信息技术方面的问题、缺少图书馆会员身份、作者和出版机构的权利之争以及如何更好地使用内容。
13. Research4Life 项目对发展中国家的科研人员开通免费的或者非常低价的访问。该项目在继续扩大，增加提供的内容，让更多的注册机构和用户参与。
14. 论文的篇平均出版成本相差甚大，取决于退稿率（导致同行评议成本）、内容量和类型、编辑加工程度等。CEPA 估计在 2010 年，出版一篇按订阅模式的，有印刷版和电子版的期刊论文，出版成本约 3,095 英镑（或 4,000 美元，不含非现金支出的同行评议成本）。CEPA 在 2018 年的新的分析显示，类似编辑活动这样的无形成本，要远远高于生产、发行等环节的有形成本，成为单篇成本的主要因素。
15. 围绕用技术和开放获取来降低成本的讨论已有很多，开放获取出版社 Hindawi 和 PeerJ 曾声称篇均成本可以降低至几百美元。而近期 PLOS 篇均成本升至 1,500 美元（从其财务报表推算），以及开放获取期刊 eLife 高达 3,000 英镑（4,000 美元）的篇均成本，让人们怀疑开放获取模式是否真能大幅降低成本。而不管怎样说，当文章总量年增长 4%，而期刊销售额年增长仅 2%，降低篇均成本的压力不可避免。

科研行为和动机

16. 虽然期刊出版方式在转型，科研人员发表论文的核心动机大体没什么变化，仍然聚焦在获得经费资助和职位晋升。
17. 科学界仍把同行评议视作学术交流的根本，即使认识到其存在不足。通常一位审稿人花 5 个小时审一篇文章，每年大约审 8 篇文章。同行评议机制面临研究论文产出增加的压力，特别是来自发展中国家的投稿的激增。投稿来源和审稿人来源暂时有失平衡，美国的科研人员贡献了全球三分之一的审稿，同时投稿只占全球的四分之一。
18. 同行评议有很多创新做法，其中渐进的创新方式比激进的更获支持。各种各样的公开评议（例如在出版前或出版后公开审稿人的姓名，或将审稿意见和论文一同刊登）正在得到出版社和资助机构的支持，虽然有证据显示这会降低审稿人接受审稿邀请的几率。“级联审稿”（**Cascade review**；把稿件和审稿意见一起转给另一期刊）方式已站稳脚跟，而“PLOS ONE”这样的巨型开放获取期刊采用的“合理但不要求重大意义”的审稿原则已趋成熟。独立于期刊（便携式）的审稿没有得到热捧，而出版后评议模式得到的支持也仍然有限，如最近“PubMed Commons”的终止。
19. 科研人员平均的年论文阅读篇数增长了几十年后，现在不再增长。目前人均每年阅读 250 篇论文，随学科而不同（医学和自然科学较多，人文社科较少），并有迹象显示总量可能下降。每篇论文的平均阅读时间的减少趋势，从 90 年代中期的 45~50 分钟到 2012 年每篇 30 分钟多一点，也可能在逆转。论文的查找和获取更多依赖搜索而非浏览，但研究者仍然继续通过多种途径来查找内容。社交媒体传播在各个学科都变得更为重要，但看来也已达峰值。平均来讲，研究人员花在出版机构的网站时间很短，常常是蹦进蹦出，直接获取要用的论文后就离开。
20. 期刊影响因子的不足还在被持续讨论，但在科学家们的眼里，那些新出现的可选用的计量指标还未能取代影响因子。越来越多的关注指向如何跟踪和显示科研成果的更广泛的经济社会影响，科研资助机构在对此提高期望。
21. 持续引发关注的还有科研和出版伦理，表现在类似出版伦理委员会（**Committee on Publication Ethics, COPE**）这样的机构变得更举足轻重，以及反剽窃技术的进步。在过去十年里，被撤回的期刊论文大大增多，而人们一致的观点是，与其说现在有更多的违规行为，不如说是现在人们的相关意识提高了。

开放获取 (OA)

