

期刊引证中科学数据引用的调研分析

张英杰¹ 彭洁^{1,2} 张新民¹ 吴国雄¹ 李馨³

(1. 中国科学技术信息研究所, 北京 100038; 2. 武汉大学信息管理学院, 武汉 430072; 3. 北京万方数据股份有限公司, 北京 100038)

摘要: 论文间的引用关系代表了作者对其他论文的观点、方法、数据和结论等的引证, 它是联接科学继承与创新的重要纽带。在对国内科技期刊编辑部对提交科学数据的态度、科学数据引用的要素、潜在的好处和面临的困难等问题进行问卷调查的基础上, 研究分析期刊编辑部在科学数据引用相关问题上的态度和现状。结果表明, 各编辑部已经充分意识到科学数据引用的潜在应用价值, 但存在着认识不足、缺少培训、规范和系统支撑不力等阻碍因素。最后归纳总结了期刊编辑部推进科学数据引用的具体方式, 并对有关数据存缴平台建设、知识产权保护 and 标准规范等方面提出了建议。

关键词: 期刊; 科学数据引用; 认知与实践; 问卷调查; 数据共享

中图分类号: G237.5

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2014.02.004

Awareness & Practice Survey on Scientific Data Citation from China's Periodicals

Zhang Yingjie¹, Peng Jie^{1,2}, Zhang Xinmin¹, Wu Guoxiong¹, Li Xin³

(1. Institute of Scientific and Technical Information of China, Beijing 100038; 2. School of Information Management of Wuhan University, Wuhan 430072; 3. Wanfang Data Co., Ltd., Beijing 100038)

Abstract: The citation between literatures is on behalf of the author's thesis ideas, methods, data and conclusions, which is an important link of inheritance and innovation of science. It designed survey questionnaire for homeland periodical editorial staff on submission attitudes, data citation elements, potential benefit and ongoing difficulties, etc.. It is found by survey that potential application value has been achieved by surveyed editorial staffs, meanwhile, which are facing less awareness, insufficient training, lack of support from system and regulation, etc. Finally, some pertinent advices also are raised around promotion methods, scientific data repository platform, regulation and rules, intellectual property, etc.

Keywords: periodicals, scientific data citation, awareness & practice, survey questionnaire, data sharing

1 引言

论文间的引用关系代表了作者对其他论文的

观点、方法、数据和结论等的引证, 它是联系科学继承与创新的重要纽带。各期刊编辑部都非常重视参考文献的引用和著录工作, 基于科技文献

作者简介: 张英杰 (1979-), 男, 博士, 中国科学技术信息研究所助理研究员, 研究方向: 科技信息资源管理。彭洁 (1965-), 女, 中国科学技术信息研究所研究员, 研究生导师, 研究方向: 科技管理与科技资源共享。张新民 (1967-), 男, 研究员, 博士, 中国科学技术信息研究所副所长, 研究方向: 科技管理与科技资源共享。吴国雄* (1970-), 男, 中国科学技术信息研究所编辑, 研究方向: 期刊评价、文献计量学。李馨 (1986-), 女, 硕士, 研究方向: 科技期刊计量。

基金项目: 国家科技支撑计划项目“电动汽车专题数据库建设”(2013BAG06B02); 中国科学技术信息研究所项目预研基金“科研项目全生命周期多源信息整合机制研究”(YY201212)。

收稿日期: 2014年2月26日。

间的引用,可以开展科技期刊、学术影响力的评价,直接关系到相关方利益,因此备受作者、期刊编辑和科研管理部门的关注^[1-2]。

根据GB 7714-2005《文后参考文献著录规则》的相关规定,常用的引用文献类型包括期刊、专著、专利、科技报告等8类,电子文献类的引用文献载体类型则包括磁带、磁盘、光盘和联机网络4类^[3]。长期积累形成的科学数据作为一种基础科研资源具有极高的科研价值,能够为相关领域的科研工作提供重要参考和支持。而在科学数据共享利用过程中,尚未形成类似于传统文献的、规范的参考引用机制,对科学数据的价值、使用情况和效果缺少可量化的衡量手段。

