

COVID-19预印本的影响力分析及其 发展启示*

匡登辉 王丽婷
(南开大学图书馆, 天津 300071)

摘要: 探讨COVID-19预印本的出版情况、传播特点和影响力, 为我国预印本平台建设、转型发展提供建议。以5 639篇COVID-19的预印本文献为主要分析对象, 对预印本平台、出版商、文献发布时间分布、预印本论文的影响力进行分析。研究表明: 随着疫情的蔓延, 预印本文献大幅激增。发布COVID-19文献的预印本平台主要有arXiv、bioRxiv、ChemRxiv、JMR Preprints、Research Square、SSRN Electronic Journal, 预印本文献支持绿色开放访问。预印本文献的引文频次与Altmetric关注分数AAS成低度正相关。针对预印本在科学交流和成果共享中的作用和地位, 提出我国预印本平台的发展建议: 推动分学科预印本平台文献资源的集成, 鼓励科研院所、学会、学术出版商与预印本平台加强协作, 积极推动预印本文献的社会影响力提升, 加强预印本论文质量品控, 加强预印本文献资源的深度开发与利用。

关键词: 预印本; COVID-19; 开放获取; 影响力评价; 知识服务

中图分类号: G353 **DOI:** 10.3772/j.issn.1673-2286.2020.12.007

引文格式: 匡登辉, 王丽婷. COVID-19预印本的影响力分析及其发展启示[J]. 数字图书馆论坛, 2020 (12): 45-50.

预印本是一份研究论文的完整草稿, 在进行同行评审之前与公众分享, 通常会被分配一个数字对象标识符 (Digital Object Identifier, DOI), 以便其他研究论文对其进行引用。这一科学交流形式的出现大大提升了科学交流的效率, 预印本颠覆了研究人员对科学发现和协作分享的方式^[1], 适应了开放、免费、共享的学术交流基本要求, 推动学术出版进入开放科学的时代。自COVID-19 (新型冠状病毒肺炎) 疫情爆发, 科学界积极参与应对这一全球公共卫生挑战, 生产出大量研究成果, 并依托预印本得到了及时传播, 对COVID-19的防控决策和政策制定、患者治疗方案设计起到了积极作用。而在SARS这一2003年发生的全球公共卫生危机中, 据Peiperl^[2]的统计, 疫情结束时仍有90%的论文还在审稿中, 没能及时发表。两相对比, 预印本这一科学交流新方式的价值更加凸显, 由此也引发了相关领域

学者对COVID-19主题预印本文献的研究兴趣。如纽约伊坎医学院精密免疫学研究所 (PrISM) 的研究人员尝试将预印本领域的“信号”与“噪声”分离, 描述其对COVID-19研究工作的贡献^[3]; Torres-Salinas^[4]针对日益增长的COVID-19研究文献, 对比分析数据库与预印本平台, 发现PubMed Central、medRxiv、SSRN (Social Science Research Network) 是预印本领导者, 但期刊仍然是科学传播的核心; Majumder等^[5]则论述了有关COVID-19预印本的传播对全球的影响; 在Wellcome基金会的支持下, Nature启动了关于COVID-19研究的快速科学预审, 快速审核与新出现疫情相关的预印本^[6]。Chahrour等^[7]与Hossain^[8]、张龙浩等^[9]、李士雪等^[10]基于文献分析, 对有关COVID-19的全球研究现状进行梳理和总结; 程冰等^[11]就新型冠状病毒肺炎疫情初期国内外201篇专业文献出版情况进行分析, 发现科学研究的

*本研究得到2020年天津市哲学社会科学规划项目“多源异构数据的学术论文评价体系构建与实践探索研究” (编号: TJTQ20-001) 资助。

专业文献主要发表在预印本文献平台和权威期刊上。

综上,关于COVID-19预印本的相关研究,集中于对预印本文献的重要性及COVID-19全球研究现状的综合等定性研究,完整性、系统性仍有所欠缺。本文拟结合COVID-19科学研究的主题,从多个维度对预印本进行定量研究,全方位评价预印本文献的影响力,以期科学研究的关注度提升,以及我国的预印本平台和科技期刊的建设、转型发展提供建议。

