

预印本服务中的质量控制方法研究*

刘静羽¹, 张智雄^{1,2}, 黄金霞¹, 彭媛媛^{1,3}

(1.中国科学院文献情报中心, 北京 100190; 2.中国科学院武汉文献情报中心, 武汉 430071;

3.中国科学院大学, 北京 100049)

摘要: 质量控制是预印本服务中的重要环节。当前国际通行的预印本质量控制措施包括“上传者 and 作者的研究身份审核”“论文格式完整性审核”和“论文内容学术性和真实性审核”三方面, 审核方式以人工审核为主、机器审核为辅。中国科学院科技论文预发布平台 (ChinaXiv) 按国际通行模式建立质量控制机制, 包括论文常规审核制、领域专家评审制、论文开放评议制、问题论文撤稿制和非诚信作者名单制等。未来可从优化平台注册和上传流程、提高机器辅助审核能力, 推动建立更加完善的开放同行评议机制, 确保质量控制政策公开透明等方面做出改进。

关键词: 预印本; 质量控制; 预印本服务; 开放获取

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2017.10.003

1 引言

预印本因其发布周期短、传播范围广等优势, 成为开放科学环境下的一种重要学术交流模式。从第一个预印本平台arXiv出现以来, 众多科学团体以及出版社相继推出预印本服务, 服务器中的论文数量迅猛增长, 2017年10月, arXiv上的论文量已经超过130多万篇, 预印本的关注度和学术交流地位不断提高。与此同时, 也应注意到, 一些“非学术”或“伪科学”论文混入预印本平台, 一部分上传者为了快速发布成果, 伪造研究身份或者违背学术道德, 在预印本平台发布虚假、不真实、抄袭的内容。这些问题导致预印本平台的内容质量下降, 预印本服务提供者信誉受到损害, 更重要的是在很大程度上影响科研人员对预印本的信任和使用, 阻碍了预印本的发展。

据Google Scholar统计, 在全球物理学和数学领域中, 被引用量较多的80%的内容都可以预先在arXiv上找到^[1], 即便如此, 仍有相当一部分科研人员对预印本的质量心存疑虑。一方面, 学术界一直将引用未经同行评议的论文(包括预印本)看作禁忌。此外, 一些不成熟的研究结果以及未经审查就在预印本系统发布等行为,

使预印本整体良莠不齐, 进一步阻碍了科研人员使用预印本进行交流。2017年, 根特大学Shawkey在社交网络上提出, 预印本的内容不应和经过同行评议的论文同等对待^[2]。另一方面, 科研资助机构虽然普遍支持预印本发展, 但对目前的预印本质量表示担忧。如NIH在2017年5月的一份声明中提到, 目前阶段性成果的整体质量确实不如经过同行评议的研究成果, 提示研究人员要对预印本的资源进行谨慎判断, 并要求发布阶段性成果的平台须具备严格、透明的关于防止论文抄袭、利益冲突、学术不端行为的政策^[3]。此外, 美国ASAPbio基金会在生物学预印本中心服务库规划中提出, 该库收录的预印本必须符合基本学术标准, 包括作者署名、学术道德规则和法律、社会标准, 必须正确标明基金支持等^[4]。因此, 只有做好预印本服务中的质量控制, 才能打消科研人员对于预印本的疑虑, 真正促进预印本的发展。

目前全球主要的预印本平台普遍认识到质量控制的重要性, 纷纷采取措施阻挡伪科学, 防止错误信息的传播, 本文通过调研6家国际知名预印本服务平台, 总结归纳出当前国际通行的预印本服务质量控制方法; 并以此为基础, 研讨分析中国科学院科技论文预发布平台ChinaXiv的质量控制机制, 提出未来可以改进的方向。

* 本研究得到中国科学院2016年传播项目“中国科学院科技论文预发布平台”(编号: 院1617)资助。

2 国际通行预印本服务质量控制方法

为深入了解当前国际预印本服务的质量控制方法,本文选取arXiv^[5]、bioRxiv^[6]、SSRN^[7]、ChemRxiv^[8]、PeerJPreprint^[9]、MDPI Preprints^[10]6家国际知名预印本服务平台,对其网站政策与质量控制相关内容,以及注册、上传论文流程等进行调研。

