

# 开放融合环境下预印本发展态势分析\*

汪庆 任慧玲

(中国医学科学院医学信息研究所, 北京 100005)

**摘要:** 本文利用Dimension平台分析预印本文献数量、主流的预印本平台及其特点; 通过文献调研法和网络调研法了解预印本平台的著作权政策和期刊的预印本政策。研究发现, 预印本文献年增长率远大于普通出版物。新型冠状病毒肺炎疫情的暴发促进了预印本文献的激增。arXiv和SSRN为两大主流的预印本平台。目前主流预印本平台均未要求作者转让著作权, 大部分出版商和学协会均接受预印本投稿且纷纷布局预印本。最后提出应推动中国预印本平台的发展, 将预印本平台纳入资源建设范畴, 利用预印本开展增值服务的建议。

**关键词:** 预印本; 学术交流; 开放获取; 资源建设

**中图分类号:** G230; G250 **DOI:** 10.3772/j.issn.1673-2286.2021.08.009

**引文格式:** 汪庆, 任慧玲. 开放融合环境下预印本发展态势分析[J]. 数字图书馆论坛, 2021 (8) : 59-64.

传统的商业出版模式完全基于订阅, 科学家需要付费才能获取研究成果, 这种模式阻碍了学术研究成果的开放共享。开放获取运动的兴起虽然极大地推动了学术研究成果的开放共享, 但其论文处理费(Article Processing Charge, APC)对于缺乏资金的作者来说仍是一种负担, 因此这种模式并不完美<sup>[1]</sup>。在此情形下, 预印本的出现开辟了科学出版的开放获取新模式<sup>[2]</sup>, 开放获取中一种新的学术交流形式开始出现。预印本是作者未向期刊投稿之前, 未经同行评议的文章手稿<sup>[3-4]</sup>。自1991年第一个预印本平台arXiv发布以来, 预印本已经历了30年的发展。据不完全统计, 目前已经出现了60多个可用的预印本平台, 几乎覆盖各个学科领域<sup>[5]</sup>。随着新型冠状病毒肺炎(COVID-19)的大流行和公共卫生危机的出现, 快速发布、可免费获取的预印本对于了解和评估有关新型冠状病毒肺炎及其影响的新信息至关重要, 预印本平台无出版时滞、传播速度快、避免同行评审偏见等优点进一步凸显。据《自然》杂志2020年相关研究显示, 科学家们经常借助预印本发布与COVID-19和SARS-CoV-2相关的研究<sup>[6]</sup>。

从20世纪中期开始, 期刊一直是学术交流的第一大主要类型。但随着预印本平台的出现, 这种传统的学

术交流生态系统将被打破。传统出版模式下, 文章只有经同行评审后才能被引用; 而预印本一经发布就可以被引用、下载和评论, 且目前主流的预印本平台大都可链接到多个社交网络和工具, 传播范围更广、传播速度更快, 提升了作者的影响力和话语权, 加速了学科研究进程。预印本促进了一个开放融合、公平公开、互动快捷的普惠型学术交流新模式的形成<sup>[7-9]</sup>。如2015年, 科学家Daniel MacArthur发布了其正在研究的预印本文章, 其预印本研究成果一经发布, 10个月就被浏览了300万次。斯坦福科学家Stephen Quake表示如果一份预印本能给两个生物学者提出新的思路, 那么与没有预印本相比, 10年内这个行业科学进展就会快5倍<sup>[2,10-12]</sup>。随着主流出版商对预印本文献的接受和认可, 预印本势必成为未来科学家开展学术交流的重要工具和平台。

## 1 预印本文献总体情况

### 1.1 预印本文献出版数量

为准确地评估预印本的总体出版态势, 笔者通过Dimension平台检索并获取相关数据。Dimensions是

\* 本研究得到国家科技图书文献中心专项任务“数字环境下NSTL资源建设策略研究”(编号: 2020XM32)资助。

Digital Science推出的一个新型的科研搜索引擎,集成了包括PubMed、DOAJ、出版社等在内的主流出版商、数据库和第三方平台的期刊论文、图书和预印本、数据集、专利、基金项目、政策文档等信息,并针对开放获取、预印本等不同类型的文献给以明确的标识。笔者在Dimensions平台上调研发现,截至2021年6月,该平台共收录各类文献信息约1.2亿条,其中预印本文献约占2.3%。