1 数据与方法

Dimensions是Digital Science联合其六大支柱产品Altmetric、Consultancy、Figshare、ReadCube、Symplectic和UberResearch,以及100多家的合作伙伴于2018年1月推出的新型学术数据平台,收录了上亿条文献信息,还提供相关的会议、基金项目、专利以及政策文本等数据^[12];链接了重要的科研资源信息,包括开放存取期刊目录DOAJ、PubMed、巴西科学在线图书馆SciELO、自然系列期刊等;提供发文量、总被引次数、篇均被引频次等常见的重要评价指标,以及学科相对引用率指标FCR、相关引用率RCR、替代计量指标值AAS (Altmetric Attention Score)等优化评价指标^[13],为科学评价研究提供了广泛的数据来源和新的解决方案。

Dimensions提供了新型冠状病毒肺炎疫情爆发以来,有关COVID-19研究的专题文献数据,包括文献的

元数据,以支持全球的科学研究。该数据集于2020年3月10日初次上线,可以通过Dimensions平台和Figshare平台免费获取,每24小时更新一次。本文采用的是2020年5月26日发布的V21 (20200526)版本数据,包含各类文献数据36 833条,预印本数据5 639条,占15.3%;具有AAS值的数据20 175篇,占54.8%。在数据清洗环节,对于缺失来源的文献,结合DOI和题录信息,补充了“Source title”信息。

完成数据采集与清洗后,从COVID-19预印本平台、出版商、论文发表时间分布、成果开放获取、预印本文献的学术和社会影响力等方面进行分析,总结COVID-19预印本的传播特点和出版情况,提出了我国的科技期刊建设、预印本平台的发展路径。

2 结果与分析

2.1 描述性统计

COVID-19预印本论文的引用指标与AAS指标间的差异较为显著,如表1所示。总体来说,引用指标完整性较好,均具有相关数据;具有有效AAS指标数据的预印本仅有1 679篇,占总量的29.8%。其中,引文量非零的预印本论文占比为12.9%,AAS非零的论文占比达29.8%。从指标数值来说,除偏度、峰度外,引文的统计学指标均低于AAS。

表1 COVID-19预印本引用与关注指标描述性统计

指标	有效	缺失	非零值	覆盖率	均值	中位数	标准差	方差	偏度	峰度
Times cited	5 639	0	728	12.9%	0.69	0.00	5.06	25.64	20.14	550.42
AAS	1 679	3 959	500	29.8%	108.14	8.00	665.96	443 500.46	14.21	244.69

2.2 预印本平台类型

数据集中涉及多个预印本平台,包括arXiv、bioRxiv、ChemRxiv、JMR Preprints、Research Square、SSRN Electronic Journal等。论文在平台间的分布较为分散,发表文献最多的是SSRN Electronic Journal,发表文献1 988篇,占35.25%;其次是Research Square,发表文献为993篇,占17.61%;位列第三的是arXiv,发表文献247篇,占4.38%。

从类型来看,发表COVID-19文献的预印本平台既有综合性的预印本平台,如American Journal Experts

(AJE)运营的Research Square;也有面向特定学科领域的预印本平台,如面向生命科学、医学领域的bioRxiv,面向社会科学和人文领域的SSRN Electronic Journal。

2.3 支持预印本文献发表的出版商

根据COVID-19预印本数据集统计结果,支持发表预印本文献的出版商主要有American Chemical Society (ACS)、Cold Spring Harbor Laboratory、Cornell University、Elsevier、JMIR Publications、Research

Square。更进一步地分析出版商与预印本平台间的关系,可以看出两者存在非常密切的关联,这些出版商均是预印本平台的拥有者或支持机构,如表2所示。

表2 预印本平台与出版商关联关系(部分)

预印本平台	关联出版商
arXiv	Cornell University
bioRxiv	Cold Spring Harbor Laboratory
ChemRxiv	American Chemical Society
JMR Preprints	JMIR Publications
Research Square	Research Square
SSRN Electronic Journal	Elsevier