近年来,伴随着期刊引用评价的逐步深入,人们的目光正投向科学数据这种独特的“参考文献”。早在2006年,微软研究院发布的《迈向科学2020年的路线图草案》(Towards 2020 Science: A Draft Roadmap)中,对科学数据出版进行了创造性地预测:“大约10年内将出现新的科学出版物类型,即为研究者专门出版有价值的科学数据(valuable scientific data)”^[4-5]。而在开放访问科学数据及其与科技文献的互操作方面,科研人员也一直保持着强烈需求^[6]。目前,数据引用正在成为科学数据应用的一种基本方式,对于数据引用的格式,国外有STD-DOI、DataCite、PANGAEA、ICPSR、Dataverse Network等项目或机构的引用格式^[7],国内中国科学院科学数据库平台也发布了自己的科学数据引用格式^[8],中科院寒区旱区研究所的中国西部环境与生态科学数据中心则提供了数据平台的科学数据引用服务^[9]。如今,在科学数据领域也出现了专业的数据引文索引(DCI),进一步推动对数据集和数据研究的发现、使用及归属,使研究人员能够在数据研究支持的文献上下文中发现、引用和查看该数据^[10]。

在大数据的背景下,我们对数字出版环境下的科学数据引用实现路径和策略进行了分析^[11],探寻作为数字出版的主体,期刊编辑部如何能够适应快速变化的科技浪潮,传播科学真知,通过建立健全的科学数据引用规则,拓展参考文献的

新功能、新服务,主动面对科学数据应用的新机遇。本文在前期调研的基础上,针对国内期刊编辑部,调研编辑部人员对开展科学数据引用的态度以及实施科学数据引用的条件保障情况。

本次调查问卷共包含16个问题,涵盖4个背景性问题,主要涉及被调查期刊的基本情况,如期刊名称,期刊所属类别以及联系方式;2个客观性问题,主要涉及被调查期刊所熟悉的国内外科学数据库;10个主观性问题,主要用于调研期刊所在编辑部对科学数据引用的态度、意愿等主观世界状况,其中包括4个单选项问题,5个多选项问题,1个开放式问题。问卷总共回收1442份,其中通过网络地址直接填写的有1364份,通过E-mail填写返回Word版问卷78份,但有6份因答题项缺失,为不合格问卷,故最终的有效问卷数为1436份。

2 调查对象背景分析

在1436份调研问卷中,共涉及905种期刊,其中填写4~6份调研问卷的杂志有36种,占总数的4%,其中《办公自动化(办公设备与耗材)》《电脑知识与技术》《电子与信息学报》3本杂志各填写了6份调研问卷;填写3份调研问卷的杂志有45本,占总数的5%;填写2份调研问卷的杂志有318本,占总数的35.1%;填写1份调研问卷的杂志有506本,占总数的55.9%。根据国家新闻出版总署最新公布的《2012年全国新闻出版业基本情况》^[12],2012年全国共出版期刊9867种,其中自然科学类期刊4953种,故本次调研所涉及的期刊占自然科学类期刊的18.3%。

从所调研期刊的类别来看,905种杂志总共涵盖了八大类,其中填写问卷最多的类别是工业技术,内含无线电电子学与电信技术、自动化技术与计算机技术、一般工业技术、矿业工程等20个子类,涉及调研问卷880份(61.27%),其次是教科文艺160种(11.14%),第三是社会科学153种(10.65%)。其余的类别分别是医药卫生72种(5.01%),基础科学71种(4.94%),农业科学54种(3.76%),哲学政法35种(2.44%),经济财政

11种(0.77%)。

从调研对象中填写人员的职务来看，总共包括了879名编辑(占61.21%)，328名责任编辑(22.84%)，118名编辑部主任(占8.22%)，111名副主编及主编(占7.73%)。

