

国内外数据期刊实践与进展研究

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14.11-6035.ncdc.2023.0001.zh

文献 DOI:

10.11922/11-6035.ncdc.2023.0001.zh

文献分类: 信息科学

收稿日期: 2023-02-10

开放同评: 2023-02-10

录用日期: 2023-03-20

发表日期: 2023-03-30

刁妍^{1*}, 孔丽华¹, 汪洋¹, 陈枢舒¹

1. 中国科学院计算机网络信息中心, 北京 100083

摘要: 数据出版呈现蓬勃发展态势, 数据期刊成为广泛实践与探索数据开放共享的重要手段。本研究聚焦具有代表性、相对成熟、不同学科范围的数据期刊, 针对数据期刊的出版实践层面, 统计了数据期刊数据论文出版量及占比情况, 并结合多年来从事数据出版工作的实践经验, 深入研究数据出版政策的制定细则和落实情况。结果表明, 在数据使用许可、数据存储和永久标识符、论文与数据关联访问、数据评审、规范数据引用等方面, 数据期刊的规定普遍较为完善, 但数据更新实践案例较少, 对数据安全和数据版权重视不足。为助力数据期刊发展, 提出了数据出版实践中需要考虑的可操作、可借鉴的细节。

关键词: 数据期刊; 数据出版; 数据版权; 数据存储; 数据引用; 数据安全; 数据更新

引言

科学数据是检验科学研究价值的天然试金石, 是重要的战略性资源^[1]。科学数据共享在检验学术真伪、增加科研成果评价渠道、增强数据重用机会、推进跨学科交流合作、有效提升论文影响力等重要性越来越受到广泛关注。公共项目资助机构、国家政府机关、科研院所等机构的数据管理计划和政策的出台如雨后春笋, 为科学数据出版的发展提供了良好生态。

独立数据出版(在数据中心或数据存储库存储发布, 不依赖出版物的数据发布)、作为论文辅助资料的数据出版(附属于出版物的数据发布)、以数据论文形式出版(作为出版物本身的数据出版)是数据出版的3种主要模式^[2]。其中, 数据论文形式的数据出版, 相比另外两种出版模式, 以论文形式组织对数据采集和生产加工过程、数据的组成结构、数据的质量控制和评估方法、数据价值、数据使用方法以及建议等, 给出了更为深入和丰富的阐述, 经过专家同评后正式出版数据论文, 更有利于对数据的可靠性、可用性、可追溯性进行验证, 从而为数据的引用和重用提供更大的可能性和便利性。

数据期刊是指出版数据论文的期刊。基于 Candela 等(2015年)^[3]、刘凤红(2019年)^[4]、Walters W H(2020年)^[5]、王卫军等(2021年)^[6]研究成果统计: 目前全球正式出版的数据期刊在160种以上, 其中纯粹型数据期刊(即出版的论文中数据论文占比50%以上)在20种以上, 分属超过20家出版机构。数据期刊整体数量呈现增长趋势, 涵盖了85个科学领域, 包括生态学, 遗传学, 生物多样性保护, 公众、环境和职业健康, 生物化学与分子生物学, 数学与计算生物学, 植物科学等多个综合领域。

* 论文通信作者

刁妍: xiyan@cnic.cn

自 2009 年首本以数据论文为主要出版任务的纯粹型数据期刊 *Earth System Science Data* 创办, 至今已有 14 年。期间, 国内外出版机构遵循数据 FAIR 原则 (《FAIR Data Principles》, 2016 年)^[7], 参考开放科学中心 (Center for Open Science, COS) 学术开放获取出版先驱提出的《期刊透明性与开放性指南》(《Transparency and Openness Promotion (TOP) Guidelines》, 简称《TOP 指南》, 2018 年)^[8]、国际数据委员会 (Committee on Data for Science and Technology, CODATA) 提出的《研究数据管理政策的当前最佳实践》(《Current Best Practice for Research Data Management Policies》, 2015 年) 等指南, 确保数据的可发现 (Findable)、可访问 (Accessible)、可互操作 (Interoperable)、可重用 (Reusable)。同时, 依据由全球土著数据联盟 (Global Indigenous Data Alliance, GIDA) 制定的土著数据治理原则 (The CARE Principles for Indigenous Data Governance, CARE), 将集体收益 (collective benefit)、质量保证 (authority to control)、责任 (responsibility)、伦理 (ethics) 原则作为数据出版实践的伦理补充^[9-10], 开展了大量的数据论文出版实践^[11-16]。本研究侧重从数据期刊的出版实践层面, 结合多年来从事数据出版的实践经验, 研究数据期刊数据论文出版情况, 深入调研和分析数据出版相关政策细则制定和落实情况, 为期刊开展数据出版实践提供先行做法和成败经验。

