

# 国际科学数据出版政策对我国建设科学数据仓储的启示\*

张藤予 丁道劲 余慧 阴姿琦  
(中国科学技术信息研究所, 北京 100038)

**摘要:** 科研和科学创新需要依靠对海量科学数据的分析、管理和重用, 科学数据的出版管理政策对于科学数据的质量控制以及开放共享具有重要作用。采用文献调研法、网络调研法、比较法, 从科学数据提交与存储、质量控制、开放共享以及强制性政策4个方面, 对国际五大学术出版商的科学数据出版政策开展调研和分析。科学数据出版政策直接影响作者如何提交以及选择何种方式共享和存储科学数据, 甚至影响作者对于科学数据仓储的认知和评价。以国际学术出版商的科学数据出版政策为导向, 提出科学数据仓储建设应优先保障数据质量, 打造开放共享的生态环境, 建立与出版商、学术界等的多元融合机制, 并不断完善其标准治理体系。

**关键词:** 科学数据; 数据出版; 出版政策; 科学数据仓储

中图分类号: G23 DOI: 10.3772/j.issn.1673-2286.2023.08.005

**引文格式:** 张藤予, 丁道劲, 余慧, 等. 国际科学数据出版政策对我国建设科学数据仓储的启示[J]. 数字图书馆论坛, 2023 (8) : 44-51.

科学数据是科研和科学创新的基础, 根据国务院办公厅2018年印发的《科学数据管理办法》, 科学数据主要包括在自然科学、工程技术科学等领域, 通过基础研究、应用研究、试验开发等产生的数据, 以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据<sup>[1]</sup>。科学数据出版一般可分为3种模式: 其一是不依靠出版物, 数据在科学数据仓储中独立出版; 其二是数据作为学术论文附属材料出版; 其三是以科学数据为全部出版内容。本文重点探讨第2种出版模式, 即科技期刊关联数据出版。随着海量密集型数据研究的兴起, 学术界对科学数据共享的呼声越来越高, 越来越多的学术出版商及学术期刊出台有关科学数据的政策, 要求作者在提交学术论文的同时存储或提交支撑论文结论的科学数据至学术期刊或指定的科学数据仓储, 使其出版的研究成果更具再验证、可挖掘和重复利用的学术价值<sup>[2]</sup>。为保障读

者尽可能地获取规范的科学数据, 学术出版商对提交内容、提交流程、数据审查要求、存储方式、利用传播以及相关隐私与伦理问题等进行说明<sup>[3]</sup>, 鼓励或强制要求作者将科学数据提交到公共科学数据仓储平台(如Figshare、Dryad等)进行存储和共享。

本文系统调研国内外科学数据出版政策相关研究进展及国际五大学术出版商的科学数据出版政策, 分析其在科学数据质量控制、开放共享等方面的要求。科学数据仓储储存科学数据并与科学数据紧密关联, 其出版政策有着诸多影响, 本文以科学数据出版政策为导向, 使其为我国建设高质量科学数据仓储提供参考。

## 1 科学数据出版政策相关研究

在科学数据出版过程中, 吴立宗等<sup>[4]</sup>提出的数据

收稿日期: 2023-07-07

\*本研究得到中国科学技术信息研究所创新研究基金“全球学术出版活动数据政策与技术标准调研”(编号: MS2023-04)资助。

出版体系框架包含5个基本环节, 秦长江等<sup>[5]</sup>调研分析了Elsevier和Wiley两大学术出版商的政策框架。总结可得, 数据出版体系框架流程主要为数据提交与存储、数据质量控制、开放共享、数据引用等。国内外相关专家、学者围绕科学数据仓储主题的期刊关联数据出版及数据出版体系框架流程的具体内容展开研究。