调研发现,目前国际主流的预印本服务对平台发布内容进行简单的质量控制,以保证论文达到基本学术要求。其包括作者的研究身份;内容真实、研究过程清晰;正确标注署名等^[11];避免伪科学、虚假等违背学术道德行为的发生,同时避免学科领域与平台主题不符的论文在平台发布。质量控制措施主要包括“上传者和研究者的研究身份审核”“论文格式完整性审核”和“论文内容学术性和真实性审核”三方面。操作层面,目前

各大预印本服务的质量控制以人工审核为主,部分平台同时采用机器辅助的方式。人工审核主要负责对内容的学术性、真实性审核,一般由领域专家承担;机器审核主要是在上传者注册身份识别、论文格式审核等方面发挥作用。如arXiv对每一篇上传的论文进行监控,利用arXiv审核系统对论文进行初步筛选,并采取措施阻止部分作者绕过上述过程而试图再次提交已经被拒的论文。对于进一步的审核,则由平台董事会学科专业委员会推荐的领域专家完成。此外,调研中大部分预印本平台支持读者对已发布内容的开放评议,不仅有助于进一步识别论文的学术性和净化预印本平台内容,也有利于论文质量的提升。目前预印本服务中对论文的质量控制主要采用审核论文上传者和作者的研究身份、审核论文格式完整性和审核论文内容的学术性和真实性三种措施(见表1)。

表 1 国内外知名预印本平台的质量控制内容

	上传者和研究者的研究身份审核	论文格式完整性审核	内容学术性和真实性审核
arXiv	邮箱后缀、所属机构、学科领域中研究学者的认可	论文符合一般学术要求,基本要素完整	学术性、学科关联性;学科领域专家参与论文评审
bioRxiv	未明确说明	具有论文基本要素,必须为PDF或一般文本格式	非剽窃;不得带有攻击性或者违背科学性的内容;未发表
SSRN	未明确说明	未明确说明	文章内容符合学术性要求;使用学术术语
PeerJPreprint	未明确说明	论文类型,字数限制;不同研究方向的论文规定了必备要素	不得发表于其他预印本平台
ChemRxiv	未明确说明	未明确说明	非剽窃;不得带有攻击性或者违背科学性的内容;化学博士参与论文评审
MDPI Preprints	上传论文需要先注册账户,提交前平台确认作者信息	不接受社论或课程作业;具备基本论文要素	所有论文接受MDPI内部审核

2.1 审核论文上传者和作者的研究身份

论文上传者和作者的研究身份审核是预印本内容质量控制的第一步,审核控制点主要包括注册邮箱、机构主页和信息、基金资助情况、同领域研究中的被认可度等。在6家平台中,有2家需要对用户注册和上传论文的进行审核,包括arXiv、MDPI Preprints。arXiv是当前全球规模最大的预印本平台,其对论文上传者的可信度进行严格审核,包括必须使用机构邮箱注册、主动声明科研身份、提交证明,推荐arXiv与ORCID的关联,审核通过的用户才可提交论文^[12]。系统将确认作者的联系方式,并进行邮箱验证,一旦发现伪造机构,将

会被立即且永久禁止提交论文。注册成功后,上传者还必须获得一位同领域科研人员的认可,才能顺利上传论文,通常这位科研人员需要在arXiv有一定的发文量。

2.2 审核论文格式完整性

格式完整性审核的内容主要包括论文基本要素是否齐全,格式是否符合平台需求。调研发现,除SSRN、ChemRxiv暂未获得论文形式审核信息外,其余预印本服务都在论文格式方面进行质量控制。如arXiv要求论文符合一般学术要求,基本要素完整;不接收只有摘要的论文,不带引文的论文、图书或评论,不包含原创、独

立研究的报告。bioRxiv要求上传文章须具备论文基本要素, 格式为PDF或一般文本格式。

2.3 审核论文内容学术性和真实性

内容学术性和真实性审核, 主要指预印本平台对论文学术性、合法性、真实性等方面的控制, 不针对研究内容本身质量进行评价。审核控制点主要为, 要具有学术性, 无剽窃、攻击性、违反法律、恶意攻击、诽谤性、涉密等内容, 无版权纠纷。以arXiv为例, 其平台政策明确说明需要“完全学术性”的内容, 并强调要与18个学科相关联。arXiv不接收带有如下问题的论文: (1) 包含煽动性、虚假内容; (2) 主题不合适或有不如实、歪曲的表达; (3) 论文标题、摘要、导言较富有戏剧性; (4) 用户在短期内提交的论文内容重复; (5) 版权纠纷作品,

