

# 科技学术期刊数据可用性声明规范 应用的现状及建议

雷 雪

中国科学技术信息研究所,100038,北京

**摘要** 随着科学数据开放共享及科技学术期刊数据出版实践的开展,数据可用性声明作为一种作者声明机制,日益成为期刊学术规范体系的一部分,并逐渐走向标准化。本文系统梳理数据可用性声明规范的产生,对作者、读者、期刊和资助机构的作用,通用的声明类型及实践开展现状,并对推进数据可用性声明规范的应用提出建议。

**关键词** 数据可用性声明;科技学术期刊;学术规范

**Application status and suggestion for data availability statement in academic journals//LEI Xue**

**Abstract** With the development of scientific data sharing and the practice of data publishing in academic journals, data availability statements(DAS) has increasingly become a part of the academic standard system of journals, and is gradually becoming standardized. In this paper, we systematically review the generation of DAS, the effect on authors, readers, journals and funding institutions, the general types of claims and the status quo of its practice, and also propose some suggestions for promoting the application of DAS.

**Keywords** data availability statements; scientific and technical academic journals; academic norms

**Author's address** Institute of Scientific and Technical Information of China, 15 Fuxing Road, 100038, Beijing, China

**DOI:**10.16811/j.cnki.1001-4314.2021.03.008

严谨、公开、可客观检验的学术规范是期刊学术质量和科研诚信的重要保障。国际出版界已出台一系列的期刊学术规范,包括研究方法、内容结构、数据与图表处理、作者贡献、利益冲突、参考文献、辅助资料、版权与使用许可、学术不端行为界定等。信息技术与科学研究的交互融合促成了科学研究第四范式,即数据密集型科学发现的到来,数据作为科学研究的主要产出,已日益成为科技学术期刊(下文简称“学术期刊”)出版的重要内容。为促进科学数据共享,规范科学数据出版,在大型出版商的推动下,数据可用性声明(Data Availability Statements)作为一种作者声明机制,日益成为学术交流规范体系的一部分,并逐渐走向标准化、制度化。尽管数据可用性声明规范已在诸多国外期刊中得到应用,并有学者对实践经验进行总结和分析<sup>[1]</sup>,但在国内期刊中的应用尚未普及,仅有王丹丹等<sup>[2]</sup>、盛小平等<sup>[3]</sup>在相关研究中有所论述。本文将

系统梳理数据可用性声明规范的产生、作用、类型及实践,并对推进数据可用性声明规范的应用提出建议。

## 1 数据可用性声明规范的产生

### 1.1 科学数据开放共享政策的推动

2004年经济合作与发展组织成员国及中国、南非共和国等34个国家签署了《开放获取公共资助研究数据宣言》<sup>[4]</sup>,倡导努力建立公共资助中获取的数字研究数据的访问机制;2007年发布的《从公共资助获取研究数据的原则和指导方针》<sup>[5]</sup>进一步明确成员国在制定科学数据政策时应考虑的原则,包括开放性、灵活性、透明度、法律符合性、知识产权保护、互操作性、安全性和可持续性等,以促进科学数据获取和共享方面的良好实践。随后诸多国际组织、政府部门、科研资助机构、社会慈善机构、大学和科研机构等主体均制定了一系列的科学数据开放共享政策<sup>[6]</sup>,部分政策与学术期刊密切相关。

如欧盟委员会2010年发布的《欧洲数字议程》<sup>[7]</sup>明确指出公共资助的研究应该通过科学数据和论文的开放获取出版进行广泛传播;2012年发布的《关于科学信息的获取与保存》<sup>[8]</sup>提出在考虑知识产权的情况下,迫切需要采取政策行动尽早向读者免费提供经过同行评议的科学出版物和研究数据;数据集应易于识别,可通过适当机制与其他数据集和出版物相联系,并提供额外信息使其能够被适当评价和使用。