统计分析Dimensions平台上2011—2020年各类文献的发展态势发现,预印本文献在所有文献(包括期

刊、图书、会议)中所占的比例较低,但是占比呈逐年增长态势(见表1)。2011年预印本文献仅10万余篇,在所有文献中占比仅为2.85%;2020年高达37万余篇,是2011年的3倍多,在所有文献中占比开始将近6%。从预印本文献的年增长率来看,2011—2015年预印本的发展较为平稳,年增长率均未超过11%;从2016年开始进入快速发展阶段,预印本文献年增长率从2016年的17.48%提高至2019年的41.37%,这个增长率远高于全部文献的增长率(未到10%)。

表1 Dimension平台上各类文献出版占比

年份	预印本文献/篇	全部出版物/种	预印本文献占比/%	预印本数量年增长率/%	全部文献年增长率/%
2011	103 366	3 631 336	2.85	10.63	2.51
2012	114 349	3 722 465	3.07	5.66	7.39
2013	120 818	3 997 598	3.02	3.98	4.96
2014	125 628	4 195 705	2.99	7.14	2.99
2015	134 602	4 321 193	3.11	9.38	5.12
2016	147 234	4 542 262	3.24	17.48	8.90
2017	172 965	4 946 435	3.50	25.06	7.06
2018	216 314	5 295 662	4.08	22.48	7.75
2019	264 943	5 706 039	4.64	41.37	9.89
2020	374 563	6 270 423	5.97	-	-

## 1.2 主流预印本平台分析

在过去10年中,预印本平台获得快速发展。根据Dimension平台上不同预印本平台收录的文献数量在整个预印本文献中的占比进行排序发现,目前数量最多的预印本平台为arXiv(1 882 535篇,占比70%);其次为SSRN(436 725篇,占比16%)、bioRxiv(123 211篇,占比4.58%)、Research Square(45 332篇,占比1.69%)、OSF Preprint(33 705篇,占比1.25%)、Preprint.org(23 239篇,占比0.86%)、MedRxiv(20 635篇,占比0.77%)、ChemRxiv(19 743篇,占比0.73%)。统计与COVID-19相关的预印本文献发现,与此相关的文献主要分布在MedRxiv(12 674篇,28.91%)、SSRN(8 126篇,占比18.54%)上,二者收录的文献量将近一半;其次为arXiv(4 260篇,占比9.72%)和bioRxiv(4 024篇,占比9.18%)。其中,第一大预印本平台arXiv是1991年康奈尔大学(Cornell University)建立的一个开放获取知识库,也是最早的预印本平台,该平台最初起源于高能物理学,目前已扩展到多个学科领域,包括定量生

物学、计量金融学、统计数据、电气工程与系统科学、经济学等。为更好地推进科学传播的创新,加强与计算和信息科学界的联系,2019年arXiv转由康奈尔大学计算机与信息科学(CIS)运营<sup>[13-14]</sup>。第二大预印本平台SSRN,即社会科学研究网,成立于1994年,于2016年5月被出版巨头Elsevier收购,最初专注于人文科学、人文法律领域,后来业务逐渐扩张至应用科学、健康科学、生命科学、物理科学。两大主流预印本平台成立时间较早,且学科覆盖范围较广。其他的综合性预印本平台还包括Research Square、OSF Preprint、Preprint.org。Research Square是2018年成立的一个覆盖医学、生物学、心理学学科的预印本平台;OSF Preprint也是于2018年由开放科学中心成立的覆盖建筑学、艺术与人文、商业、教育、工程、法律、生命科学、医学与健康科学、物理科学和数学、社会和行为科学的多学科综合性预印本平台;Preprint.org是一个接受全学科文献的预印本平台。