## 1.1 科学数据出版质量控制政策

Jones等<sup>[6]</sup>通过具体分析Springer Nature和Taylor & Francis两大学术出版商的数据出版政策, 提出学术出版商在制定标准化科学数据政策时需要考量的关键性因素, 包括提高数据的可用性和向作者传达政策的合规性。Vlaeminck等<sup>[7]</sup>以346种经济学和商业研究期刊为样本, 研究期刊获取基础研究数据和计算代码的政策, 指出科学数据出版政策的制定和实施需要时间和经济成本, 商业研究期刊并不完全有意愿制定类似的指导方针和流程。在我国, 方恒等<sup>[8]</sup>将国外科学数据期刊的质量控制政策归纳为描述符格式、引用规范要求、内容类型、同行评议4个方面, 通过对比认为国内科学数据期刊发展相对滞后, 相关政策多为建议或引导性政策, 缺乏强制性。许洁等<sup>[9]</sup>将科学数据出版质量控制分为数据提交、数据存储、数据评审、数据引用和数据评价5个阶段, 并比较研究了*Scientific Data*和《中国科学数据》在存储、评审和引用方面的数据政策, 指出目前数据出版中存在存储设施建设不足、评审体系不健全、数据知识产权侵权严重等问题。傅天珍等<sup>[10]</sup>认为科学数据出版质量控制涉及数据格式、内容、注释以及同行评议等, 并强调注重隐私保护在科学数据出版中的重要性。

在数据出版评审方面, 熊泽润等<sup>[11]</sup>通过调研典型数据期刊, 提出更快建设科学数据出版同行评议标准体系与机制、开发审查工具、加强评审专家激励政策等建议, 从而进一步构建完善的数据出版同行评议体系。屈宝强等<sup>[12]</sup>分析了数据同行评议在3种不同的数据出版模式下的不同特点, 讨论数据同行评议的影响因素、时间成本、工具和系统等关键问题。

## 1.2 科学数据开放共享政策

从科学数据开放共享的影响因素识别角度, Christian等<sup>[13]</sup>通过问卷调查法及内容分析法对编辑和

作者对学术期刊科学数据开放政策的认知和了解情况进行调研, 发现尽管发布数据政策的期刊数量有所增加, 政策要求作者公开报告结果背后的数据, 但作者并不总是这样做, 提交的数据也并不总是符合质量标准。Kim等<sup>[14]</sup>通过在美国进行全国性调查, 指出来自资助机构和期刊的制度压力以及数据存储库和元数据标准的可用性都对数据共享规范产生重大影响, 而数据共享规范对数据共享行为具有重大影响。

从科学数据开放共享的政策内容角度, Hardwicke等<sup>[15]</sup>利用相关分析法和因果推断法, 定量评估了*Cognition*引入强制性开放数据政策对数据可用性、可重用性和可重复性的影响。方恒等<sup>[8]</sup>认为科学数据开放共享政策涉及提交出版、数据确权、出版费用等方面, 并分别总结归纳了国外数据期刊的对应政策。何青芳<sup>[3]</sup>调研了美国、英国、澳大利亚3个典型国家的科学数据公开共享政策, 将学术出版机构发布科学数据共享政策的方式分为科学数据仓储存储和伴随论文同时发表。张晓青等<sup>[16]</sup>指出国外科学数据共享政策内容分为3类: 其一是对数据质量控制、数据获取和开放共享、数据保护和保存、数据科学管理等的政策要求; 其二是对隐私、传统知识、敏感数据、版权及数据所有权的政策规定; 其三是其他描述说明, 如对科学数据开放共享目的、开放共享原则、开放共享范围、开放共享角色和责任、监管和实施的说明。温芳芳<sup>[17]</sup>指出, 国外的科学数据开放共享政策包括科学研究资助政策、数据质量控制与元数据标准政策、数据存储及版权保护政策, 且涉及开放共享方式和时间、数据获取重用、数据共享成本、法律与伦理、基础设施与技术支持、培训与指导等方面, 具有参与主体广泛、公共资助机构发挥领导力、各政策制定者统一意见并相互合作等特点。

综上, 国内外学者围绕科学数据的质量控制政策、开放共享政策等的研究已取得一定进展, 但是研究侧重点各有不同, 分散于科学数据出版的各个环节。科学数据仓储平台是数据作者与科研用户的桥梁, 是科学数据出版的载体。需要综合分析科学数据出版政策, 推进我国本土的科学数据仓储成为受到国际学术出版机构信任的、具有权威性的高水平科学数据仓储。

## 2 国际五大学术出版商的科学数据出版政策分析

按照学术期刊出版规模, Elsevier、Springer Na-

ture、Wiley、Taylor & Francis以及SAGE分列全球五大学术出版商。通过文献、网络调研,对五大学术出版商的数据出版政策要点进行归纳分析,以明确其在科学数据提交、存储、使用、共享等方面的政策要求。