又如,英国国家科研与创新署(UK Research and Innovation, UKRI)于2018年对原英国研究理事会2015年发布的《研究数据管理最佳实践指南》进行修订,进一步明确获取UKRI资助的研究数据必须遵循7项共同原则<sup>[9]</sup>,包括应以一种有助于其他研究人员理解和重用的方式记录并公开足够的元数据,提供如何获取数据的信息,使其易于访问和发现;研究数据的发布受到法律、道德和商业限制,研究机构的政策和实践应在研究过程的所有阶段都考虑这些因素;为确保研究团队在收集和分析数据方面的努力得到适当认可,研究团队有权在一定期限内特权使用所收集和分析的数据;为认可产生、保存和共享关键研究数据集的研究

人员的智力贡献,使用的任何研究数据都应确认其来源,并遵守数据访问的条款和条件等。

上述政策,一方面推动了世界范围内科学数据开放共享实践的开展,另一方面也提供了科学数据出版和管理的总体原则和行动指南。

## 1.2 学术期刊数据出版实践的开展

科学数据是科学研究不可或缺的组成部分,数据与文献的关联由来已久,对于文献具有重要意义。不仅可以作为文献的结论来源和复核基准,部分学者的研究表明,数据链接还可以显著提高文献被引。以天体物理学领域为例,天体物理学数据系统(ADS)提供出版文献到相关数据集的链接,允许用户直接访问与文献相关的数据<sup>[10]</sup>。Dorch等<sup>[11]</sup>以2000—2014年在主要天体物学期刊上发表的论文为调研对象,按照是否具有数据链接,将ADS中的记录进行分类计量,结果表明,与没有数据链接的论文相比,有数据链接的论文平均被引次数要多得多,即存在引文优势的论文与数据有关。

在大数据时代背景及开放科学运动推动下,科学数据作为重要的战略资源,逐渐成为科研产出的“一等公民”,具有独立的身份识别、属性描述、监控机制、溯源流程<sup>[12]</sup>。数据出版正是使科学数据能够被社会广泛而有效利用的一种手段,即通过一定的公共机制发布科学数据集,使公众能够在一定规则下发现、获取、重用和评价。期刊界和科学界对科学数据及其出版认识也在不断提高。2013年Nature系列期刊面向生命科学文章实施了18项审查,其中包括需要提供数据和代码的可用性声明,并强烈建议通过公共存储库共享数据<sup>[13]</sup>。2014年3月PLoS推行新的数据开放政策,要求作者提交论文时,必须提交一份基于PLoS政策的数据可用性声明,共享科研数据及与之相关的元数据和方法<sup>[14]</sup>。2014年6月美国国立卫生研究院与Nature出版集团和《Science》就研究结果的可重复性进行联合研讨会,提议期刊要求作者将支持论文结论的所有数据作为同行评议的内容在出版时公开;鼓励共享软件,并要求在手稿中至少有一个声明,描述软件是否可用及如何获取<sup>[15]</sup>。

为进一步推进数据出版标准化,明确论文数据存储、访问及共享等相关问题,2016年3—5月,Springer Nature将数据可用性声明以标准规范的形式引入《Nature Neuroscience》等5种期刊,成功试行2个月;2016年9月,《Nature》采用数据可用性声明规范,使其期刊政策与标准化的Springer Nature研究数据政策框架相一致<sup>[1]</sup>;随后所有发表在《Nature》以及Nature家族生命科学期刊上的研究论文,都被要求包括如下信

息:支持他们研究的数据是否可用,以及他人如何获取这些数据。自Springer Nature引入并实施数据可用性声明规范后,Elsevier、Wiley、Taylor & Francis等其他大型出版商也推出了类似举措。

## 2 数据可用性声明规范的作用

Springer Nature、Elsevier、Wiley、Taylor & Francis等出版商普遍认为,科学数据指的是验证作者研究发现的观察或实验结果,涵盖与作者研究项目相关的一系列有用材料,包括但不限于原始或处理过的数据文件、软件、代码、模型、算法、方案、方法、材料等,表现形式可以为电子表格、文档、图像、视频或音频等。