要求上传到arXiv的内容必须拥有完全的版权。此外, bioRxiv、SSRN、ChemRxiv等都有类似的质量要求。

从实践效果看, 当前全球各领域预印本平台纷纷出现, 论文数量不断增长, 上述质量控制措施为预印本服务有序健康发展提供有效保障, 是当前国际预印本服务的通行做法。

3 ChinaXiv质量控制机制

ChinaXiv由中国科学院于2016年推出, 面向全国科研人员, 提供自然科学领域的中英文预印本服务。为保障平台内容的学术性、严谨性、权威性, ChinaXiv按国际通行做法建立了质量控制机制, 包括论文常规审核制、领域专家评审制、论文开放评议制、问题论文撤稿制和非诚信作者名单制等(见图1)。

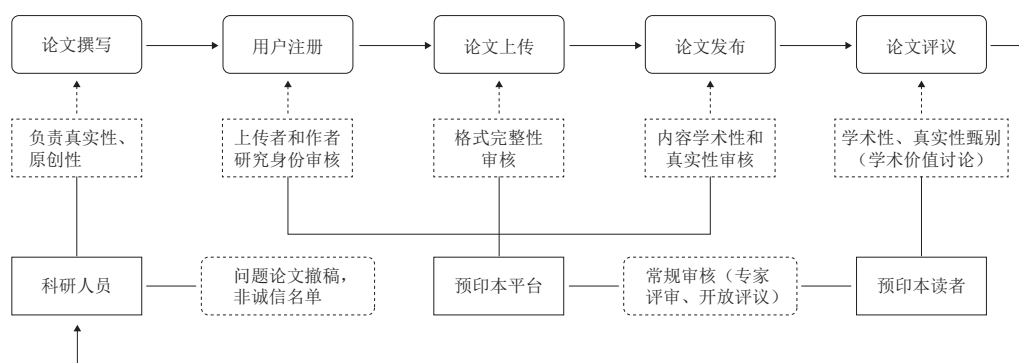


图1 ChinaXiv预印本质量控制机制

3.1 论文常规审核制

ChinaXiv设立质量控制审核组负责每日论文预印本的审核发布, 审核分为初审、终审两级, 目的是最大限度保证论文审核的公平公正。同时, ChinaXiv制定《中国科学院科技论文预发布论文审核工作指南》和《中国科学院科技论文预发布平台论文准入规范》(简称《论文准入规范》), 以保证论文审核工作有序、高效进行。其中, 《中国科学院科技论文预发布论文审核工作指南》主要对论文审核工作的流程和操作进行规范, 《论文准入规范》有效帮助审核人员判断哪些论文可以发布。论文常规审核工作, 主要包括以下三个方面。

(1) 识别论文提交者和作者的研究身份。《论文准入规范》明确要求上传者 and 作者具备研究身份。ChinaXiv建议用户使用机构邮箱注册, 并进行邮箱和手机验证^[13]。注册时建议作者可以选择提交ORCID、机构主页、学术

主页、简要的学习和科研经历, 以及在国内同行评议期刊上发表过的论文等材料。一旦发现论文上传者的研究机构、身份造假或未在正式刊物上发表过同主题的同行评审论文, 那么论文将不能在ChinaXiv发布。

(2) 审核预发布论文的格式完整性、内容是否存在重复或剽窃、要求不涉及版权问题。ChinaXiv系统能够自动检查论文题目、摘要、关键词、参考文献等必备要素的完备性, 以及论文是否与平台已有论文重复、是否涉嫌剽窃。机器判断结果主要用来辅助人工审核。同时, 上传至平台的论文, 作者应保证不涉及版权纠纷。

(3) 甄别预发布论文内容的研究性、科学性。ChinaXiv在平台中明确声明, 作者提交的科技论文应当具有科学性、创新性, 具有一定学术水平和发布价值, 坚决抵制违反学术道德和规范的学术不端行为, 不抄袭、不剽窃、不造假, 不涉及恶意攻击、诽谤、涉密等不良信息和非法内容。同时, 《论文准入规范》明确提出, 对于论

文中有不当言论或者是伪科学的论文,不能在ChinaXiv平台发布。

通过上述三个方面的常规审核,ChinaXiv平台的日常运营和内容质量得到有效保证。

3.2 领域专家评审制

除论文质量管理组外,ChinaXiv还充分发动科研人员力量参与预印本服务质量控制工作,包括ChinaXiv科学咨询委员会成员和中国科学院研究所具有副高级以上职称的科研人员等。当常规审核无法对论文的研究性、科学性作出判断时,则启用专家评审流程,由领域专家向ChinaXiv工作组提供审核意见,然后由ChinaXiv项目组做出论文是否发布的决定。