五大学术出版商发布的科学数据出版政策框架主要包括数据提交与存储、数据质量控制、数据开放共享以及政策实施强制性4个部分。

第一部分为科学数据提交与存储,主要包括提交要求和存储位置。出版商在真实性、可用性、格式、引用等方面对作者提交的科学数据提出要求,作者可选择出版商认可和推荐的科学数据仓储。第二部分为科学数据质量控制,主要包括格式校验和同行评议。首先期刊通过在线校验工具等对标识符以及数据格式进行校验,接着同行业审稿人根据期刊指南中的要求对数据进行评审与指导。第三部分为科学数据开放共享,主要包括版权管理和共享方式,作者可选择是否认可通过科学数据仓储实现的数据共享。第四部分为政策实施强制性。出版商制定从宽松到严格的科学数据出版政策,供旗下期刊选择。

## 2.1 科学数据提交与存储政策

### 2.1.1 数据提交要求

(1) 真实性。科学数据不仅承载科研信息,还是科学研究信息来源的一部分,其在科学信息传播、学术活动、科研交流和科学文献管理等方面都有关键的作用。科学数据应真实地反映科学研究活动的过程和结果,真实是对科学数据最基本的要求。为此,国际五大学术出版商制定了相应的政策。Springer Nature要求对电子图像做尽可能少的处理:允许一定程度的图像处理,但图像必须能够正确地表示原始数据,且符合学科领域的相关标准。Wiley明确指出,若发现数据伪造或未经证实的错误,论文可能被撤稿。

(2) 可用性声明。数据可用性声明(Data Availability Statements)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<sup>[18]</sup>。目前,数据可用性声明作为明确数据作者公开数据的简单声明,逐渐成为科学数据政策体系的一部分。Springer Nature指明数据可用性声明应包含的内容:如果数据在存储库中,需包括超链接和持久标识符;如果数据不能共享(为了保护研究参与者的隐私),需

要对此作出解释;还包括研究中生成的原始数据和支持结果与分析的任何二次数据的重用规则。Springer Nature在2016年发布声明称,对所有在其旗下期刊上发表的论文,包括在生命科学期刊上发表的论文,作者均需要提供数据可用性声明。Elsevier旗下期刊在投稿过程中会提示作者直接创建数据可用性声明,并发布于论文大纲的“研究数据”部分。Taylor & Francis要求作者在稿件内附上数据可用性声明,并放置在“参考文献”之前。Wiley、SAGE也均不同程度要求数据作者按照提供的数据可用性声明的规范格式来提交数据可用性声明。

(3) 格式规范性。科学数据出版的前提是规范的数据格式。数据格式规范包括视频/图像/表格文件格式规范、数据论文的具体内容格式规范、数据集文件大小限制、数据丰度以及数据类型规范。科学数据出版的对象既包括广义性数据及所描述的元数据,也包括视频图像、建模模型和报告等,这些不同形式的科学数据能够起到丰富数据内容、佐证科学数据及解释说明的一系列作用。

(4) 引用规范性。论文引用其他数据集、引用本期刊数据集均在数据引用规范范围之内。各学术出版商通过实施FORCE11制定的《数据引用原则联合声明》<sup>[19]</sup>,提倡作者像引用其他论文和网站信息一样引用数据集。引用的数据集成为参考文献的一部分,作者遵循数据引用规范,使用工具标记数据,并使用数据对象标识符关联数据。Springer Nature在期刊指南中指明,在参考文献列表中可引用公开可用的数据集,规范的数据引用需包括一个持久标识符(如DOI、PID)或有指向的链接,以供读者验证。Taylor & Francis、Wiley等也提供了具体数据引用语句规范样例。

### 2.1.2 数据存储位置

国际五大学术出版商要求作者将科学数据提交到自建的科学数据平台或其他可信任的科学数据仓储,多数学术出版商允许作者将科学数据存储至公共科学数据仓储。例如,Springer Nature认可的科学数据仓储包括通用型平台Figshare、Dryad及我国自主科学数据仓储科学数据银行(Science Data Bank, ScienceDB),以及7个学科大类(化学和化学生物学、生物科学、地球环境和空间科学、材料学、健康科学、