数据可用性声明是一种简单且可互操作的机制,用于明确作者是否共享数据、数据存储位置及访问条件等。在宏观层面,数据可用性声明可增加数据透明度,推进科学数据开放共享,加快数据出版进程,规范数据出版流程,构建更加开放的研究环境,促进更快和更有效的科学发现。在微观层面,数据可用性声明对于作者、读者、期刊和资助机构来说,都具有重要意义。1)对于作者,通过数据可用性声明提供所使用的存储库链接,便于他人访问和获取,能够因为所做的工作而获得相应的荣誉,提升论文可信度,带来更多的引用,也可能带来新的合作机会;2)对于读者,通过数据可用性声明能够准确、快捷地获取共享数据,提升学术交流效率,重用和重新解释数据很容易,可以避免重复实验,从而有更多的机会进行元分析和新知识提取;3)对于期刊,能够为作者提供依据和标准,使作者更好地遵从期刊数据政策和科学数据管理要求,鼓励科学实践和建立信任机制,同时体现出期刊的规范性和严谨性,提升期刊的学术声誉;4)对于资助机构,有利于监测出版商和研究团体遵守数据政策的情况,将其作为证据来支持资助机构政策,同时改善学术界以外人员获得科学数据的途径,提升公众对研究的信心,减少重复工作的资金,从而使公共财政受益。

## 3 数据可用性声明规范的类型

鉴于Springer Nature、Wiley、Taylor & Francis 3大学术期刊出版商旗下分别有3 000余种<sup>[16]</sup>、1 600余种<sup>[17]</sup>、2 700余种<sup>[18]</sup>期刊(截至2021年2月),覆盖科学、技术、工程、医学、人文和社会科学的所有领域,影响力大,且提供明确的数据可用性声明规范,具有很好的示范引领作用和代表性,因此笔者以上述出版商官网发布的数据政策为样本进行分析。数据可用性声明的类型如表1所示。

当数据可用和可链接时,作者需要在他们的参考

文献中提供数据引用。3大出版商也同时提供了数据可用性声明标准模板供作者参考。在声明模板上,相

似或相同声明类型对应的格式也具有趋同性,如表2所示。

表1 数据可用性声明类型

声明描述	Springer Nature <sup>[19]</sup>	Wiley <sup>[20]</sup>	Taylor & Francis <sup>[21]</sup>
数据在使用 DOIs 发布数据集的公共存储库中公开可用	√	√	√
数据在不使用 DOIs 发布数据集的公共存储库中公开可用	√	√	√
数据来自公共领域资源	√	√	√
数据由于商业限制而共享受限	√	√	√
数据因隐私/伦理限制不公开,但应合理要求可从通信作者处获取	√	√	√
数据受第三方限制	√	√	√
数据可根据作者要求由其提供	√	√	√
数据共享不适用——没有新的数据生成,或完全描述理论研究	√	√	√
数据可在文章和/或其补充材料中获取	√	√	√
数据生成于中央、大型设施中,可根据要求提供	√	—	√
研究期间产生的数据受数据共享授权约束,并且可以在不使用 DOIs 发布数据集的公共存储库中使用	√	—	—
作者选择不共享数据	—	√	√
非数字数据	√	—	√

表2 数据可用性声明样例

出版商	声明类型/描述	声明模板/实例文本
Springer Nature	从公共资源中派生并在文章中可访问的数据集	本研究分析的数据集可在[存储库名称]获取,存储[数据标示,如 doi:10.7910/DVN/HEWGDD][参考号]。这些数据来自以下公共领域资源:[资源列表和它们的URLs]
Wiley	来自公共领域资源的数据	支持本研究结果的数据可在[存储库名称]中获取,位于[URL/DOI],参考号[参考号]。这些数据来源于以下公共领域的可用资源:[资源列表和URLs]
Taylor & Francis	数据来自于公共领域	支持本研究结果的数据可在[存储库名称][URL/DOI]获取,参考号[参考号]。这些数据来源于以下公共领域的可用资源:[资源列表和URLs]