1 研究方法和思路

本研究选取具有代表性、相对成熟、不同学科范围的 25 本国内外数据期刊为研究对象, 首先通过期刊官网调研了数据期刊的基本情况 (见表 1); 随后以上述期刊官网及 DOAJ (Directory of Open Access Journals) 数据库作为数据来源, 对其数据论文出版情况 (包括创刊时间或开始刊发数据论文的时间、数据论文发文量占期刊总发文量的比例、数据论文的出版量以及年度出版变化趋势) 进行统计 (其中 *Journal of Chemical and Engineering Data* 由于统计源数据访问受限, 为保证调研数据的准确性, 暂不计入本次统计范畴); 再从 25 本调研期刊中筛选出分属不同学科和出版机构的 15 本期刊, 依据其官网相关文字描述、相关文献以及实施方式, 深入研究其数据版权、数据存储、数据评审、数据引用、数据安全、数据更新等数据出版相关政策细则制定和落实情况; 然后针对其中存在问题与挑战进行探讨, 并结合数据期刊出版现状及其数据出版实践进行分析, 以期为期开展数据出版实践提供可操作、可借鉴的细节。本文中所指的纯粹型期刊为从期刊发表数据论文起始年份统计至 2022 年年底, 累计数据论文发文量与累计总发文量的比值超过 50% 的期刊, 混合型期刊则为占比低于 50% 的期刊。

表 1 调研期刊基本情况表

Table 1 Basic information of the data journals involved in this study

期刊名称	主办/出版机构	学科范围	是否 OA	创刊时间 (年)	出版状态	期刊类型	官方网址
<i>Journal of Chemical and Engineering Data</i>	American Chemical Society	材料科学	否	1956	正常	*	https://pubs.acs.org/journal/jceaax

期刊名称	主办/出版机构	学科范围	是否 OA	创刊时间 (年)	出版 状态	期刊类 型	官方网址
<i>BMC Genetics</i>	Springer Nature	遗传学	是	2000 (2017 年 12 月始发 数据论 文)	正常	混合型	https://www.ncbi.nlm.nih.gov/pmc/journals/3993/
<i>BMC Research Note</i>	Springer Nature	生命科学	是	2008 (2017 年 12 月始发 数据论 文)	正常	混合型	https://bmresnotes.biomedcentral.com/
<i>ZooKeys</i>	Pensoft	动物分类 学	是	2008 (2011 年 始发数据 论文)	正常	混合型	https://zookeys.pensoft.net/
<i>Earth System Science Data</i>	Copernicus Publications	地球科学	是	2009	正常	纯粹型	https://www.earth-system-science-data.net/home.html
<i>PhytoKeys</i>	Pensoft	植物分类 学	是	2010 (2012 年 始发数据 论文)	正常	混合型	https://phytokeys.pensoft.net/
<i>GigaScience</i>	华大基因、 Oxford University Press	基因组 学、生命 科学	是	2012	正常	混合型	https://academic.oup.com/gigascience
<i>F1000 Research</i>	F1000 Research Ltd (Taylor & Francis 成员)	生命科学/ 医学	是	2012	正常	混合型	https://f1000research.com/
<i>Journal of Open Archaeology Data</i>	Ubiquity Press	考古学	是	2012	正常	纯粹型	https://openarchaeologydata.metajnl.com/
<i>Biodiversity Data Journal</i>	Pensoft	生物多样性	是	2013	正常	混合型	https://bdj.pensoft.net/
<i>Genomics Data</i>	Elsevier	基因组学	是	2013 (2015 年)	2018 年后 并入	混合型	https://www.sciencedirect.com/journal/genomics-data

期刊名称	主办/出版机构	学科范围	是否 OA	创刊时间 (年)	出版 状态	期刊类 型	官方网址
				始发数据 论文)	Data in Brief		
<i>Geoscience Data Journal</i>	Wiley	地球科学	是	2013	正常	纯粹型	https://rmets.onlinelibrary.wiley.com/journal/20496060
<i>Dataset Papers in Science</i>	Hindawi	综合	是	2013	2017 年停 刊	纯粹型	https://www.hindawi.com/journals/dpis/contents/
<i>Journal of Open Psychology Data</i>	Ubiquity Press	心理学	是	2013	正常	纯粹型	https://openpsychologydata.metajnl.com/
<i>Open Health Data</i>	Ubiquity Press	健康和医学	是	2013	正常	纯粹型	https://openhealthdata.metajnl.com/
<i>Data in Brief</i>	Elsevier	综合	是	2014	正常	纯粹型	https://journals.elsevier.com/data-in-brief
<i>Scientific Data</i>	Springer Nature	综合	是	2014	正常	纯粹型	https://www.nature.com/sdata/
中国科学数据	中国科学院	综合	是	2015	正常	纯粹型	http://www.csdata.org/
<i>Journal of Open Humanities Data</i>	Ubiquity Press	人文	是	2015	正常	纯粹型	https://openhumanitiesdata.metajnl.com/
<i>Open Data Journal for Agricultural Research</i>	多所大学和研究 基金会	农业和食 品安全	是	2015	2022 年未 出刊	纯粹型	https://odjar.org/
<i>IUCrData</i>	International Union of Crystallography	结晶学及 相关领域	是	2016	正常	纯粹型	https://iucrdata.iucr.org/
<i>Chemical Data Collections</i>	Elsevier	化学	否	2016	正常	纯粹型	https://www.sciencedirect.com/journal/chemical-data-collections