物理学、社会科学)内的学科领域型科学数据仓储。Springer Nature仅要求作者将部分类型的数据(包括DNA序列数据、基因组组装数据等)提交给指定公共科学数据仓储,如果没有合适的特定科学数据仓储,作者可以将数据提交给通用数据仓储,包括由大学、科研机构为其附属研究人员提供的任何通用数据仓储。Taylor & Francis向作者提供10个参考的通用数据仓储,作者也可使用FAIRsharing和re3data.org网站进行选择。Wiley并未列出明确的科学数据仓储列表,但也要求作者通过FAIRsharing和re3data.org网站自行检索选择合适的科学数据仓储。Elsevier则在2015年启动开放数据试点项目,后续推出Mendeley Data平台,其与ScienceDirect上的文章链接,通过文章读者可以访问数据集,促进科学数据的开放共享。

## 2.2 科学数据质量控制政策

### 2.2.1 格式校验

科学数据出版的前提是通过便捷的方式把科学数据转换成机器可访问、识别、存储的统一标准格式。Springer Nature在有既定研究团体授权的情况下,对于强制存储的数据进行检查。Elsevier在将数据集纳入仓储库之前,会按照其特定的学科标准来判断和验证数据集的格式是否存在问题,其出版的*Genomics Data*通过Editorial Manager投稿系统控制数据格式:数据作者可以按照数据格式规范要求,利用该投稿系统对已提交的论文进行调整和标准化处理,该系统在任意地点都可通过网络访问。Wiley旗下的数据期刊*Geoscience Data Journal*要求在出版前检查是否分配相应的数字对象标识符(DOI或PID)。

### 2.2.2 同行评议

有效的科学数据质量把控和数据可信性保障需要依靠同行评议。由于科学数据存在读取难、演绎难的特点,对科学数据进行同行评议的难度要比普通论文同行评议大得多,对于科学数据的审查也较为少见。

*Nature*要求数据评审组至少包括1位数据标准评审专业人员,其负责把控科学数据的质量,包括数据集的可重用性、严谨性和完整性,重点评估每个数据描述符和相关数据集的技术质量和完整性。Springer Nature

只选定部分期刊进行科学数据同行评议,重点评审数据可用性声明,评审依据是Springer Nature对数据可用性声明制定的相应政策要求。负责同行评议的人员可以要求查看底层数据或代码,从而对论文的学术水平进行深入评估。Elsevier注重评审数据的时效性,要求期刊编辑在评审阶段对科学数据保密,同时需要初步评估数据的客观性及数据来源的真实性等。Elsevier要求评审人员主要审定数据是否完整、可用,评审不涉及数据的内容和价值。Elsevier认为应侧重审查科学数据的透明度,并认为科学数据论文的数据价值和潜在重用价值是需要重点关注的。

## 2.3 科学数据开放共享政策

### 2.3.1 版权管理

目前开放共享数据资源普遍遵循知识共享许可协议(Creative Commons License,即CC协议)。根据CC协议以及不同的开放程度和开放领域限制,出版机构与数据作者明确说明相应的权利许可情况。协议的限制越少,数据重用的可能性就越大。很多数据期刊由数据作者自由选择不同开放程度的许可协议,如Springer Nature旗下的数据期刊*Scientific Data*由数据作者自行选择包括CC-BY、CC-BY-NC、CC-BY-NC-ND的许可协议。Springer Nature要求数据使用遵循CC0协议,并遵循存储该数据的相应科学数据仓储的许可要求。Elsevier推出的Mendeley Data平台要求研究人员提交原始科学研究数据作为补充文件,其文件遵循CC-BY协议提供,Elsevier旗下大多数数据期刊(如*Data in Brief*)采用CC-BY或CC-BY-NC-ND协议。Taylor & Francis建议作者根据FAIR原则公开提供与其研究相关的所有数据。Wiley出版的部分期刊对提交的科学数据采用CC0协议,且同意科研人员对科学数据进行挖掘和元数据分析。

科学数据的版权归属是版权管理的重要部分,但国际出版商政策中较少出现版权归属声明。大多数期刊由作者保留版权,如:*Scientific Data*声明作者保留所有版权,遵循科学数据仓储相关协议;*Data in Brief*表示作者保留无限制使用数据的权利。少部分期刊,如Taylor & Francis旗下部分期刊默认作者放弃所有权利,作品进入公有领域,并明确声明采用CC0协议。