3.3 论文开放评议制

ChinaXiv遵循“先开放,后评议”原则,所有预印本发布后,将接受来自平台注册用户的评议。平台提供星级评分,而撰写评论包括公开评论、匿名评论、仅发给作者的评论功能。评议内容经审核团队审核后发布。对于开放评议结果中差评较多的论文,将进入再审核,甚至撤稿的流程。

3.4 问题论文撤稿制

ChinaXiv维护作者在平台上的论文发布权及被访问权,同时要求论文作者遵守基本学术道德。一旦发现内容剽窃嫌疑、作者身份造假、研究内容有失科学性的情况,将报请平台负责人对论文内容进行审核,确定是否撤稿并发布相关申明。

3.5 非诚信作者名单制

为杜绝非科技类、伪科学论文在平台发布,ChinaXiv基于审核工作中发现的学术非诚信作者,如被撤稿并被认定为非诚信的作者等,建立非诚信作者名单,并由平台负责人核定。非诚信名单中的作者,其论文不再进入ChinaXiv审核与发布流程。

通过上述机制的建立与实施,ChinaXiv形成以平台为主导,作者、用户、专家多方参与的预印本质量控制模式,实现从论文上传、审核到发布的“全流程”质量控制。

ChinaXiv自2016年6月上线为科研人员提供稳定的预印本服务,截至2017年10月底,平台累计发布4 600多篇论文。

4 ChinaXiv质量控制机制的改进

从当前国际预印本的发展态势看,一方面,各领域预印本平台纷纷出现,论文数量不断上升,预印本在当前学术交流中扮演越来越重要的角色;另一方面,各界对预印本平台运营规范程度的要求进一步提高。在这样的背景和趋势下,ChinaXiv应考虑从技术、政策等多方面进一步改进和优化质量控制机制,为科研人员提供持续、稳定、高质量的预印本服务。具体可从如下方面深入研究和改革。

(1) 进一步优化作者注册和论文上传流程,提高机器辅助审核能力。随着投稿数量的增加,为保证预印本平台在短时间内完成大量论文的质量把关,需借助技术手段提高审核工作效率,同时提升用户体验。一方面,在作者注册阶段优化流程,将作者需提交的信息项进行元数据化,规范作者在线填写。同时,ChinaXiv可以考虑与数据库、学术主页平台(如Research ID)、研究者身份识别平台(如ORCID)等合作,在预印本平台中嵌入个人学术成果、学术信息的查询入口,不仅便于作者更快捷地查找个人相关学术信息,也有利于联合相关平台对作者身份进行认证。另一方面,论文初审环节参考arXiv做法,完善现有机器自动审核论文格式、重复性等功能,提高准确性;同时,使系统能够自动识别并筛查论文是否符合平台要求,节省初审时间;与外部数据源合作,扩大论文重复性检测的范围。

(2) 推动建立更加完善的预印本开放同行评议机制。“先开放,后评议”,在科研成果正式出版前快速得到同行的评价和反馈既是预印本的一大优势,也是预印本质量控制的重要手段。要想充分发挥开放评议机制在预印本质量控制中的作用,不能只限于在平台上设置论文评论、留言区,而应当从评议方式、内容处理、政策引导等方面多下功夫,充分发挥开放评议的价值。如ChinaXiv应将用户对论文的评级、评论内容,在流程上与日常审核工作、领域专家审核有效对接。同时,参考社交网站的评论发言政策(如问答系统Quora),规范预印本平台的开放评议行为,加强对评论内容的分析,提高论文审核的效率。

(3) 确保平台预印本质量控制政策的公开透明,为作者提供更好指导。预印本平台的规范运营和审核机

制的公开透明,是科研人员和资助机构的普遍要求。如NIH于2017年5月发布的一份声明中明确提出“建议预印本平台确立透明公开的审核机制,对论文进行一定程度的质量控制,以保障科研人员对高质量学术内容的需求”^[3]。此外,在国际预印本服务的发展过程中,平台运营方经常受到作者的质疑,认为预印本服务存在偏见。如日内瓦大学的一位量子物理学家怀疑arXiv存在“拉黑”作者的情况,原因是其两名学生的一篇关于黑洞的文章被arXiv拒稿。该物理学家认为这篇论文虽不够严谨,但作为预发布论文来说是合适的^[14]。也有不少人指责arXiv将部分论文放到一些活跃度和关注度不高的分类下,阻碍了学术交流。即使这些质疑的合理性有待商榷,但反映的问题值得重视。一方面,ChinaXiv在未来发展过程中,应重视平台各类质量控制政策的公开,且注意政策内容的细化,将其放置于网站中易于查找的位置;另一方面,加强审核人员的培训,做到规范操作,对审核全过程如实记录并存档。此外,平台还应通过质量控制政策的设置,指导和引导作者提交符合平台要求的预印本,从前端加强预印本质量控制,也能够帮助科研人员更好地使用预印本开展学术交流。