期刊名称	主办/出版机构	学科范围	是否 OA	创刊时间 (年)	出版 状态	期刊类 型	官方网址
<i>Data</i>	Multidisciplinary Digital Publishing Institute (MDPI)	自然科学 及部分社 会科学	是	2016	正常	纯粹型	https://www.mdpi.com/journal/data
全球变化数 据学报	中国科学院	地球科学	是	2017	正常	纯粹型	http://www.geodoi.ac.cn/
地球大数据 (<i>Big Earth Data</i>)	中国科学院、 Taylor & Francis	地球科学	是	2017	正常	混合型	https://www.tandfonline.com/journals/tbed20

注：“状态”一栏的调研时间为2023年1月17日。“*”指由于该刊为非开放获取期刊，无精确统计数据划分纯粹型或混合型。“OA”指Open Access（开放获取）。

2 数据期刊出版现状

从2009年起，数据论文的出版开始受到期刊界的广泛瞩目；2012年开始进入纯粹型数据期刊的集中创刊期，也有一些传统研究型期刊加入数据论文的出版。至2022年，数据期刊的发展呈现多极态势，有的稳步上升，有的波动较大，有的则呈现下降趋势甚至合并、停刊。

2.1 创刊时间

数据出版随着信息技术及科研范式的转变应运而生。从图1可见，2013–2016年为纯粹型数据期刊较为集中的创办期，调研的15本纯粹型数据期刊中，有12本于该阶段创办。

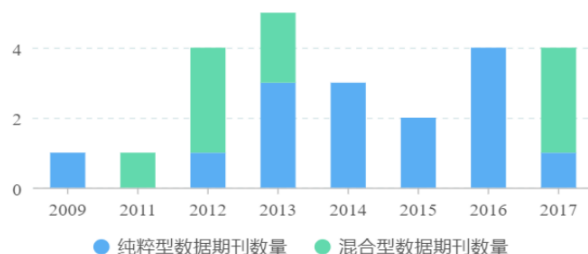


图1 数据期刊刊发数据论文的年份

Figure 1 The years in which the data journal published data papers

2.2 数据论文发文量占比

图2统计了数据期刊出版数据论文占总发文量的比例，纯粹型数据期刊主要集中在80%–100%。其中 *Journal of Open Archaeology Data*、*Dataset Papers in Science* 的数据论文占比达100%，其次占比95%以上的有 *IUCrData* (99.88%)、*Data in Brief* (99.76%)、*Chemical Data Collections* (96.41%)、*Open Health Data* (96.30%)、*Open Data Journal for Agricultural Research* (96.3%)、《中国科学数据》(95.30%)。混合型数据期刊中有一半以上数据论文占比在5%以下，其他主要集中在15%–30%的区间。