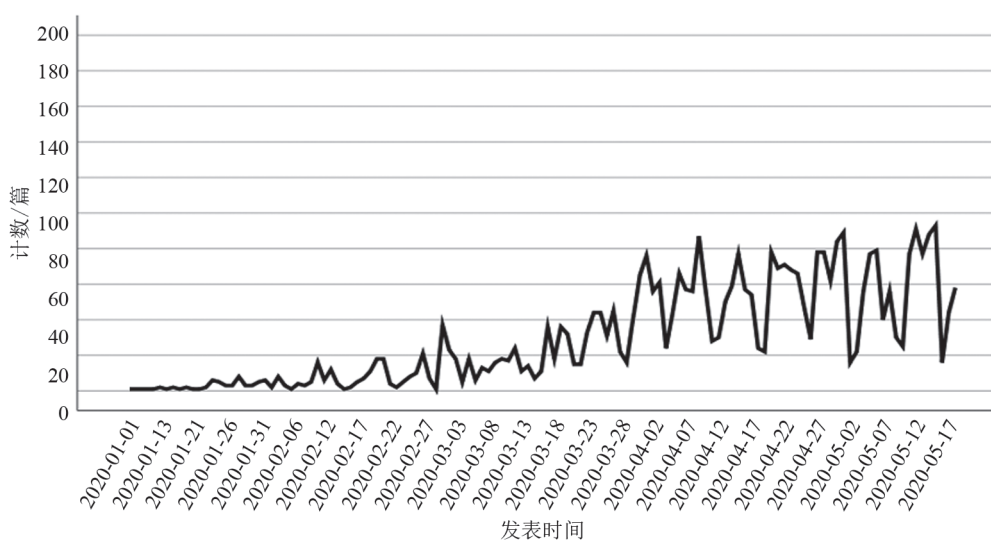


图1 COVID-19预印本文献发表时间分布

2.5 开放获取情况分析

OA是学术界为实现科研成果快速网络传播而采取的行动,预印本作为公开、确凿的数据记录,除确定研究成果的首发权,还可以为疫情防控、临床诊疗、应急处置、科学决策提供重要的依据。预印本文献都是开放获取的,其中65.16%的稿件递交到平台经简单审核后,读者即可自由免费访问,34.82%的稿件发表后即开放获取。

2.6 COVID-19引文频次与AAS相关性分析

相关系数 r 常用来刻画两变量间相关关系的强弱,本文主要分析引文频次与AAS指标间的相关性。如表1

2.4 COVID-19预印本文献发表时间分布

关于COVID-19的预印本文献,疫情爆发的最初几天是零散出现,但自2020年1月26日以后发文数量呈增长的态势,特别是2020年3月18日之后,预印本文献数量增长速度进一步加快(如图1所示),这说明随着疫情的蔓延,越来越多的科学家加入到新冠病毒的研究当中并选择通过预印本渠道公开其研究成果,如COVID-19基因组序列、临床特征、治疗效果评价等,不少重要研究成果都是先发表在bioRxiv或medRxiv预印本系统上,大大加快了知识传播速度。

所示,由于引用和AAS指标的偏度值均大于0,说明样本数据不是来自于正态分布的整体^[14],在相关性分析时需要选择斯皮尔曼相关分析法^[15],检验结果见表3。

表3 不同类型COVID-19文献的引文频次与AAS相关性分析

文献类型	文献数量(N)	占比/%	被引频次VS AAS(r)
出版物	20 175	100	0.399**
论文	18 321	90.81	0.398**
图书	108	0.54	-0.043
章节	32	0.16	-0.103
专论	35	0.17	0.626**
预印本	1 679	8.32	0.447**

注:**在0.01级别(双尾),相关性显著

从表3可知,全部出版物及论文、图书、预印本等各类文献的引文频次与AAS的斯皮尔曼相关系数均为正值,说明引文频次与AAS属于正相关。仅从数值来看,引文频次与AAS相关性最大的是专论,相关性系数达0.626;预印本文献次之,相关性系数为0.447;全部出版物的引文频次与AAS的相关系数为0.399,略高于论文的引文频次与AAS的相关系数0.398,这几项数据均在0.01水平上显著相关。图书或章节的引文频次与AAS的相关系数为负值,表明图书或章节的引文频次的变化趋势相反,即引文频次越高,AAS数值越低。