### 2.3.2 共享方式

科学数据的开放共享对科研的发展起着重要的推进作用。五大学术出版商都不同程度地强调科学数据集的真实性、原创性、科学合理性，并建议遵循现有的开放协议。随着不同科学数据仓储开放共享数据，科学数据出版从收费模式逐渐转变为免费模式。Springer Nature声明首选公共科学数据仓储来存储提交的数据，通过公共科学数据仓储进行开放共享。

## 2.4 科学数据出版政策实施强制性

通过对五大学术出版商的调研，发现Springer Na-

ture、Elsevier所制定的科学数据政策内容相对较为详尽，政策条目较明确，Taylor & Francis、Wiley和SAGE的政策条目相对较少且较为宽松。科学数据政策类型不同、条目不完全相同，但整体看五大学术出版商都采取从鼓励向强制过渡的数据政策，出版数据质量控制以及共享程度要求越来越严格。

以Springer Nature为例，其采取4种类型的数据政策<sup>[20]</sup>，如表1所示。4种类型的数据政策均强制要求“数据通过数据仓储共享”及“允许引用数据”。类型一对其他条目均无要求，类型二建议提供数据标识符和数据可用性声明但不强制，类型三强制作者提供数据标识符和数据可用性声明，类型四对所有条目均有强制要求。

表1 Springer Nature数据政策类型

政策条目	类型一	类型二	类型三	类型四
数据通过数据仓储共享	强制	强制	强制	强制
允许引用数据	强制	强制	强制	强制
数据标识符	无要求	建议	强制	强制
数据可用性声明	无要求	建议	强制	强制
检验数据存储位置	无要求	无要求	建议	强制
数据引用规范	无要求	无要求	建议	强制
数据同行评议	无要求	无要求	建议	强制
出版过程与数据仓储的集成	无要求	无要求	建议	强制

Elsevier采取5种类型的数据政策<sup>[21]</sup>，其对于数据共享的要求比Springer Nature宽松。Elsevier与很多特定领域的数据仓储都有合作关系，可以提供文章和数据库之间的链接。如果投稿时已有数据链接，作者可以在文章或投稿系统里添加数据标识符。如果没有合适的特定领域的数据仓储，作者可以通过Mendeley Data平台共享数据。Taylor & Francis采取5种类型的数据政策<sup>[22]</sup>，其“开放数据+FAIR”政策对使用许可做了特别要求：科学数据需遵循CC-BY、CC0或其他等效协议。Wiley采取4种类型的数据政策<sup>[23]</sup>，SAGE采取3种类型的数据政策<sup>[24]</sup>。相对而言，Taylor & Francis、Wiley、SAGE的政策类型简单、政策相对宽松。

通过调研分析，发现国际学术出版商主要通过规范数据格式与规定存储位置、采用同行评议和格式校验控制质量、制定版权管理政策和规定开放共享方式、实施不同强制性的政策4个策略，构建出较为科学有效的数据出版政策框架。

## 3 对我国建设高水平科学数据仓储的启示

在我国，科学数据的开放共享已经逐步得到重视。一方面，政府和科研机构出台了一系列政策规范：除国务院办公厅印发的《科学数据管理办法》外，还有中共中央办公厅、国务院办公厅印发的《关于进一步弘扬科学家精神加强作风和学风建设的意见》，国家卫生健康委员会、科技部、国家中医药管理局联合发布的《医学科研诚信和相关行为规范》，中国科学院制定实施的《中国科学院科学数据管理与开放共享办法》，中国农业科学院制定的《中国农业科学院农业科学数据管理与开放共享办法》；另一方面，我国陆续建成了一批科学数据仓储，形成“国家人口健康科学数据中心”“国家地球系统科学数据中心”等20个国家级学科领域科学数据中心，涉及生物学、天文学、地震材料、医学、高能物理、农业、林业、计量学、地理学、气象学等领

域<sup>[25]</sup>。科学数据仓储在学术出版主体和数据作者之间起着重要作用, 国际学术出版商对于科学数据的政策重点围绕科学数据仓储主题展开, 随着国际科学数据仓储的发展, 一批具有国际影响力的科学数据仓储逐步形成。聚焦国内, 中国科学院计算机网络信息中心在2015年运行上线并于2021年正式命名的ScienceDB被Springer Nature、Elsevier、Cell Press和美国地球物理学会(American Geophysical Union, AGU)旗下期刊收录到推荐的公共科学数据仓储名单中<sup>[26]</sup>。