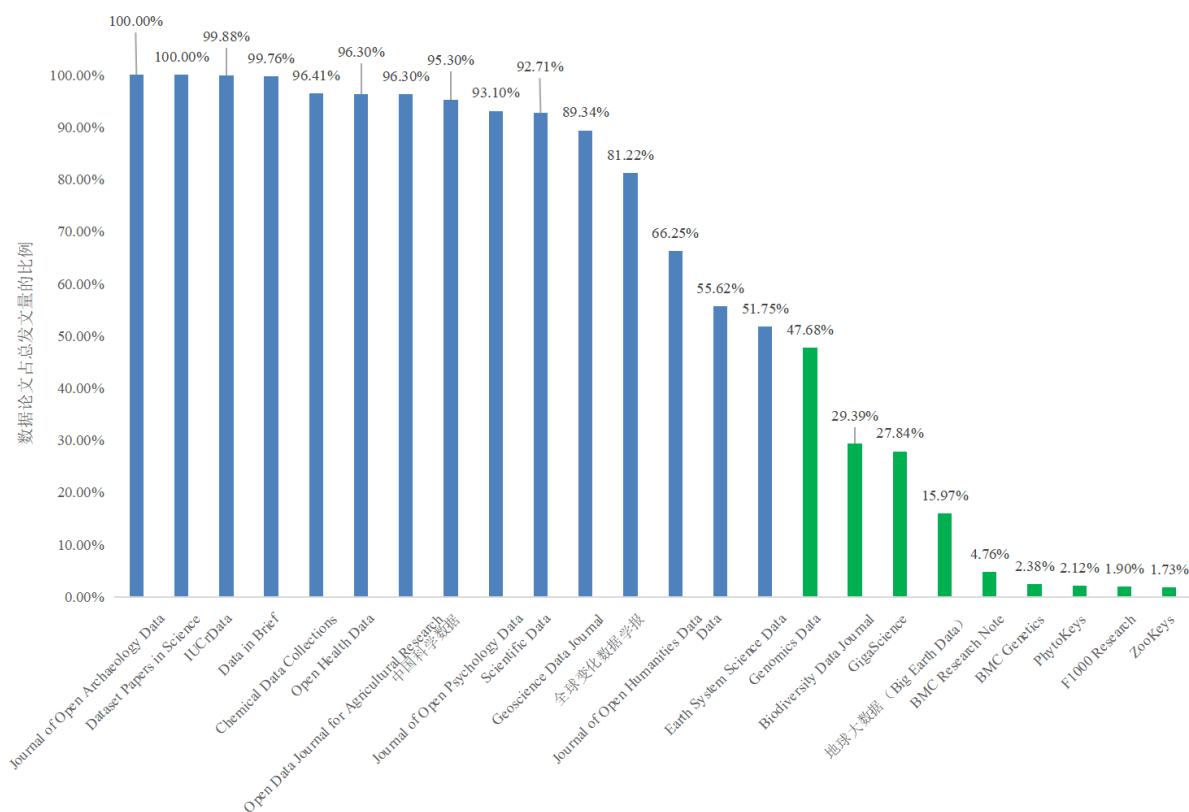


图2 数据期刊刊发数据论文占总发文量的比例

Figure 2 Ratio of published data papers to total number of published papers in data journals

2.3 数据论文出版量

图3所示,从数据期刊刊发数据论文的年份算起至2022年(如停刊,则统计截止到停刊年份),年均数据论文发文量最高的是 *Data in Brief*, 高达866.00篇;其次是 *Scientific Data* (258.44篇)、*IUCrData* (232.57篇)和 *Genomics Data* (96.00篇);国内两本纯粹型数据期刊数据论文发文量也较为靠前。其中, *Genomics Data* 于2018年并入 *Data in Brief*, *Dataset Papers in Science* 于2017年停刊, *Open Data Journal for Agricultural Research* 在2022年未出刊。

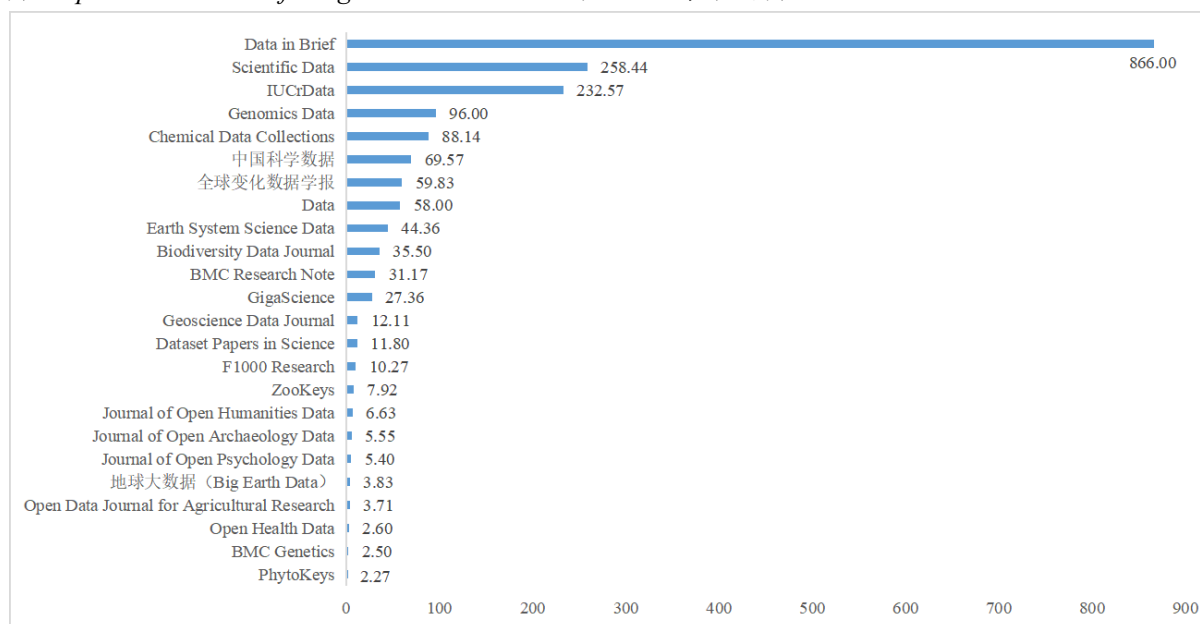


图3 数据期刊年均数据论文发文量(限刊发数据论文期内)