2.7 高影响力COVID-19预印本论文分析

被引频次排名前10的COVID-19预印本论文中,9篇来自bioRxiv,1篇来自medRxiv;排名居首论文的主题是新型冠状病毒种类鉴定及命名,被引频次高达187。AAS排名前10的论文中,7篇出自bioRxiv,2篇出自SSRN Electronic Journal,1篇出自ChemRxiv;AAS得分居首(14 949)论文的主题是新型冠状病毒刺蛋白中独特的插入物与HIV-1 gp120和Gag蛋白不可思议的相似性,总体来看,专门面向生命科学和医学领域的bioRxiv体现出更强的影响力;然而,被引频次前10和AAS得分前10的论文并不完全一致,排名上也差异较大,如AAS得分居首的论文仅被引16次。

另外,从影响力数据积累情况看,在论文刚发表的较短时间内,社交媒体的关注和论文的使用数据快速积累,数据量在短时间内相当大,远超这一阶段被引频次。因此,社交媒体指标和使用数据在论文刚发表后不久便可以对论文的影响力展开评价,但论文的被引量通常在论文发表2~3年之后达到引用峰值,而且更多体现的是其在学术同行中的学术影响力^[16]。由此可以看出,被引次数测度的是论文的长期影响力,反映学术水平,以AAS为代表的社交媒体指标更多地在于反映论文的短期影响力和社会影响力。

3 讨论及启示

开放获取运动的开展,使得优质的科学成果更多、更快地上传至预印本平台,越来越多的原创性的最新成果发表在开放获取的数字出版平台上,通过社交网络扩散,实现快速传播,我国需要构建自主可控的预印本系统和开放出版的学术生态^[17]。鉴于预印本已成为重

要的开放学术资源,总结分析COVID-19预印本论文,对我国搞好预印本平台建设,具有重要启示。

(1) 推动分学科预印本平台文献资源的集成。研究发现,COVID-19主题的预印本分布在多个生物医学领域的预印本平台(包括bioRxiv、medRxiv、arXiv、PeerJ PrePrints等),以及综合性预印本平台、开放获取网站上(如ChinaXiv、PLOS、BioMed Central、Springer Open等),这充分显示出学者发布论文的途径是多元的。但是,鉴于不同平台的实力差异,对平台文献的内容与形式特征揭示度差异较大;而且文献的分散分布也不利于用户的文献利用。因此,很有必要将这些开放、免费的资源整合起来,根据研究领域的不同,做成分学科的预印本平台文献资源集成系统,实现相关预印本信息的一站式检索、利用,构建科学共同体自治的新型科研成果交流和共享平台,与传统的基于期刊的成果发布方式形成有效互补。

建立学科预印本资源集成,对于出版机构,对接预印本,建立优质的稿源仓储,优先筛选优质稿件,从而提高期刊的影响力;对于科研院所,可以彰显机构的学术影响力,促进机构与业界的学术交流;对于图书馆,可以开放获取免费、优质的学术资源,提高服务能力和资源保障水平;对于基金资助机构,为保护和加快资助的学术成果的广泛传播,体现资助效益。因此,鼓励各类科学研究、学术出版相关机构主动加强与预印本平台的合作,积极探索协同互利机制,形成良好生态。

(2) 鼓励科研院所、学会、学术出版商与预印本平台加强协作。本研究发现,发表COVID-19文献的预印本平台与学术研究、出版机构的关系密切,如SSRN的背后是全球最大的学术出版商Elsevier、Research Square背后是美国著名高校Cornell University,并形成了深度合作、多方共赢机制。