除了这些综合性的预印本平台,其他学科领域也纷纷推出各自的预印本平台,如2013年由美国冷泉港实验

室推出的生命科学领域的预印本平台bioRxiv, 2019年由冷泉港实验室、耶鲁大学和BMJ创立的医学预印本平台MedRxiv。这些预印本平台在发起的时候就得到了业内众多科学家、学协会和出版商的支持, 成为各自专业领域内重要的学术交流平台。一些预印本平台还与出版商有合作, 如Research Square为投稿到Springer Nature的稿件提供“*In Review*”服务。若Research Square上的预印本发表到Springer Nature期刊上以后, 则该预印本平台会提供最终发表版本的链接。bioRxiv和medRxiv也会自动添加预印本手稿最终发表版本的链接<sup>[15]</sup>。

### 1.3 预印本平台的特点分析

(1) 突发事件凸显预印本出版速度快的优势。如在生命科学领域, 从投稿到最终录用大约需要6个月甚至更长的时间。文章在预印本发布后, 到期刊最终录用生成DOI (Digital Object Identifier, 数字对象唯一标识) 的中位时间是166天 (约5个半月)<sup>[5]</sup>。可见, 传统商业模式下, 出版周期较长。相比之下, 预印本文献一般只需要几天的时间就可以完成发布。而一些预印本平台发布速度更快, 如Preprint.org预印本平台声明大部分的预印本24小时内可以发布。因此越来越多的科学家选择在预印本平台上发布COVID-19相关文献。笔者在Dimensions平台利用title/abstract包含关键词“COVID-19”或“SARS-CoV-2”或“coronavirus”检索2019—2020年COVID-19相关文献, 并分析COVID-19预印本文献在所有COVID-19出版物中的占比发现, 2020年相关预印本3万余篇, 在全球COVID-19相关文献中的占比高达17.10% (见表2)。

表2 Dimension平台上各类COVID-19相关文献分布情况

年份	COVID-19 相关预印 本/篇	COVID-19 相关OA出 版物/条	全球COVID-19 相关文献/篇	预印本出版 物占比/%
2019	53	922	1 110	4.77
2020	33 548	166 477	196 233	17.10

(2) 社交媒体和替代计量凸显预印本的开放融合优势。相对于期刊而言, 预印本平台包含的学科主题较为广泛, 且提供相关的替代计量学指标, 包括文摘浏览次数、下载次数、评论次数等。如SSRN提供下载次数、文摘浏览次数以及排名, 还直接链接到社交媒体 (如

Twitter、Facebook) 和替代计量学软件 (如PlumX), 支持一键分享; Preprint.org统计了观看次数、下载次数、评论次数; MedRxiv统计了PDF下载次数、HTML格式全文浏览次数、文摘浏览次数、社交网络分享次数; ChemRxiv统计了观看次数、下载次数。有些还专门对最新预印本文献或者关注度较高的文章列出了专栏, 如Preprint.org预印本平台会列出热度较高的文章和最新发表的文章。

(3) 灵活多样的数据格式满足不同传播途径需求。分析主流的预印本平台文献的格式可以发现, 这些平台都提供PDF格式, 而一些平台 (如bioRxiv) 除了提供PDF格式外, 还支持分享、邮件, 支持使用BibTex、Bookbends、EasyBib、Endnote、Medlars、Mendeley、Papers、RefWorks等多种格式下载。任何人都可以免费下载和阅读预印本平台上的文章, 满足了不同传播途径的需求。

(4) 多维度审核指标保障预印本的出版质量。为控制预印本文献的质量, 主流的预印本平台均会对预印本文献作者的身份、内容的学术性和真实性、格式的完整性等方面进行一定程度的审核<sup>[16]</sup>。如arXiv明确表示如果作者提交的预印本手稿内容缺乏独创性、新颖性或重要性将被拒绝, 格式不正确或者学科与该平台不符的手稿、重复提交的手稿都将被拒绝。如bioRxiv在审核的过程中要求预印本手稿须具有论文基本要素, 必须为PDF或一般文本格式, 不得出现带有攻击性或者违背科学性的内容; SSRN则要求文章需使用学术专业术语, 符合学术性相关要求。据统计, 约2/3的预印本文章最终都发表在同行评议的期刊上<sup>[17]</sup>, 可见预印本文献的质量还是较好的。