Figure 3 Annual average numbers of data papers published in data journals since the publication of data papers

2.4 年度出版变化趋势

图 4 显示, 从纯粹型数据期刊出版的数据论文占比的年度变化来看, 绝大多数期刊在经过初创期震荡之后, 中后期数据论文占比基本维持在一个相对稳定的水平, 例如 *Data in Brief* 维持在 99% 以上, *Chemical Data Collections*、*Scientific Data*、*IUCrData*、《中国科学数据》等维持在 90% 以上, *Data*、*Earth System Science Data*、*Journal of Open Humanities Data* 维持在 50%–70% 的范围内。此外, *Journal of Open Archaeology Data*、*Dataset Papers in Science* 在出刊的年份, 数据论文占比均在 100%。

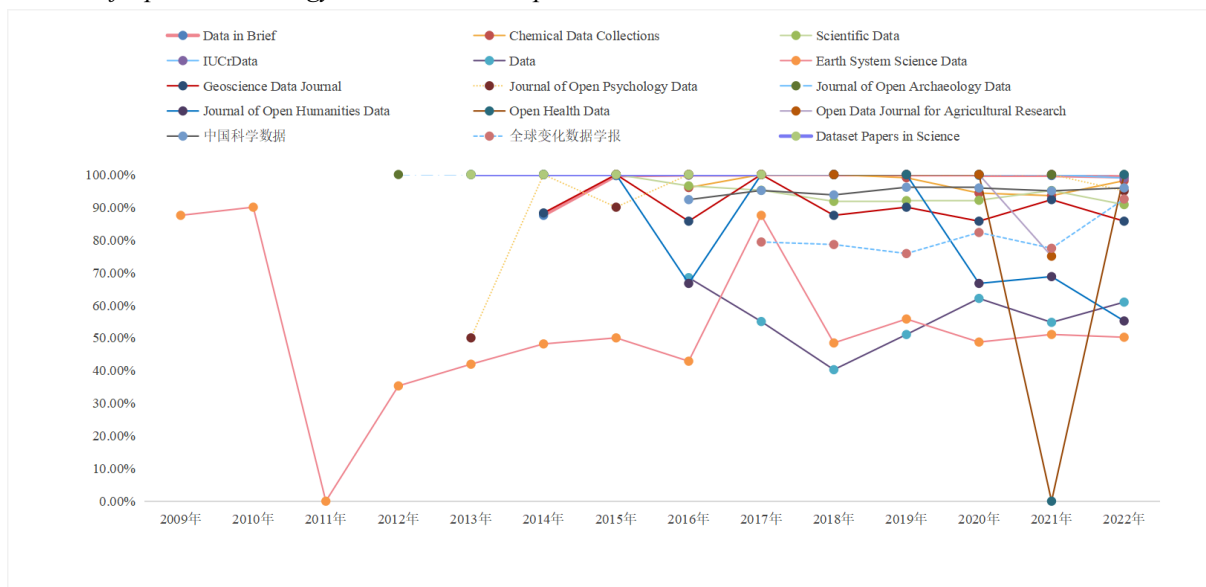


图 4 纯粹型数据期刊数据论文占比的年变化

Figure 4 Annual changes of the proportion of data papers in pure data journals

混合型数据期刊中, *GigaScience* 的数据论文占比呈现先上升、后下降直至稳定的趋势, 从 2012 年的 5.56% 增加到 2017 年的 40.94%, 自 2018 年以后趋于稳定, 维持在 25% 左右; *Genomics Data* 从混合型逐渐转型为纯粹型; *BMC Genetics*、地球大数据 (*Big Earth Data*) 的数据论文占比呈现小幅上升趋势; *F1000 Research* 则维持在 2% 左右 (图 5)。

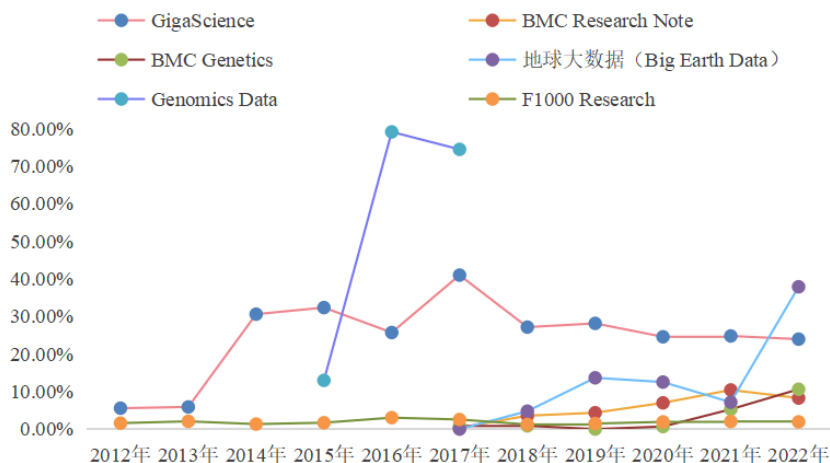


图 5 混合型数据期刊数据论文占比的年变化

Figure 5 Annual changes of the proportion of data papers in mixed data journals

数据论文出版量较大的期刊发文量波动较大, *Scientific Data* 呈现明显的上升趋势, 其 2022 年数据论文出版量比 2021 年增加了 153.68%; *IUCrData* 呈现下降趋势; *Data in Brief* 发文量在 2015 年后远超其他期刊, 不同年份的数据论文出版量有增有减, 年度变化波动较大 (图 6)。