(3) 积极推动预印本文献的社会影响力提升。COVID-19预印本文献的被引频次与AAS相关性分析表明,两者存在正相关关系,相关系数为0.447,为中度相关。因此,预印本文献发表后,通过提升其社会影响力,将有助于提升其学术领域的影响力。基于此,我国的预印本平台建设时,需要主动接入主流社交媒体,如Facebook、LinkedIn、微博、微信等,借助社交媒体,有效拓宽传播渠道,扩大信息受众,增强专业知识的渗透力^[18],提升传播效率,进而形成高社会影响力提升论文学术影响力。高学术影响力吸引更多学者将更多成果通过预印本平台发表的良性循环,进一步提升预印本平

台的影响力。

(4) 加强预印本论文质量品控。本文结果显示, 预印本大多不是研究论文的最终形式, 应利用CrossRef等基础信息提供商, 从而提升论文的显示度和引用规范。据解贺嘉等^[19]对国外预印本平台的研究, 预印本和论文最终发表版本差异较小。刘静羽等^[20]的研究则认为, 当前国际通行做法是由上传者 and 作者对研究内容、论文格式完整性、学术性和真实性作出保证。但笔者认为, 对不同学科的预印本, 其论文品控侧重点不同, 如医学类的论文, 实名注册的作者必须声明伦理审查、临床试验登记、患者知情书、资金来源和利益冲突信息; 医学编辑决定论文的接收与否, 标识“未经同行评议”。

(5) 优化预印本平台用户体验, 加强预印本文献资源的深度开发与利用。预印本作为迅速发展的学术交流方式, 内容可以免费开放获取。用户访问平台后就可以查看、浏览、检索、下载、评论最新科研成果, 形成“投稿→审核→发布→评审→修改→发表”各环节协同联动的出版新模式, 必然要求论文元数据规范准确, 支持跨系统传输、交换、调用和互认; 要求简化系统的内部操作流程, 提供更加符合用户需求的功能, 促进更大范围内的预印本资源开放和共享。此外, 还需要开展针对不同层次的用户需求, 深度开发和有效利用预印本资源, 开展个性化、差异化的用户互动、信息定制服务, 如可以开展基于研究主题或课题组的Alert提醒服务、RSS推送等。对于录用的稿件, 需要进一步加强与期刊的合作, 精准关联论文的不同版本, 对接Dimensions、PubMed、Scopus、Web of Science (WoS) 核心合集等数据平台, 打通期刊、预印本等资源的有机集成、整合和关联, 发布每篇文章的相关数据(如论文被下载、浏览、转发、社交媒体提及、开放评论等), 支持预印本文献的自由检索、获取、分析、探索预印本可持续发展的机制。

4 结语

COVID-19爆发后, 一些科学家在预印本平台推出的创新性、争议性研究成果使得预印本成为全球焦点, 频繁出现在学术新闻和社会新闻中, 预印本已然成为国际科研生态的重要一员, 其不仅能够确认首发权, 而且其透明、多元的评审机制, 开放获取的传播方式也能够加快创新成果的传播速度、提升学术交流的广度。我国应主动融入国际潮流的新型学术交流模式, 借鉴国外