本次调查所涉及的905种期刊，涵盖了各个层次的期刊编辑人员，涉及理工农医等主要期刊门类，问卷的填写区域除宁夏、青海、西藏三地外，已经覆盖全国所有省份，具有一定的代表性。其中在线填写问卷最多的基本省份分别是山东(848份，占59.05%)、北京(213份，占14.83%)、湖北(43份，占2.99%)。

3 期刊对科学数据引用的认知与基础

(1) 期刊对科学数据提交的态度

针对“贵刊对作者提交论文时需要提交论文的数据集持什么态度？”这一问题，在5个选项中，有1325份(占92.27%)要求提交数据集，其中有515份(占35.86%)选择了“作者自愿提交”，有489份(占34.05%)选择了“要求部分数据强制提交”，有225份(占15.67%)选择了“制定措施鼓励提交”，而只有96份(占6.69%)“采取措施强制提交”，其余的111份(占7.73%)则选择了“无所谓，目前还不清楚”。

在回答“增加科学数据引用是否会增加编辑部的工作量？”时，有614人选择了“不会”(占42.76%)，有703人选择了“会”(占48.96%)。

(2) 期刊对选择科学数据集的存缴平台意向

针对期刊文章所涉及科学数据集的存缴意向，各编辑部更倾向于要求作者将相关科学数据集提交到“编辑部自建的数据存储平台”，共有889人(占61.91%)选择了此项，此外有811人(占56.48%)选择了“期刊专业领域内的专业数据中心”，而“CNKI或万方数据等传统文献数据库平台”和“第三方公益性数据存储中心”分别有685人(占47.7%)和533人(占37.12%)。由此可见，鉴于科学数据的专业性强，与期刊文章联系紧密，期刊编辑部更倾向于在这个过程中扮演“主角”，如《水文》杂志，其中的水文信息通

过监测(测)—传输(报)—整理(整)—计算(算)等途径得到的科学数据，十分珍贵，在《水文》期刊中占有较大篇幅，非常期待能有机会将这些科学数据进行统一的摘录、收集和存储。而传统的文献数据库出版商和第三方在其中只能扮演“配角”的平台作用。这一方面为传统文献出版商开拓此项应用提供了启示；另一方面也为各期刊编辑部提供了机遇，需要专注出版，创新出版模式。如以GigaScience杂志为例，“GigaScience采用一个崭新的模式，即允许作者将其研究成果的摘要(summary)和数据发布，无需提出结论(conclusion)。^[13-14]”这种模式越过了对研究数据进行分析 and 讨论的工作，使研究者能够迅速将数据共享，这将在很大程度上缩短数据公布的时间。

(3) 阻碍作者提交科学数据集的因素

在我们的调查中，共设计了7个因素，统计后发现“没有便捷的数据提交工具或软件”成为阻碍科学数据提交的首要因素，共有919人(占64%)选择了该项，而“数据有一定的保密性”“数据太大，找不到合适的科学数据存缴中心”“还需要进一步支撑未来的研究，不愿意公开”位于阻碍因素的第二梯队，分别有869人(占60.52%)、845人(占58.84%)和827人(占57.59%)，最后“数据质量不高”“太麻烦，增加了负担”和“数据提供方没有授权”是最次要的阻碍因素，分别有500人(占34.82%)、442人(占30.78%)和396人(占27.58%)。从中我们可以看到，期刊编辑部若想开展科学数据引用，需要开发建设方便易用的数据提交工具，形成专业领域内的影响力，此外，从现状来看要合理解决科学数据保密与知识产权问题还需要持续的探索。

(4) 修订参考文献著录规则(GB/T7714-2005)的必要性

在现行的期刊中，普遍采用了国标《参考文献著录规则(GB/T7714-2005)》^[3]。它是2005年3月23日，由国家质量监督检验检疫总局和中国标准化管委会发布，以代替原有的GB/T 7714-