参考文献

- [1] 徐佳宁.国际电子预印本发展现状实证研究[J].图书馆建设,2010(10):42-46.
- [2] Should we cite preprints[EB/OL].(2017-05-14)[2017-10-27].http://fossilsandshit.com/should-we-cite-preprints/.
- [3] NIH.Reporting Preprints and Other Interim Research Products[EB/OL]. [2017-08-15].https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html.
- [4] TOMN W G,VICKERY B A,KURAS J,et al.Chemical e-journals, chemical e-preprints[J].Online Information Review,2002,26(3):164-171.
- [5] Cornell University Library.arXiv[EB/OL].[2017-09-16].https://arxiv.org/.
- [6] Principles of Transparency and Best Practice in Scholarly Publishing [EB/OL].[2017-09-16].http://www.trjemergmed.com/files/principles-of-transparency-and-best-practice-in-scholarly-publishing.pdf.
- [7] SSRN.Frequently Asked Questions[EB/OL].[2017-09-16].https://www.ssrn.com/en/index.cfm/ssrn-faq/#paper_eligibility.
- [8] ChemRxiv.ChemRxiv Beta:The Preprint Server for Chemistry[EB/OL]. [2017-09-16].https://chemrxiv.org/.
- [9] Peer J Preprint.Preprints(not yet peer-reviewed)[EB/OL].[2017-09-16]. https://peerj.com/preprints-search/.
- [10] Preprints.What is a “Preprint”?[EB/OL].[2017-09-16].https://www.preprints.org/policy_and_process.
- [11] WAGER E,KLEINERT S.Responsible research publication: international standards for authors[M].Singapore: Imperial College Press,2010.
- [12] Fledgling site challenges arXiv server[EB/OL].(2017-07-15)[2017-09-16]. http://physicsworld.com/cws/article/news/2009/jul/15/fledgling-site-challenges-arxiv-server.
- [13] 夏莉霞,方卿.国外开放存取期刊的质量评价与控制研究综述[J].信息资源管理学报,2011(2):49-54.
- [14] MERALI Z.ArXiv rejections lead to spat over screening process[EB/OL]. (2016-01-29)[2017-09-16].http://www.nature.com/news/arxiv-rejections-lead-to-spat-over-screening-process-1.19267.

作者简介

刘静羽,女,1989年生,硕士,馆员,研究方向:开放资源建设。

张智雄,男,1971年生,博士,研究员,博士生导师,通讯作者,研究方向:信息系统和智能信息处理,E-mail: zhangzhx@mail.las.ac.cn。

黄金霞,女,1972年生,博士,副研究馆员,硕士生导师,研究方向:开放资源建设。

彭媛媛,女,1993年生,硕士研究生,研究方向:开放资源建设、数据评价。

Study on the Quality Control Methods of Preprint Services

LIU JingYu¹, ZHANG ZhiXiong^{1,2}, HUANG JinXia¹, PENG YuanYuan^{1,3}

(1.National Science Library, Chinese Academy of Sciences, Beijing 100190, China; 2.Wuhan Library, Chinese Academy of Sciences, Wuhan 430071, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Quality control is the critical part of preprints service. The international practice of quality control includes three aspects: verifying the scholar identification of registers and authors, examining the format integrity of papers, reviewing whether the content is academic and faithful, or not. Manual review is the primary channel of quality control, while the machine review plays as the supplement. Following the above international practice, ChinaXiv established its preprints quality control mechanism, including the mechanism of conventional review, subject experts' review, open peer review, articles' withdraw, dishonest authors' list. Later, the mechanism can be improved by optimizing the workflow of registration and upload of articles, enhancing the ability of machine auxiliary review, promoting to set up a perfect open peer review mechanism, insuring to make the quality control policy public.

Keywords: Preprint; Quality Control; Preprint Services; Open Access

(收稿日期: 2017-09-29)