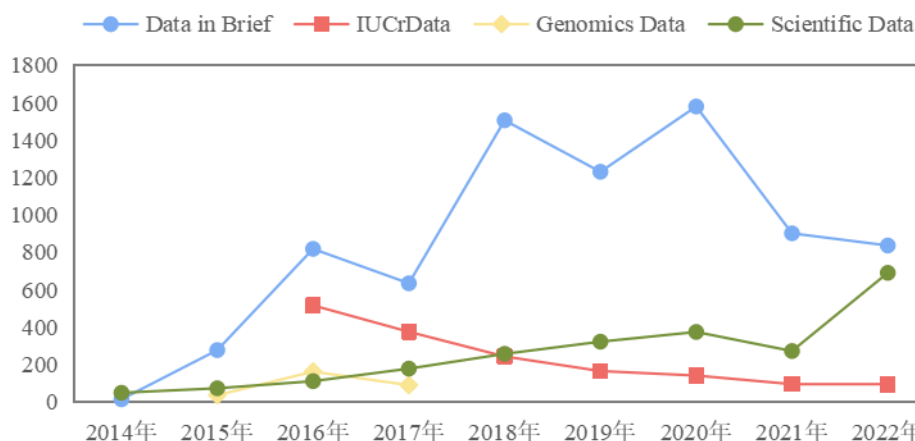


图6 发文量较大的数据期刊数据论文出版量年变化

Figure 6 Annual volume changes of data papers in data journals with a large number of publications

数据论文出版量相对较小的期刊中，《Earth System Science Data》、《中国科学数据》、*Data*、*F1000 Research*、*Journal of Open Humanities Data*、《地球大数据（Big Earth Data）》、*BMC Genetics* 等期刊整体上看，数据论文发文量处于上升期；*Chemical Data Collections*、*BMC Research Note*、*GigaScience* 呈现先上升、后下降的趋势；《全球变化数据学报》发文相对平稳；*Open Health Data*、*Dataset Papers in Science* 呈下降趋势；*Journal of Open Archaeology Data*、*Geoscience Data Journal*、*Journal of Open Psychology Data* 年度波动较大；此外，《Earth System Science Data》、*Open Health Data* 均在个别年份出现未刊发数据论文的情况；*Open Data Journal for Agricultural Research* 发文量较小且呈下降趋势，2019、2022 年均出现未出刊的情况（图 7）。