大部分科学数据仓储的国际知名度及国际参与度较低, 尚未得到国际知名出版商的认可, 数据关联出版模式的学术可信度、行业认可度都不理想, 有关科学数据存储、数据开放共享、数据版权、数据管理与发表、质量控制等的政策较为模糊, 学术出版主体与科学数据仓储之间的关联度较低, 甚至我国科研用户对其研究领域内的科学数据仓储也没有足够的认知<sup>[27]</sup>。我国科学数据仓储建设应注重以下几个方面, 从而在推动科学进步中发挥更大的作用, 并更好地参与国际学术交流, 提升中国的科学话语权。

### 3.1 坚持数据质量优先

数据结构复杂、数据量增长较快。国际学术出版商在推荐和收录科学数据仓储时, 均以科学数据仓储的质量保障体系为标准, 其服务的安全稳定性是重要评判依据。科学数据仓储不能够一味地追求海量数据而缺乏质量把控。

在收集海量科学数据后, 如何保障科学数据的质量是值得关注的命题。科学数据仓储需要通过标准化、规范化建设, 进行有效的质量保障。除了在元数据规范、格式校验、标识符规范、同行评议、版权管理、引用规范等方面不断借鉴并形成体系政策, 还需持续形成更加健全的行业标准政策。从长远角度看, 要通过科学数据仓储来提高科学数据质量及服务科技创新, 整个科学研究环境中需要形成一个闭环: 从科学数据仓储角度, 科学数据仓储通过提供智能数据处理工具、同行评议辅助工具等, 更为有效地为作者、评议人员、用户提供更加友好、便捷、智能的服务; 从专家同行评议角度, 规范完善同行评议规章制度, 探索同行评议新颖模式, 对科学数据质量进行有效把控; 从数据生产者角度, 保护作者隐私和知识产权, 评估、评价通过共享发表数据产生的效益, 推动科研人员产出高质量科

学数据, 增加其共享数据的动力。

### 3.2 打造开放共享生态

国内的开放共享生态环境还未建立完善, 一是由于国内开放共享观念转换不到位、不成熟; 二是由于科学数据作者及科研人员对于科学数据公开出版模式不认可、不积极; 三是由于在开放共享的环境下版权制度存在缺陷, 同时明确的数据引用机制未建立。多方面导致开放共享生态不完善, 影响数据作者对于传播共享数据的积极性。

开放共享是保障科学数据仓储影响力的有效举措。随着大数据时代的发展, 国家发布一系列政策方针, 有意识地营造数据共享的社会大氛围, 科学数据必然会逐步走向开放共享。《科学数据管理办法》明确“开放为常态, 不开放为例外”的政策原则。国内出版机构应尽快加入开放共享的大环境, 积极遵循《科学数据管理办法》。《科学数据管理办法》明确了责任主体, 促使广大普通的科研人员介入数据开放共享, 无需顾虑承担共享数据的成本, 如此一来带动整个开放生态的繁荣。在开放共享的大环境中, 期刊及学术出版商首先应主动引导数据作者共享数据, 提供简便的数据上传流程及清晰的仓储选择方案。其次, 可在评价机制中加入数据质量、数据引用、数据重用等指标, 通过量化每个数据集产生的影响, 激励数据作者创作并共享数据。最后, 逐渐形成权益保护政策机制, 在数据出版过程中保障数据所有者权益, 打消科研人员顾虑, 提高其共享数据的积极性。

### 3.3 建立多元融合机制

国际知名科学数据仓储在国际组织、学术界、出版商、技术商等利益相关方的共同推动下取得了长足的发展。国际五大学术出版商制定了全面规范的、分层级的、不同强制程度的数据出版政策, 鼓励越来越多的期刊制定并实施数据出版政策, 将不同类型数据期刊推荐给对应的仓储平台。每种期刊应根据自身特点及实际情况遵循特色化的出版政策, 推动数据的公开、共享与传播, 促进科学数据的重用。