## 2 预印本平台的著作权政策

正式投稿的期刊文章一经发表就无法进行再次修改。相较之下, 预印本平台一般允许作者对已上传的手稿再次修改、补充和完善后再次上传, 所以一篇预印本文章也会存在多个版本。一般而言, 预印本手稿首次上传后就会生成一个DOI号, 之后不管文章修改多少次, 均使用同一个DOI以方便同行对该文章的引用。根据《中华人民共和国著作权法实施条例》第二十三条规定, 著作权自作品完成创作之日起产生, 并受著作权法的保护。因此, 对于预印本文献而言无论文章手稿修改多少次, 有多少个版本, 该文章的著作权自论文完成之

日起就已自动产生,对论文的再次修改和补充对著作权均无影响<sup>[18]</sup>。除了特殊法律和相关协议约定之外,预印本的著作权理论上属于作者。通过对主流预印本平台的调研发现,目前均未要求作者将著作权转让给平台,但须与平台签署相关知识共享协议以将相关的论文著作权许可给平台。

作者与预印本平台签订的知识共享许可条款一般是针对单篇预印本文献而言,但针对海量的预印本文献,一般来说大部分平台均支持对这些文章进行文本与数据挖掘,如bioRxiv、ChemRxiv均支持对预印本文章的文本和数据挖掘。

除了著作权许可协议外,一些预印本平台还有专门的元数据许可,元数据一般包括预印本文献的标题、作者、摘要及其他相关信息。如arXiv遵循的知识共享CC0 1.0,即在法律允许的范围内,放弃其在全球范围内根据版权法对作品的所有权利,包括所有相关和邻接权,将作品献给公有领域。因此,第三方机构也可以对这一部分的元数据信息进行再利用。

### 3 期刊预印本政策

笔者调研出版商的预印本政策发现,总体来看目前所有的开放获取出版商均接受预印本文章,绝大部分商业出版商和学协会均接受预印本文章,不同出版商关于预印本文章的政策可以在SHERPA/RoMEO上找到<sup>[19]</sup>。从出版商目前的预印本政策来看,可以分为如下4类。

(1) 无预印本政策限制。大部分的出版商和学协会均接受所有的预印本平台的文章,并无任何特殊的限制条件,如Emerald集团出版物、剑桥大学出版社、冷泉港实验室出版社、Springer Nature、美国微生物学会。

(2) 仅接受非商业性预印本平台的文章。包括美国科学促进会、美国医学物理协会、美国计算机协会、贝尔斯坦研究所、英国医学杂志公司、马萨诸塞州医学学会、芝加哥大学出版社、美国生物化学与分子生物学学会。

(3) 接收所有预印本文章的投稿,但要求经同行评审或编辑修改后的文章版本不得再次上传到预印本平台,或要求投稿的文章需在原有预印本文章的基础上进行明显地修改完善,新增重要的内容,以将预印本文章与正式发表的文章区别开来。如美国癌症研究协会、美国心脏协会、美国精神病学协会出版物、美国胸

科学会、欧洲分子生物学组织出版社。

(4) 明确表示不接收。调研发现,目前仅有两个出版机构(均在医学领域)明确表示不接收预印本临床研究文章,即英国骨与关节外科编辑协会(British Editorial Society of Bone & Joint Surgery)、骨与关节外科杂志公司(the Journal of Bone and Joint Surgery Inc.)。

## 4 预印本发展的启示

通过上述对预印本文献的发展态势、著作权政策以及期刊的预印本政策分析,我们发现预印本目前仅占学术交流系统中的一小部分,但从2016年以后一直呈快速增长态势,其年增长率远远高于普通出版物,势必成为未来学术交流的一种重要平台和工具<sup>[20]</sup>。预印本发展给我们带来的启示如下。

### 4.1 推动中国各学科专业预印本平台的发展

从预印本发展态势来看,预印本平台最先在国外兴起,目前主流的预印本平台均分布在国外。从国内的情形来看,目前少数预印本平台(包括中国科技论文在线、中国预印本服务系统、ChinaXiv),均为包含多学科的综合预印本服务平台,各个学科领域未形成有影响力的专业性预印本服务平台。据SCI数据库统计,2018年世界科技论文总数为206.97万篇,其中中国论文占20.20%,仅次于美国,连续10年排在世界第2位。可见,中国科学家发表的高质量论文很大一部分都分布在国外的期刊上。此种情况下,中国教育部、高校、图书馆、学协会应该联合起来共同推动中国预印本平台的快速发展,以确保重要的学术成果首先在国外的预印本平台上发布。中国医学科学院图书馆作为国家科技图书文献中心的医学中心分馆也应联合中国医学科学院/北京协和医学院、中华医学会等机构共同推动中国生物医学印本平台的建设,确保优质的生物医学论文在国内预印本平台首发,确立生物医学领域科学家学术交流的话语权。