1987。该标准非等效采用国际标准ISO 690 和ISO 690-2, 规定了各个学科、各种类型出版物的文后参考文献的著录项。标准中确定的文献类型有数据库DB一项, 而缺失科学数据。2012年, 基础科学数据共享网发布了专门的《TR-REC-069 科学数据引用规范》^[8], 作为项目标准, 该标准界定了基础科学数据共享网的科学数据的引用规范, 起到了良好的示范作用。故在是否有必要针对科学数据的著录, 对GB/T 7714-2005进行修订以增加科学数据著录项的调研中, 有982人(占68.39%)认为有必要修订该标准, 其中更有315人(占21.94%)认为“很有必要”。有422人(占29.39%)认为“条件还不成熟, 暂时没必要”, 只有32人(占2.23%)认为“没必要”。

(5) 科学数据引用的著录元素

在科学数据的描述性元数据中, 包括了数据生产者、生产时间、数据名称、数据发布单位以及数据唯一标识符等元素。在《TR-REC-069 科学数据引用规范》^[8]中, 也规范了科学数据的引用元素, 包括作者、名称、发布机构、发布年份、传播机构、传播时间、唯一标识符和解析地

址8个必选元素以及版本这一可选元素。虽然元素名称在实际应用中各不相同, 但哪些是科学数据引用的必备元素, 各元素的重要性如何? 在调研中, 我们设计了10个元素, 结果发现位列前三的分别是数据名称(1001人, 69.71%)、数据创建者(836人, 58.22%)、数据授权方式(578人, 40.25%)、详细的排名情况如图1所示。这为进一步设计修订相关规范提供了需求依据。

(6) 科学数据引用的作用

在发表科技文献时, 科研人员已经形成了列出参考文献的习惯, 但是对于文中所引用的数据却并没有留意, 尚未普遍形成引用科学数据的意识, 对规范科学数据引用能够给科研人员带来的潜在好处认识也不充分^[14]。在本次调研中, 从期刊编辑人员来看, 规范地引用科学数据, 认为“可以使其他科研人员更好地去验证自己的研究或他人的成果”, 共有1107人(占77.09%), 另有998人(占69.5%)认为“可以使科学数据引用更加快捷, 更易检索及访问”, 其次则是“可以获得大量当前还无法实现的新的研究机会”, 有912人(占63.51%), 详细的情形如表1所示。