预印本平台建设经验, 提升我国的预印本建设水平, 保障国家科学交流的主导权, 营造开放的、互联的、普惠的学术生态圈, 促进学术交流。

参考文献

- [1] KUPFERSCHMID K. Preprints bring ‘firehose’ of outbreak data [J]. *Science*, 2020, 6481 (367): 963-964.
- [2] PEIPERL L. Preprints in medical research: Progress and principles [J]. *PLoS Medicine*, 2018, 15 (4): e1002563.
- [3] SCHANOSKI A S, RODRIGUEZ A, MOREIRA A, et al. Advancing scientific knowledge in times of pandemics [J]. *Nature Reviews Immunology*, 2020, 20 (6): 338.
- [4] TORRES-SALINAS D. Daily growth rate of scientific production on Covid-19. Analysis in databases and open access repositories [EB/OL]. (2020-04-14) [2020-07-13]. <https://arxiv.org/abs/2004.06721>.
- [5] MAJUMDER M S, MANDL K D. Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility [J]. *The Lancet Global Health*, 2020, 8 (5): e627-e630.
- [6] MICHAEL A, JOHANSSON D S. Fast peer review for COVID-19 preprints [J]. *Nature*, 2020, 579: 29.
- [7] CHAHROUR M, ASSI S, BEJJANI M, et al. A bibliometric analysis of COVID-19 research activity: a call for increased output [J]. *Cureus*, 2020, 12 (3): e7357.
- [8] HOSSAIN M M. Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping [J]. *F1000 Research*, 2020, 9: 374.
- [9] 张龙浩, 李柏宏, 贾鹏, 等. 新型冠状病毒 (SARS-CoV-2) 全球研究现状分析 [J]. *生物医学工程学杂志*, 2020, 37 (2): 236-249.
- [10] 李士雪, 单莹. 新型冠状病毒肺炎研究进展述评 [J]. *山东大学学报 (医学版)*, 2020, 58 (3): 19-25.
- [11] 程冰, 檀博, 孟连生, 等. 新冠肺炎疫情初期国内外专业文献出版情况分析 [J]. *科技与出版*, 2020 (4): 40-45.
- [12] Digital Science. Dimensions: A collaborative approach to enhancing research discovery [EB/OL]. [2020-08-04]. <https://www.digital-science.com/resources/portfolio-reports/dimensions-collaborative-approach-enhancing-research-discovery/>.
- [13] MALEA E O, LÓPEZ-COZAR E D. Dimensions: re-discovering the ecosystem of scientific information [J]. *El*

- Profesional De La Información, 2018, 27 (2) : 420-431.
- [14] 刘江涛, 刘立佳. SPSS数据统计与分析与应用教程 [M]. 北京: 清华大学出版社, 2017: 47.
- [15] 迈克·塞沃尔. 链接分析: 信息科学的研究方法 [M]. 孙建军, 译. 南京: 东南大学出版社, 2009: 34.
- [16] GLÄNZEL W, THIJS B, DEBACKERE K. Productivity, performance, efficiency, impact—What do we measure anyway? [J]. Journal of Informetrics, 2016, 10 (2) : 658-660.
- [17] 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.
- [18] 田杰, 王雅娇, 王佳. 学术期刊在突发公共卫生事件中的社会责任与思考——以新冠肺炎疫情为例 [J]. 中国科技期刊研究, 2020, 31 (3) : 241-247.
- [19] 解贺嘉, 刘筱敏. 国外预印本平台研究述评 [J]. 科技与出版, 2020 (6) : 21-27.
- [20] 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.

作者简介

匡登辉, 男, 1981年生, 硕士, 副研究馆员, 研究方向: 科学计量学、科学评价。
王丽婷, 女, 1983年生, 硕士, 馆员, 通信作者, 研究方向: 文献计量学, E-mail: Wanglitingnk@163.com.

Impact Evaluation and Development Enlightenment of Preprint Based on COVID-19

KUANG DengHui WANG LiTing
(Library of Nankai University, Tianjin 300071, China)

Abstract: This paper discusses the publishing situation, dissemination characteristics, impact evaluation of COVID-19 preprint, and provides suggestions for the future academic emergency publishing service, the construction and transformation of China's preprint platform. Taking 5 639 COVID-19 preprints as the main analysis object, the influence of preprinted paper was analyzed on the preprint platform, publisher, preprinted document release time sequence, and the influence of preprinted paper from both the reference and attention levels. The results show that with the spread of the epidemic, the preprinted papers have increased dramatically. Preprint platforms for the publication of COVID-19 include arXiv, bioRxiv, ChemRxiv, JMR Preprints, Research Square, SSRN Electronic Journal, and preprint support green open access. The citation count of preprint is positively related to the Altmetric Attention Score at a low degree. Aiming at the role and status of preprint in scientific exchange and achievement sharing, this paper puts forward its development proposals: develop the integration of sub-discipline preprint systems; attach importance to cooperation and promotion with scientific research institutes, societies and periodicals, create a win-win situation; use social media to effectively broaden communication channels; strengthen the quality control of preprinted papers; promote the construction of preprint platforms, and strengthen active content disclosure and personalized knowledge services.

Keywords: Preprint; COVID-19; Open Access; Impact Evaluation; Knowledge Services

(收稿日期: 2020-11-07)