数据可用性声明的格式和位置会受到期刊独特风格的影响,如大多数 Elsevier 期刊都在手稿提交流程中,指引作者直接创建数据可用性声明,发布于文章大纲“研究数据”部分;Taylor & Francis 的大部分期刊则要求数据可用性声明在文章原稿内提交,放置于“参考文献”部分之前。但一般情况下,任何声明都应包括数据类型/描述、数字对象标识符(DOI)、使用条款(如禁令、许可协议、数据共享协议)等内容。

#### 4 数据可用性声明规范的实践

数据可用性声明是出版商或学术期刊数据共享政策内容的一部分。通常学术期刊出版商在数据共享政策中对数据可用性声明的提供做出规定。仍以上述3大学术期刊出版商为例,数据共享政策与数据可用性声明的关系如表3所示。

由表3可知,3大学术期刊出版商均采用分级的数据共享政策,数据共享强度从鼓励到披露再到强制要求,层层推进。数据共享强度较低的政策,一般鼓励或强烈建议提供数据可用性声明;数据共享强度较高

的政策,数据可用性声明是强制要求提供的。相应地,出版商旗下的期刊可根据学科性质、内容质量和办刊风格选择适合的数据共享政策等级,对应提供数据可用性声明。以 Springer Nature 为例,据统计,截至2021年2月,其旗下采用标准化数据共享政策的期刊共计1707种,其中646种期刊选用类型1,528种期刊选用类型2,524种期刊选用类型3,9种期刊选用类型4,意味着528种期刊被鼓励提供数据可用性声明,533种期刊被强制要求提供数据可用性声明<sup>[24]</sup>。

国内机构主办的学术期刊中,彭琳等<sup>[25]</sup>以中国科学院主办的65种SCI收录的英文科技期刊为研究对象分析其采用的数据政策,调查结果显示,只有8种期刊的数据政策中提到了数据可用性声明,且均与Nature和Elsevier出版商具有合作关系。国内数据出版处于起步阶段,但一些示范性的数据期刊已推出相应的政策,如《中国科学数据(中英文网络版)》,在其“出版伦理与学术不端声明”中提出了提供“数据可用性声明”的要求<sup>[26]</sup>;《全球变化数据学报(中英文)》在其数据论文模板中明确规定了数据可用性声明的位

置,设置在摘要、关键词和 DOI 号之后,正文之前<sup>[27]</sup>。

表3 数据可用性声明与数据共享政策的关系

出版商	鼓励/强烈建议提供数据可用性声明	要求/强制提供数据可用性声明
Springer Nature <sup>[22]</sup>	类型2:鼓励数据共享和数据证据共享	类型3:鼓励数据共享并要求提供数据可用性声明 类型4:鼓励数据共享、数据证据共享,以及要求数据进行同行评审
Wiley <sup>[20]</sup>	类型1:鼓励数据共享	类型2:期望数据共享 类型3:要求数据共享 类型4:要求数据共享和同行评审数据
Taylor & Francis <sup>[23]</sup>	类型1:基本模式	类型2:应合理诉求共享 类型3:公开获取 类型4:开放数据 类型5:开放并遵循 FAIR 标准

现今数据可用性声明规范已在部分期刊中得到应用,但实践中仍存在期刊数据可用性声明的描述不清晰,或者使用者对期刊相关要求理解不到位、数据信息获取不充分的情况,从而引发一定的知识产权纠纷。如2021年2月10日,上海交通大学研究者在《Science Advances》在线发表题为《A methylo-trophic origin of methanogenesis and early divergence of anaerobic multi-carbon alkane metabolism》的研究论文;2月11日,蒙大拿州立大学的 Roland Hatzenpichler 即声明该文未经允许使用了包括其本人在内的一些 PI 的数据<sup>[28]</sup>;2月17日,该文作者申明因在分析中无意包含了一些禁止使用的数据,为保护数据生成者的利益,主动撤回稿件<sup>[29]</sup>。因此,数据可用性声明规范仍需在实践中进一步推广、加强和完善。