22. 一些新的商业模式在市场上扎根，期刊出版行业变得更多样化，也更竞争激烈。OA 让人们从网上免费获取原始研究的成果，也没有太多的版权和许可上的限制。OA 有 3 种做法：开放获取出版（“金色 OA”，包括全部或者混合的开放获取期刊），延迟开放获取，和作者自存档（“绿色 OA”）。
23. 开放获取期刊目录（Directory of Open Access Journals）列出了大约 11,811 种（其中 9,172 种是英文的）完全 OA 的期刊。在 Scopus 或 Web of Science 这样的选择性的文摘索引数据库里，OA 期刊与其他期刊相比被收录的几率仍多少小些，这可部分归因于 OA 期刊的历史通常较短，规模也较小（除了一些特例）。这样一来，每年出版的总共 300 万篇论文中，OA 的论文数占比要远小于 OA 期刊种数的占比。
24. 在 2016 年，大约三分之一的学术文献来自合法的和可持续的开放获取资源。最新的估计是 OA 期刊发表的论文数占论文总数的 15~20%（而 OA 期刊种数占比为 26~29%），另有占比 10~15% 的论文通过出版商的网站延迟 OA 或通过自存档 OA。
25. 金色 OA 有时被当作是收取论文处理费（APC, article processing charge）的商业模式的同义词，但严格来说它仅指文章一出版就立即开放获取的期刊。Scopus 数据库收录的相当部分的金色 OA 期刊并不涉及论文处理费，而是按其他模式（如机构的支持或赞助）。论文处理费模式本身也日益复杂，有各种收费办法（比如按文章篇幅）、折扣、预付费和机构会员制、混合出版物的抵扣和捆绑销售安排、阅读-发表一揽子协议等。
26. 基于论文处理费的金色 OA 有很多潜在的优点，在某些领域得到了很多支持。该模式可随科研产出的增长而增长，有在系统范围内降低成本的潜力，论文再用得以简化。在某些地区，资助机构会报销出版费，但就算通常有资助机构的支持，大学里的费用支出实施细则仍有待理顺。OA 出版费用最终会由谁买单还不清楚：目前该费用低于论文出版的历史平均成本；完全 OA 论文的出版费仍低于混合 OA，虽然差距在缩小。将订阅支出转向 OA 的呼吁在增加，但研究密集的国家和大学仍担心对预算的影响。
27. OA 出版带来了一种新型期刊的出现，即所谓的巨型期刊。以 PLOS ONE 为例，巨型期刊有三个特点：完全开放获取和较低的出版费；基于“合理但不要求重大意义”的、快速而不精挑细选的同行评议（考察该科研工作的合理性，而不考虑其是否有重大影响、对特定领域的重要性及相关度这样的更主观的指标，来评审稿件是否录用发表）；非常宽的学科范围。巨型期刊出版的文章数量在持续增长，但增速比以前慢。Scientific Reports（科学报告）取代了 PLOS ONE 成为巨型期刊的代表。
28. 资助机构在学术交流中扮演了日益重要的角色。他们强调责任和传播，要求衡量和提高投入产出。这些因素促使他们支持和要求开放获取（以及相关的、争议较小的数据共享）。从最近的发展来看，资助方和政策制定者越来越有意干预 STM 市场，无论是通过自建出版平台、强化 OA 要求，还是调整激励机制以影响作者投稿时选择期刊。
29. 对绿色 OA 和机构仓储的负作用的担忧还在，但有所减轻。绿色 OA 缺乏独立的可持续的商业模式，意味着该模式的存在，依赖于其不损害（订阅模式的）期刊。两方面的观点都有证据支持，一边说绿色 OA 可以增加下载和引用，另一边说期刊论文有长的使用半衰期和学科间有巨大差异。实践中，更多关注已转向社交协作网络（SCN）和盗版网站对订阅的潜在破坏性影响上。