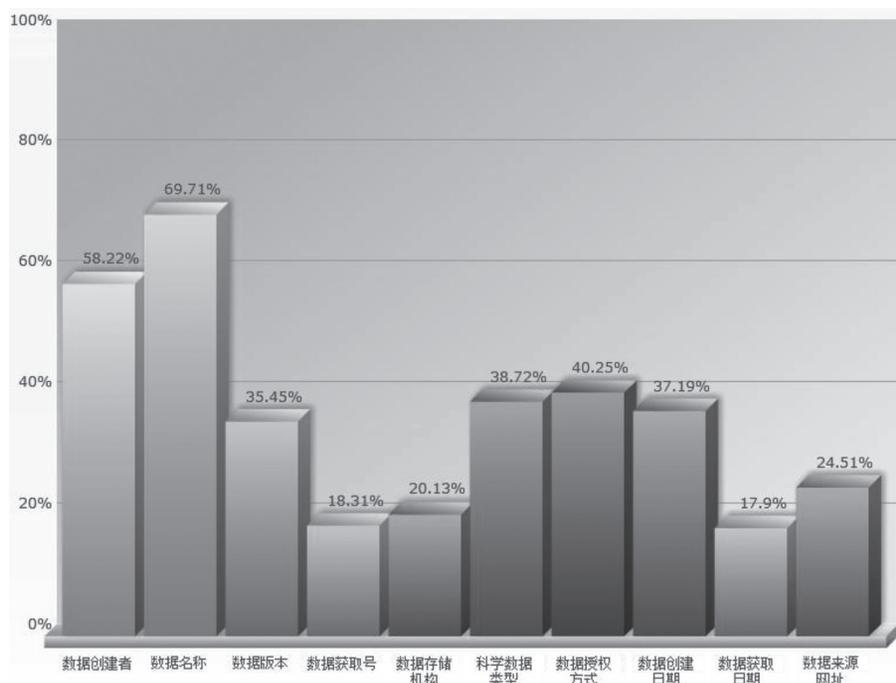


图1 科学数据引用的著录元素序列图

表1 科学数据引用的作用

选项	小计	比例
可以使其他科研人员更好地去验证自己的研究或他人的成果	1107	77.09%
可以使科学数据引用更加快捷，更易检索及访问	998	69.5%
可以获得大量当前还无法实现的新的研究机会	912	63.51%
可以保证数据（或数据相关信息）链接的持久性和唯一性	840	58.5%
可以通过共享和被引用数据获得更高的信誉和认可度	689	47.98%
可以认识领域内的同行	356	24.79%
可以表现出良好的科学素养	298	20.75%
得不到什么直接好处	46	3.2%

4 开展科学数据引用的建议

针对“您对推进在科技期刊中开展科学数据引用有什么建议？”这一问题，在1436份有效问卷中，总共收到267份反馈，筛选出44个无效回答后，得到223个有效回答。在对这223个建议进行归纳总结后，大体可分为以下几类。

（1）开展科学数据引用的价值与意义

开展科学数据引用可以鼓励创新，杜绝抄袭剽窃等学术腐败；引用更加规范化，尊重数据提供者劳动和权利；尽快实现资源共享，体现作者研究成果；能有效遏制学术造假等问题；希望能方便作者、编辑及读者；数据引用是科技期刊的生命所在；对科研工作者的评价方式多元化，增加数据所有者的尊重和引用标注；达到资源共享，真正提高研究水平；研究成果的客观和透明。

（2）针对推进科学数据引用的方式

建议的推进原则是“要顺应形势，逐步试点推广；合理引导，科学规划管理；服务读者，鼓励促进发展”。科学数据引用是一项基础性的工作，需要在充分调研国内外情况的基础上，广泛征求同行和专家意见，全面考虑，建议考虑结合不同学科领域的科技期刊的实际情况，对科学数据库的标准及引用等相关问题分别进行分类处理；自然科学类期刊适用，社会科学类期刊条件成熟时推行；加强政策配套，如建立健全相关法

律法规，实现有法可依，有章可循；建立奖励机制，对实施数据引用的期刊进行奖励（非物质奖励）。

（3）针对数据存缴平台

在期刊中开展科学数据引用，需要建立学科内统一、权威、安全的数据存储平台，提供方便快捷的上传方式，开放编辑部投稿平台与数据存储平台的链接接口；收录进入科学数据库的数据要科学、合理，需要制定统一著录标准和方法，组织多学科专家定期评价，保证数据来源的可靠和数据的真实有效；期刊科学数据引用，保证对数据的及时更新，保证数据的科学性、准确性、实用性，鼓励作者提交数据详细说明；建立数据库可以通过其他途径，比如第三方数据平台整合科学数据，分类存放，集中度越高越好。

（4）针对知识产权

推进科技期刊中开展科学数据引用的意义是重大的，但必须首先建立科学的著作权保护体系，明确权利义务，否则将引发诉讼纠纷，也不利于良好科研风气的培养；数据也应该由原始作者提供并拥有相应权利。

（5）针对标准规范

需要规范的科学数据引用格式与统一的科学数据引用标准，为作者提供常用标准及论文数据引用规范；期刊科学发展，制订相关规范（国标），规范执行标准，对科技期刊编辑规范进行必要的更新或补充，进一步丰富和完善《参考文