## 5 推进数据可用性声明规范的建议

### 5.1 优先在高质量期刊中强制推行规范

部分研究者基于调查分析期刊数据政策和期刊特点的关系,发现制定数据政策的期刊学科分布不均衡,但存在一些共性现象,影响因子高的期刊更重视数据出版。如 Nicole 等<sup>[13]</sup>、Paul 等<sup>[30]</sup>学者的调查研究表明,高影响因子和具有数据共享需求的期刊之间存在显著关联,影响因子越高的期刊,采用的数据共享政策往往越强硬,低影响力期刊很少采用数据共享政策,或者并不强制要求数据共享,相应的数据可用性声明是否采纳也具有同样规律。究其原因,知名期刊可能处于更好的位置,更愿意也更易于对作者施加新的要求和实践,作者对数据政策的接受度也更高。因此,在数据可用性声明规范推进中,期刊主管部门和主办部门可优先选择在高质量数据出版期刊中试行,进而强制使用,以点带面,形成业界共识,从而促进数据可用性声明规范的普及应用。

### 5.2 加强科研人员和期刊编辑的认知和培训

虽然采用数据可用性声明的目的是明确作者数据共享与否及存储位置,但出版商必须考虑规范所适用的学科文化,制定清晰、准确的数据可用性声明政策,确保科研人员了解其义务是什么,需要怎么做才能遵守规范,以及若使用数据不当会面临的知识产权风险;也必须对编辑人员进行培训并提供相应资源,这样才能为科研人员提供良好的支持服务。Grant 等<sup>[1]</sup>研究表明,Nature 系列期刊引入数据可用性声明后,编辑处理文章的时间会有所增加(平均每篇增加 15~20 min),但考虑到数据可用性声明的重要性和益处,对期刊出版工作流程的影响及增加的时间是合理和可接受的。从出版时间成本的角度考虑,出版商需要对科研人员和期刊编辑提供支持和培训,以推动数据可用性声明的尽快普及。

### 5.3 促使数据尽快纳入科研评价体系

中国科学院院士郭华东认为,“科学家科研贡献和学术声望评价机制不改变,是阻碍科学数据开放共享实践的一个根本问题”<sup>[31]</sup>。即数据共享的障碍之一是,它要求科研人员付出额外的工作,却没有多少回报,缺乏激励机制。而数据引用有可能改变这一点,它们可以很容易地纳入当前基于论文引用的奖励系统,承认作者贡献,给予作者信誉。因此,在制度层面上,提高科研人员对数据共享的认知,将数据尽快纳入科研评价体系,推动数据引用的分析和统计工作,提供科研人员数据共享的有效激励,从根本上加快科学数据共享、数据出版的进程,自然就推进了包括数据可用性声明规范在内的数据共享政策的落地和实施。

## 6 参考文献

- [1] GRANT R, HRYNASZKIEWICZ I. The impact on authors and editors of introducing data availability statements at nature journals [J]. International Journal of Digital