### 4.2 将预印本纳入资源建设范畴

新的学术交流生态系统下,预印本平台的重要性毋庸置疑。加之受疫情冲击,全球经济形势不容乐观。在

此情形下,按照党中央、国务院“过紧日子”的要求,国内图书馆的经费也进一步缩减。如天津大学图书馆就因为经费的大幅缩减,暂停订购大量论文数据库<sup>[21]</sup>。因此,在经费紧张的情况下,图书馆更应该有效地利用预印本文献资源,并将其纳入资源建设的范畴。预印本文献与期刊等正式出版物的主要区别在于其首发权的认定,预印本文献对于推动某一学科领域的学术创新具有重要的意义。因此,图书馆很有必要重点凸显预印本在研究成果首发、学术交流、学术创新方面的价值,一方面将国内外重要的预印本平台整合至本馆的检索系统中,并针对预印本平台开展相关的培训和宣传工作,让更多的学者了解并学会利用预印本平台上丰富的文献信息资源开展教学科研和学术交流工作;另一方面,鼓励和引导更多的专家和学者将自己的最新研究成果发布到国内预印本平台,保证国内优秀研究成果在国内的首发权。通过对国内外主要预印本平台调研发现,与商业数据库不同,预印本平台一般均有PDF格式可供免费下载,可以避免法律侵权的风险,一些预印本平台还提供了批量下载接口,且一些预印本平台的元数据也进行了开放;加之目前国内外有影响力的预印本平台较为分散,因此图书馆可以对预印本平台上的文献进行采集、保存和整合,实现对不同平台预印本文献的一站式检索和集成服务<sup>[1, 22]</sup>。

### 4.3 利用预印本平台开展增值服务

图书馆可以针对主流的预印本平台提供培训、学术新闻快讯等增值服务,帮助教师和科研工作者拓展学术视野,提高学术素养,为其开展科研项目立项、研究论文选题、学术合作、跟踪学科发展动向等提供参考和借鉴。很多预印本平台都注重统计文章的各种替代计量学指标,一些平台还将关注度较高的文章专门列出来,并支持将文章在各种社交媒体上进行分享。因此,图书馆可以针对预印本平台上关注度较高、学术影响力较大的预印本文献开展重点报道、全文推送、社群分享等增值信息服务,也可以针对热门话题(如新冠肺炎相关研究)和前沿学科(如人工智能)开展专题信息服务,让国内的专家和读者了解该领域的最新国际动向。此外,图书馆也可以系统地整理国际上一些期刊的预印本政策并对其进行宣传和公开,让广大学者了解主流的期刊对预印本的接纳和包容态度,并更积极地融入预印本文献的生态系统建设中。