技术

30. 几乎所有的 STM 期刊都上网了，就专业期刊而言，绝大多数的文献利用都是基于电子版的。有些一般期刊还在同时提供纸质版，包括学协会会员期刊和某些领域（如人文领域和某些执业领域）的期刊。已发展成熟的研究领域（如非执业领域）的期刊纷纷停印纸版，或者转向按需印刷。
31. 开放科学，有时被称为开放学术以便涵盖人文领域，已经取代并包含开放获取作为学术交流转型的总原则。对此尚没有令人满意的普遍接受的定义，因为在利益相关者是谁的问题上几乎没有共识。由联合国教科文组织资助的 OSI 把出版商视作利益相关者，而一些基于图书馆的组织却不。所有定义都强调，整个研究周期的开放性使开放科学比 OA 更重要。
32. 社交网络和社交媒体在学术交流中的影响还不及在一般消费者网络中那么大。考虑到包括时间在内的实际使用成本，若没有十分明确的益处，更多使用可能得不偿失。另外也因为质量和信任：科研人员对未经同行评议、成果权属不清的学术交流方式仍持小心谨慎的态度。虽有这些挑战，Twitter 的使用很普遍，虽然科学家仍把电子邮件作为主流的通信方式。在社交协作网络（SCN）方面，ResearchGate 是首选站点，但科研人员更多是在维护其科研简历，而不是在通信。出版界在说服 ResearchGate 当没有一般使用权就不要提供全文时取得过一些成功，但其很多内容仍然是非法发布和托管的。
33. 移动设备（智能手机和平板电脑）快速而广泛的应用尚未显著改变大多数研究人员与期刊内容的交互方式：2014 年来自移动设备的访问在大多数 STM 平台的流量占比不到 10%（在临床医学等某些领域该比率明显较高），但这种情况在改变。医疗领域专业人员最快采纳此类服务，通常是为获取摘要信息、参考手册或学习材料，而不是刊载原始研究论文的期刊。然而，对于大多数研究人员而言，“真正的工作”仍是在笔记本电脑或个人电脑上完成。现在，大多数的主要出版商都将其大部分期刊（如果不是全部）为移动设备的访问做了优化。
34. 数据密集型研究的爆发性增长，促使出版机构要解决出版物与科研数据双向关联的问题，以支持数据挖掘和管理有可能作为出版物单元的数据集。在研究数据联盟的引领和协调下，这方面有快速的发展：大多数研究资助机构已建立或强化了有关数据存储或共享的政策；数据仓储在数量和类型方面均呈现增长（包括对数据孤岛的仓储）；发起了 Scholix 行动，以系统地制订哪些数据支持文献，哪些文献引用数据。出版机构也同时响应和密切配合这些学术共同体引领的项目，开发关于期刊的标准的数据保存和共享政策，推出数据引用政策，在出版中链接或包含数据，创办先驱性的数据期刊和服务，和开发类似科睿唯安的数据引文索引的数据发现服务。
35. 文本和数据挖掘继续在生命科学行业从细分领域的应用开始扩散，这将有可能改变科学家利用文献的方式。最近欧洲层面的讨论聚焦在版权和许可管理上，其推广仍受制于一系列挑战，包括内容聚合、挖掘的成本和技术技能要求、缺乏对科研人员的激励。虽然已有若干进展，如 CrossRef 的 TDM 工具、PLSclear、版权清算中心为 TDM 集成论文内容的服务，许可授权的框架（如 STM 标准许可条款）和程序（权利获准）、TDM 的内容获取和集成，但这个领域仍不成熟。
36. 资助方和科研机构对绩效评估越来越重视，相应的，从元数据（如下载量、被引数）来对科研活动进行分析的信息服务，以及像 CRIS 工具（Current Research Information System 当前科研信息系统）这样的新的软件服务呈现增长态势。近年出现了一批新的网络规模的发现和分析工具，有望提升我们对科研投入、活动和产出的关系的理解。

37. 人工智能的发展和在各行业的应用，总体上也影响到出版业，部分原因是因为它可以提高成本效率。人工智能囊括了语义技术的大部，被视作引领性原则。有关区块链的谈论很多，尚未在实践中证明有什么用处。出版商一直在提供同行评审和编辑加工等服务，但争取作者的竞争加剧，科研全球化和新的赋能技术在推动作者服务的扩展，并更加注重提升作者的体验。在线协同写作还不能宣告成功，虽然最近对 **Annotations**（标注）的强调让人们看到希望。
38. 也许学术基础设施的最大变化是预印服务器的开发，以及预印本在生物和化学等领域得到越来越多的使用，这些领域此前对此几乎没有兴趣。期刊现在一般不把预印本作为在先出版。有些人担心预印本（可以版本更新）可能成为归档记录版本的选择，从而削弱出版商的商业模式。