- Curation, 2018, 13(1): 196
- [2] 王丹丹, 刘清华, 葛力云. Springer Nature 科研数据政策标准化工作实践及启示[J]. 图书情报工作, 2020, 64(18): 137
- [3] 盛小平, 郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作, 2020, 64(22): 25
- [4] Declaration on access to research data from public funding [EB/OL]. [2021-02-19]. [http://www.unitus.it/public/platforms/32/cke\\_contents/3148/oecd\\_declaration.pdf](http://www.unitus.it/public/platforms/32/cke_contents/3148/oecd_declaration.pdf)
- [5] OECD. OECD principles and guidelines for access to research data from public funding [EB/OL]. [2021-02-19]. <http://www.oecd.org/science/inno/38500813.pdf>
- [6] 温芳芳. 国外科学数据开放共享政策研究[J]. 图书馆学研究, 2017(9): 91
- [7] European Commission. A digital agenda for Europe [EB/OL]. [2021-02-19]. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>
- [8] European Commission recommendation on access to and preservation of scientific information[J]. JLIS.it, 2012, 3(2): 9
- [9] Common principles on research data [EB/OL]. (2020-10-26) [2021-02-19]. <https://www.ukri.org/apply-for-funding/before-you-apply/your-responsibilities-if-you-get-funding/making-research-data-open/>
- [10] EICHHORN G, ACCOMAZZI A, GRANT C S, et al. From the literature to the data: linking to on-line data from the ADS[J]. Bulletin of the American Astronomical Society, 2001(33): 1320
- [11] DORCH B F, DRACHENTM, ELLEGAARD O. The data sharing advantage in Astrophysics [EB/OL]. [2021-02-19]. <https://arxiv.org/pdf/1511.02512.pdf>
- [12] 孔丽华, 习妍, 张晓林. 数据出版的趋势、机制与挑战[J]. 中国科学基金, 2019(3): 237
- [13] NICOLE A V, JESSICA M, MELISSA A H, et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? [EB/OL]. (2017-04-25) [2021-02-19]. <https://peerj.com/articles/3208.pdf>
- [14] 吴蓉, 顾立平, 刘晶晶. 国外学术期刊数据政策的调研与分析[J]. 图书情报工作, 2015, 59(7): 100
- [15] National Institutes of Health. Principles and guidelines for reporting preclinical research [EB/OL]. [2021-02-19]. <https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>
- [16] Home to the world's most influential journals [EB/OL]. [2021-02-12]. <https://www.springernature.com/gp/products/journals>
- [17] Find the right journal to publish your research [EB/OL]. [2021-02-12]. <https://authorservices.wiley.com/author-resources/Journal-Authors/find-a-journal/index.html>
- [18] About Taylor & Francis Group [EB/OL]. [2021-02-12]. <https://taylorandfrancis.com/about/>
- [19] Data availability statements [EB/OL]. [2021-02-12]. <https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>
- [20] Wiley's data sharing policies [EB/OL]. [2021-02-12]. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>
- [21] Data availability statements [EB/OL]. [2021-02-12]. <https://authorservices.taylorandfrancis.com/data-sharing-policies/data-availability-statements/>
- [22] Research data policy types [EB/OL]. [2021-02-12]. <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096>
- [23] 数据共享政策 [EB/OL]. [2021-02-12]. <https://authorservices.taylorandfrancis.com/wp-content/uploads/2019/04/Data-sharing-policies-in-Chinese.pdf>
- [24] Find your journal's data policy and services [EB/OL]. [2021-02-12]. <https://www.springernature.com/gp/authors/research-data-policy/springer-nature-journals-data-policy-type/12327134>
- [25] 彭琳, 韩燕丽. 我国科技期刊数据政策分析及启示: 以中国科学院主办英文期刊为例[J]. 中国科技期刊研究, 2019, 30(8): 873
- [26] 《中国科学数据(中英文网络版)》出版伦理与学术不端声明 [EB/OL]. [2021-02-12]. <http://csdata.org/p/static/1327/>
- [27] 全球变化数据学报数据论文模板: 中文 [EB/OL]. (2019-12-10) [2021-04-12]. <http://www.geodoi.ac.cn/WebCn/DocList.aspx>
- [28] 'No malicious intent': authors retract week-old Science Advances paper based on embargoed data [EB/OL]. (2021-02-17) [2021-04-12]. <https://retractionwatch.com/2021/02/17/no-malicious-intent-authors-retract-week-old-paper-based-on-embargoed-data/>
- [29] WANG Y Z, WEGENER G, WILLIAMS T A, et al. Retraction of the research article: "a methylophilic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism" [J]. Science Advances, 2021, 7(8): eabh1051
- [30] PAUL S, MARIANNE B, JANE H S A, et al. Research data sharing: developing a stakeholder-driven model for journal policies [J]. Journal of the Association for Information Science and Technology, 2015, 66(12): 2448
- [31] 田瑞颖. 《中国科学数据》: 从根源唤醒沉睡的数据 [EB/OL]. (2021-01-18) [2021-02-12]. [http://www.cas.cn/cm/202101/t20210118\\_4774891.shtml](http://www.cas.cn/cm/202101/t20210118_4774891.shtml)  
(2021-02-20收稿;2021-04-19修回)