献著录规则》国家标准。

(6) 针对实施中可能的困难

我国的法制法规和执行力均不健全,社会诚信缺失和学术腐败的存在,对推进在科技期刊中开展科学数据引用将产生很大的阻碍;科学数据来自于科学研究,这是科学研究成果的一部分,是有知识产权的,是否共享应由有所属权的人决定;数据引用中的数据可靠性是很重要的问题,特别是对于国家、省部级资金资助的项目,尚无明确规定;开展科学数据引用是一项长期的工作,不仅需要技术的发展,还需要人们意识上的跟进,培养作者和期刊,让其认识到开展科学数据引用的重要意义;数据造假较多,实施起来会有困难,提交数据面临着规范化、国际化以及如何保证不会被滥用的问题;在科技期刊中开展科学数据引用会增加作者的工作量及论文篇幅,给作者带来麻烦并造成经济负担。

5 结论

通过本次针对期刊编辑部的“科学数据引用实践及认知现状调研”,我们发现科学数据及科学数据引用对期刊编辑人员而言,还是一个新鲜事物,但各编辑部也已充分意识到了其潜在的应用价值,有542人(占37.74%)认为在论文中添加科学数据引用“肯定是”下一步的发展趋势,对于部分期刊,更是有772人(占53.76%)认可这一趋势。

e-Science时代的科学技术发展,凸现了科学数据的战略重要性。科学数据引用才刚刚开始,不管是引用规范实践,还是数据库平台建设,甚至于科学数据出版等新兴业态的创新,都还需要下一步积极尝试。汤森路透推出的DCI是国际数据库商的积极探索,这无疑吹响了科学数据引用的商业号角,但期刊作为一种重要的出版媒介,无疑将在未来的科学数据引用中占据重要地位。可以预见,未来的科学数据引用必将在传

统出版媒体、新兴网络媒体、数据库内容出版商、公共科学数据库平台以及科研人员等的共同努力下,稳步向前推进。

参考文献

- [1] 朱大明. 参考文献的引用动机[J]. 科技导报, 2013, 31(22): 84.
- [2] 马智峰. 参考文献的引用及影响引用的因素分析[J]. 编辑学报, 2009, 21(2): 23-25.
- [3] GB/T 7714-2005, 文后参考文献著录规则[S]. 2005.
- [4] 钱鹏. 高校科学数据管理研究[D]. 南京: 南京大学, 2012.
- [5] Towards 2020 Science: A Draft Roadmap [EB/OL]. [2013-01-20]. http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/background_overview.htm.
- [6] 张英杰, 彭洁, 张新民, 等. 科研人员引用科学数据的问卷调查[J]. 中国科技资源导刊, 2013(1): 41-46.
- [7] 王卫华, 胡良霖, 沈志宏, 等. 科学数据引用规范的研制[J]. 中国科技资源导刊, 2013(1): 36-40. DOI: 10.3772/j.issn.1674-1544.2013.01.007.
- [8] TR-REC-069 科学数据引用规范[EB/OL]. [2013-01-20]. <http://www.nsd.cn/pronsdchtml/1.compservice.standards/pages/3440.html>.
- [9] 寒区旱区科学数据中心[EB/OL]. [2013-11-20]. <http://westdc.westgis.ac.cn/>.
- [10] 汤森路透发布 Data Citation Index 促进对全球研究数据的发现[EB/OL]. [2013-01-20]. <http://ip-science.thomsonreuters.com.cn/press/press20121102/>.
- [11] 彭洁, 贺德方, 张英杰. 数字出版环境中科学数据引用的实现路径及策略调查分析[J]. 出版发行研究, 2014(4): 57-61.
- [12] 2012年全国新闻出版业基本情况[EB/OL]. [2013-01-20]. <http://www.gapp.gov.cn/govpublic/80/684.shtml>.
- [13] 大数据杂志《GigaScience》正式发刊, 实现文章与数据结合的创新学术发表模式[EB/OL]. [2013-01-20]. <http://www.ebiotrade.com/newsf/2012-7/201271291628859.htm>.
- [14] GigaScience: 基因研究, 数据先行[EB/OL]. [2013-01-20]. <http://www.science-weekly.cn/skhtml-news/2012/1/1658.html>.