各个出版主体需加强合作, 科学数据仓储应与出版机构进行深度合作, 从而推动论文关联数据的出版。首先, 应支持我国自主建设的科学数据仓储发展, 从

源头更好地保障无障碍的合作共建;其次,打通双方平台,使得作者能够一站式提交科研论文及相关数据,便于科学研究人员和数据评审人员撰写和评审数据论文;最后,评审工作分工应更加明确,数据技术性和科学性审查分别由科学数据仓储管理人员和期刊编辑、数据专家来负责,从而提高数据的重用性和科学性。另外,要积极关注并及时跟踪国际数据出版政策新动态,引进和借鉴成功经验,对标国际数据期刊,大力发展我国的数据期刊。根据学科特点和期刊办刊特色,综合考虑数据标准、数据提交与发布流程、格式校验、同行评议、科学数据仓储、数据引用及评估评价等,建立特色化的科学数据出版政策体系。在实践层面,可以采取从小到大、由少及多的相对缓和的政策推广方式,为今后大数据环境下的海量密集型数据研究奠定坚实基础。

### 3.4 完善标准治理体系

与国外相比,我国仅有少数期刊拥有数据出版政策,数据标准化体系未能得到国内科研机构和科研人员的足够关注和重视。目前,我国正向文化强国不断迈进。要成为文化强国首先要成为出版强国,但没有出版业的标准化就没有出版强国。数据出版是出版业的重要部分,建设完善其标准治理体系是必要举措。

根据数据出版领域全球治理的需求,从宏观角度,以目前国内外相关实践为基础,逐步实现对数据出版领域标准化体系的系统谋划、重点推进、广泛参与、全面应用;从微观角度,重点突出推进我国学术期刊关联数据提交的政策要求。研制完善数据出版标准化体系工作,在科研组织层面,是对数据论文规模增长和质量提升的积极响应措施;在科技治理决策层面,是我国实现科学数据出版后发先至,乃至赶超发展的必要基础性工作;在中国式现代化建设战略层面,是实现高水平科学数据共享、推动学术出版高质量发展以及建设科技强国的有力支撑。

## 4 结语

本文梳理与总结了国际五大学术出版商目前的数据出版政策。与传统学术期刊相比,科学数据期刊是新兴出版类型且仍处于发展阶段,数据期刊出版模式的复杂程度更高,故本文的结论仍有一定的局限性。当前,随着开放科学运动的不断推进和全球科研模式的

深刻变革,对科学数据出版及科学数据仓储的关注度逐渐升高,科学数据期刊将推出更为有效的出版政策,从而促进科学数据仓储建设稳步推进。因此,在开放共享的环境下,应注重科学仓储质量控制,建立多元融合机制,完善标准治理体系,并高度关注国际数据期刊和科学数据仓储的发展动态,以持续推进我国科学数据仓储的稳步发展。

### 参考文献

- [1] 科学数据管理办法[EB/OL]. [2023-03-06]. [https://www.gov.cn/zhengce/content/2018-04/02/content\\_5279272.htm](https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm).
- [2] PIWOWAR H A, CHAPMAN W W. A review of journal policies for sharing research data[C]//Proceedings ELPUB 2008 Conference on Electronic Publishing. 2008.
- [3] 何青芳. 国外科学数据管理政策的调查与分析[J]. 上海高校图书情报工作研究, 2016, 26(2): 9-13.
- [4] 吴立宗, 王亮绪, 南卓铜, 等. 科学数据出版现状及其体系框架[J]. 遥感技术与应用, 2013, 28(3): 383-390.
- [5] 秦长江, 吴思洁, 王丹丹. 学术期刊出版机构数据政策框架研究[J]. 出版科学, 2021, 29(6): 46-54.
- [6] JONES L, GRANT R, HRYNASZKIEWICZ I. Implementing publisher policies that inform, support and encourage authors to share data: two case studies[J]. Insights: The UKSG Journal, 2019, 32: 1-11.
- [7] VLAEMINCK S, HERRMANN L K. Data policies and data archives: a new paradigm for academic publishing in economic sciences? [EB/OL]. [2023-03-06]. <https://www.semanticscholar.org/paper/Data-Policies-and-Data-Archives%3A-A-New-Paradigm-for-Vlaeminck-Herrmann/abd81db56d9041b6babca48ba39b17475a87ed51>.
- [8] 方恒, 宋立荣. 国内外期刊机构科学数据出版政策研究[J]. 情报杂志, 2022, 41(3): 181-190.
- [9] 许洁, 唐文辉, 夏心悦. 面向实践的数据出版现状分析与对策探讨[J]. 中国科技期刊研究, 2020, 31(11): 1331-1337.
- [10] 傅天珍, 陈妙贞. 我国学术期刊数据出版政策分析及建议[J]. 中国出版, 2014(23): 31-34.
- [11] 熊泽润, 宋立荣. 科学数据出版中同行评议的问题思考[J]. 中国科技资源导刊, 2022, 54(5): 21-29.
- [12] 屈宝强, 王凯. 数据出版视角下的科学数据同行评议[J]. 图书馆杂志, 2017, 36(10): 71-77.
- [13] CHRISTIAN T M, GOOCH A, VISION T, et al. Journal data