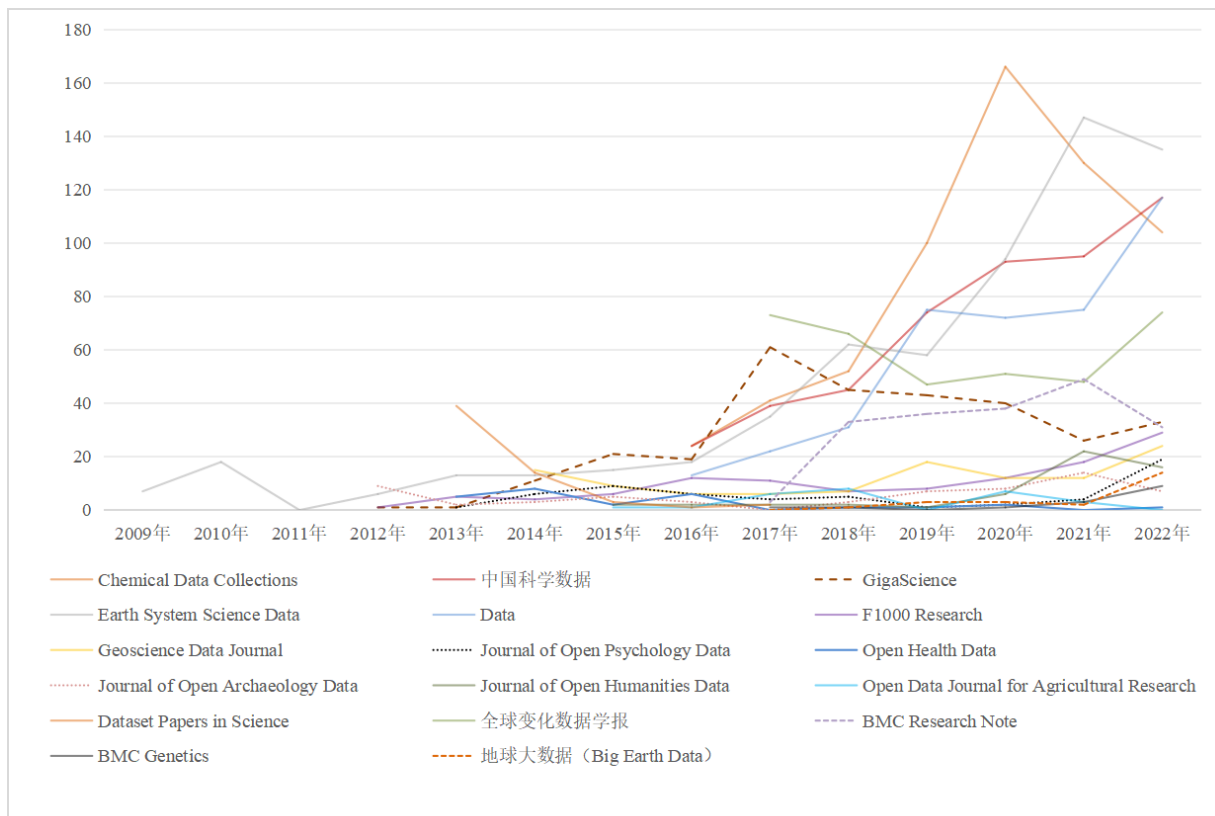


图7 发文量相对较小的数据期刊数据论文出版量年变化

Figure 7 Annual volume changes of data papers in data journals with a relatively small number of publications

数据论文发文量和占比的年度变化规律反映了各数据期刊的出版策略、出版趋势的差异。在明确期刊发展定位后，数据期刊普遍倾向于将数据论文的出版占比维持在相对恒定的数值，呈现出以

下 4 种出版策略：一是定位为纯数据期刊，只接收数据论文，其他发布内容仅限编者按（Editorial）、勘误（Correction）等非学术论文，如出版量较小的 *Journal of Open Archaeology Data*、*Dataset Papers in Science*，出版量较大的 *IUCrData*、*Data in Brief* 等。二是纯度较高的数据期刊，主要接收数据论文，兼发布少量研究论文（Research articles）、评论（Comment）、数据服务（Data Services Paper）、综述（Review Article）等类型文章，如出版量较小的《中国科学数据》、*Geoscience Data Journal*，出版量较大的 *Scientific Data*、*Genomics Data* 等。三是纯度较低的数据期刊，以接收数据论文为辅，有计划地主动将数据论文的占比控制在较低比例，如 *GigaScience*、*BMC Research Note* 等。四是偶发数据论文的期刊，数据论文占比较小，出版量不定、随机性较大，如 *F1000 Research*。同时可以看出数据期刊以下 3 类出版趋势：一是数据论文出版量逐渐萎缩甚至衰败的期刊，如 *Dataset Papers in Science*、*IUCrData*。*Dataset Papers in Science* 的数据论文出版量逐年减少，2013 年创刊、2018 年停刊，创刊高峰时期数据论文的出版量为 39 篇，停刊前的 3 年内每年出版量不足 3 篇。*IUCrData* 的出版量也在逐年减少，从高峰时期的 516 篇下降到 2022 年的 93 篇。二是发展势弱的期刊，如 *Journal of Open Archaeology Data*、*Open Health Data*、*Open Data Journal for Agricultural Research* 等数据期刊虽纯度高，但是稿源稳定性欠佳，时高时低、时有时无。三是强势发展的期刊，如 *Earth System Science Data*、*Scientific Data*、*Data*、*Data in Brief*、《中国科学数据》。数据论文的出版量整体出现持续增长的趋势。一般情况下，稿源的稳定性与期刊的质量和投入、出版机构的实力和支持程度密切相关。*Earth System Science Data* 是纯粹型数据期刊中创办最早及 2022 年影响因子最高的期刊，通过持续良好的发展，吸引了越来越多的优质稿源。*Scientific Data* 的 2022 年影响因子在纯粹型数据期刊中位列第二，出版机构为 Springer Nature，创刊即成为焦点，出版机构成熟的办刊模式增进了该刊的持续成长。*Data* 是国际知名学术开放获取出版先驱 MDPI 强力打造的数据期刊，数据论文出版量稳定增长。《中国科学数据》通过不断组织优秀数据集专题的出版，保证数据论文出版的质量和数量；多渠道多平台传播、优质服务和强强联合等方式汇聚各方优势资源，从而实现稳定的增长。