### 参考文献

- [1] 开放获取与开放科学之间的冲突: APCs是问题的关键,预印本是解决方案的关键 [EB/OL]. [2021-06-12]. <https://mp.weixin.qq.com/s/f5U-GbzK9g6zE9Kg02wSaw>.
- [2] 曾建勋. 重视预印本及其系统生态建设 [J]. 数字图书馆论坛, 2020 (10): 1-2.
- [3] ChemRxiv. Chexiv frequently asked questions [EB/OL]. [2021-06-27]. <https://chemrxiv.org/f/faqs#faq-9>.
- [4] 唐桂芬. 预印本发展和研究探析 [J]. 出版与印刷, 2020 (2): 61-69.
- [5] 预印本的第二波浪潮: 出版商如何生存 [EB/OL]. [2021-06-27]. <https://baijiahao.baidu.com/s?id=1651869367347503525&wfr=spider&for=pc>.
- [6] How a torrent of COVID science changed research publishing in seven charts [EB/OL]. [2021-06-27]. <https://www.nature.com/articles/d41586-020-03564-y>.
- [7] 吴家睿. 预印本: 学术交流的“破坏性创新” [J]. 科学, 2019, 71 (2): 37-39.
- [8] 曾建勋. 开放融合环境下NSTL资源建设的发展思考 [J]. 大学图书馆学报, 2020, 38 (6): 63-70.
- [9] BJORK B C, SOLOMON D. The publishing delay in scholarly peer-reviewed journals [J]. Journal of Informetrics, 2013, 7 (4): 914-923.
- [10] 张智雄, 黄金霞, 陈雪飞, 等. 科技预印本库的政策动向与政策挑战 [J]. 中国科学基金, 2019, 33 (3): 219-228.
- [11] Arxiv Looks to the Future With Move to Cornell CIS [EB/OL]. [2021-06-12]. <https://cis.cornell.edu/arxiv-looks-future-move-cornell-cis>.
- [12] Frequently asked questions (FAQ) [EB/OL]. [2021-06-12]. <https://www.biorxiv.org/about/FAQ>.
- [13] 刘静羽, 张智雄, 黄金霞, 等. 预印本服务中的质量控制方法研究 [J]. 数字图书馆论坛, 2017 (10): 15-19.
- [14] ABDILL R J, BLEKHMANN R. Tracking the popularity and outcomes of all bioRxiv preprints [J]. Elife, 2019 (8): 914-923.
- [15] 罗娇, 刘晶晶. 预印本平台著作权政策内容的构建 [J]. 中国科技期刊研究, 2020, 31 (1): 24-31.
- [16] Preprints policy [EB/OL]. [2021-06-12]. [https://pt.abcdef.wiki/wiki/List\\_of\\_academic\\_journals\\_by\\_preprint\\_policy](https://pt.abcdef.wiki/wiki/List_of_academic_journals_by_preprint_policy).
- [17] 国际学术出版商纷纷布局预印本 [EB/OL]. [2021-06-12]. <https://mp.weixin.qq.com/s/oJ6bsVIjHWxPHuMbmBlug>.

- [18] 提早发行科研成果预印本 (Preprints) 的利与弊 [EB/OL]. [2021-06-12]. [https://www.sohu.com/a/219861946\\_100093702](https://www.sohu.com/a/219861946_100093702).
- [19] BERLIN S. If the papers don't come to the journal [J]. EMBO Reports, 2018 (19): 1-21.
- [20] 预印本平台火了, 同行评议已如鸡肋? 详解预印本的前生今世 [EB/OL]. [2021-06-12]. <https://mp.weixin.qq.com/s/azxzP5-HboIEUomxD8PRRg>.
- [21] 因经费大幅缩减, 天津大学暂停订购大量论文数据库: 科研汪太难了 [EB/OL]. [2021-06-12]. [https://www.sohu.com/a/441672966\\_120363032](https://www.sohu.com/a/441672966_120363032).
- [22] 侯志江. arXiv预印本服务研究及启示 [J]. 图书馆工作与研究, 2019 (2): 47-54.

## 作者简介

汪庆, 女, 1988年生, 硕士研究生, 馆员, 研究方向: 资源建设与服务。

任慧玲, 女, 1971年生, 硕士, 研究馆员, 通信作者, 研究方向: 资源建设与服务, E-mail: ren.huiling@imicams.ac.cn。

The Development Trend of Preprints in an Open and Convergent Environment

WANG Qing REN HuiLing

( Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100005, China )

Abstract: This article used the Dimension platform to analyze the number of documents on the preprint platform, mainstream preprint platforms and their characteristics; understood the copyright policy of the preprint platform and the preprint policy of the journal through the literature research method and the network research method. The study found that the annual growth rate of preprinted documents was much higher than that of ordinary publications. The outbreak of the new crown epidemic had contributed to a surge in preprint documents. arXiv and SSRN were the two mainstream preprint platforms. At present, mainstream preprint platforms do not require authors to transfer their copyrights. Most publishers and academic associations accept preprint submissions and have distributed preprints. Finally, it was proposed that the development of China's preprint platform should be promoted, the preprint platform should be included in the scope of resource construction, and the suggestion that preprints should be used to develop value-added services.

Keywords: Preprints; Academic Exchanges; Open Access; Resource Construction

(收稿日期: 2021-06-18)