- policies: exploring how the understanding of editors and authors corresponds to the policies themselves[J]. PLoS ONE, 2020, 15 (3): e0230281.
- [14] KIM Y S, BURNS C S. Norms of data sharing in biological sciences: the roles of metadata, data repository, and journal and funding requirements[J]. Journal of Information Science, 2016, 42 (2): 230-245.
- [15] HARDWICKE T E, MATHUR M B, MACDONALD K, et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition[J]. Royal Society Open Science, 2018, 5 (8): 180448.
- [16] 张晓青, 盛小平. 国外科学数据开放共享政策述评[J]. 图书馆论坛, 2018, 38 (8): 147-154.
- [17] 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): 91-101.
- [18] 雷雪. 科技学术期刊数据可用性声明规范应用的现状及建议[J]. 编辑学报, 2021, 33 (3): 271-275.
- [19] MARTONE M. Data citation synthesis group: joint declaration of data citation principles[EB/OL]. [2023-05-18]. <https://doi.org/10.25490/a97f-egykc>.
- [20] Springer Nature. Research data policies[EB/OL]. [2023-03-06]. <https://www.springernature.com/gp/authors/%20research-data-policy>.
- [21] Elsevier. Research data[EB/OL]. [2023-03-06]. <https://www.elsevier.com/about/policies/research-data>.
- [22] Taylor & Francis. Understanding our data sharing policies[EB/OL]. [2023-03-06]. <https://authorservices.taylorandfrancis.com/data-sharing-policies/>.
- [23] Wiley. Wiley's data sharing policies[EB/OL]. [2023-03-06]. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>.
- [24] SAGE Publishing. Research data sharing policies[EB/OL]. [2023-03-06]. <https://uk.sagepub.com/en-gb/asi/research-data-sharing-policies>.
- [25] 无数据、不论文, 但数据要怎么保存? [EB/OL]. [2023-06-23]. <http://zhishifenzi.com/news/multiple/12228.html>.
- [26] Science Data Bank[EB/OL]. [2023-06-28]. <https://www.scidb.cn/>.
- [27] 袁雅琴, 胡晓彦, 佟继周, 等. 大数据开放背景下的我国空间科学数据出版实践[J]. 中国科技资源导刊, 2022, 54 (1): 89-96.

## 作者简介

张藤予, 女, 硕士, 馆员, 研究方向: 信息资源建设、数字出版, E-mail: zhangty@istic.ac.cn.

丁道劲, 女, 博士, 副研究馆员, 研究方向: 数字图书馆、数字出版。

余慧, 女, 硕士, 研究实习员, 研究方向: 信息资源建设。

阴姿琦, 女, 硕士, 研究实习员, 研究方向: 信息资源建设。

Enlightenment of International Scientific Data Publishing Policies on Building Scientific Data Warehousing in China

ZHANG TengYu DING QiuJin YU Hui YIN ZiQi  
(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract: Scientific research and innovation rely on the analysis, management, and reuse of massive scientific data. The publication and management policies of scientific data play an important role in the quality control and open sharing of scientific data. This article adopts literature research, network research, and comparative methods to conduct research and analysis on the scientific data publishing policies of the five major international academic publishers from four aspects: scientific data submission and storage, quality control, open sharing, and mandatory policies. The policy of scientific data publishing directly affects how authors submit, share, and store scientific data, and even affects authors' understanding and evaluation of scientific data warehousing. This article is guided by the scientific data publishing policies of international academic publishers, and proposes that the construction of scientific data warehousing should prioritize ensuring data quality, creating an open and shared ecological environment, establishing a diversified integration mechanism with publishers and academia, and continuously improving its standard governance system.

Keywords: Scientific Data; Data Publishing; Publication Policy; Scientific Data Warehousing

(责任编辑: 王玮)