3 数据期刊出版实践

数据论文是用来描述数据的论文，在形式上与传统研究型论文相似，本质上又与后者存在较大差异。通常数据期刊出版实践中除了会面临与传统期刊相同的出版伦理和学术不端、OA 类型、是否免费下载，论文的归档、索引和长期保存方案，学科领域内的特殊要求等方面的抉择外，还需考虑科学数据及数据论文的科普、数据唯一永久标识符、数据可用性声明、数据版权、数据评审、数据引用、数据安全、数据更新处理流程等方面的问题。以下是 15 种调研期刊的相关情况。

3.1 数据使用许可和版权

共享协议对于保护科学数据权属，明确数据的重用目的和方式至关重要。数据论文正式出版之前，作者应明确数据的权属关系，数据的使用条件（包括是否保留数据署名权、是否允许商用、是否要求在本数据基础上生产的新数据采用本数据的共享方式）等，以便对数据采取不同的开放共享类型。

3.1.1 数据使用许可类型

开放知识基金会（Open Knowledge Foundation）汇总了当前被广泛使用的适用于数据的共享协议，即使用许可证（License），主要有表 2 中的以下几种（<http://opendefinition.org/licenses/#Data>）。包含 BY（Attribution）的使用许可证均需保留数据作者署名权或标明数据来源，SA（ShareAlike）是指衍生数据应遵循原数据的使用许可方式继续共享。其中，CC0 1.0、PDDL 1.0 是最宽松的协议，视为数据作者放弃所有权利，数据进入公有领域；CC-BY 4.0、ODC-By 1.0 均要求使用者在尊重数据作者署名权的前提下使用数据；CC-BY-SA 4.0、ODbL 1.0 则在前述使用许可证的基础上增加了要求

新数据以原数据使用许可证进行共享的限制。

表 2 适用于数据的使用许可证
Table 2 Licenses suitable for data

简称	英文全称	署名权 (BY)	相同方式 共享 (SA)	内容概要
CC0 1.0	Creative Commons CCZero	放弃	放弃	放弃所有权利，数据进入公有领域
PDDL 1.0	Open Data Commons Public Domain Dedication and Licence	放弃	放弃	放弃所有权利，数据进入公有领域
CC-BY 4.0	Creative Commons Attribution 4.0	保留	放弃	允许使用者在尊重数据作者署名权的前提下以任何媒介或格式共享（复制、分发），以任何目的（含商业目的）改编（修改、转换）数据
ODC-By 1.0	Open Data Commons Attribution License	保留	放弃	适用于数据库，在标明数据库来源的前提下，可以共享（复制、分发和使用），创建，改编（修改、转换和构建）新数据或数据库
CC-BY-SA 4.0	Creative Commons Attribution Share-Alike 4.0	保留	保留	允许使用者在尊重数据作者署名权、新数据采用原数据相同方式共享的前提下以任何媒介或格式共享（复制、分发），以任何目的（含商业目的）改编（修改、转换）数据
ODbL 1.0	Open Data Commons Open Database License	保留	保留	适用于数据库，在标明数据库来源、新数据采用原数据相同方式共享的前提下，可以共享（复制、分发和使用），创建，改编（修改、转换和构建）新数据或数据库

CC0 1.0、CC-BY 4.0、CC-BY-SA 4.0 均属 CC 系列知识共享许可协议，此外还有非商业性使用（Non Commercial, NC）、禁止演绎（No Derivs, ND）授权要素，CC BY-NC 4.0、CC BY-NC-SA 4.0、CC BY-ND 4.0、CC BY-NC-ND 4.0 协议，共计 4 个授权要素，7 种许可类型。

《潘顿原则》（《Panton Principles》）强烈禁止使用限制商业再利用或通过排除为特定目的或特定个人或组织使用而限制衍生数据的许可证。认为这些许可证使数据无法有效地集成和重用，也无法防止用于支持数据保存的商业活动。强烈建议科学数据，特别是公共资助的数据，通过使用公共领域专用许可或知识共享零豁免（强烈建议 PDDL 或 CC0）将已发表科学的基础数据明确地投入公共领域，并确保遵守《实现开放获取数据的科学共享协议》（Science Commons Protocol for Implementing Open Access Data）和开放知识/数据定义^[17]。《关于科学数据的北京宣言》（The Beijing Declaration on Research Data）也主张不应该对公共资金资助的科学数据的重用施加任何限制，让研究人员选择最低限制性的许可证^[18]。

数据出版机构和存储库一般自行选用和推荐适合的使用许可证，如 Pensoft 推荐使用 ODC-By

1.0、CC0 1.0、PDDL 1.0^[19], Elsevier 指定的 Mendeley 数据库提供了 CC0 1.0、CC-BY 4.0、CC BY-NC 4.0 选项^[20], 科学数据银行提供了 CC0、CC-BY 4.0、CC BY-SA 4.0、CC BY-NC 4.0、CC BY-NC-SA 4.0、CC BY-ND 4.0、CC BY-NC-ND 4.0 和 ODbL, 共计 8 种数据许可协议^[21]。

3.1.2 期刊数据使用许可和版权情况

调研发现, *IUCrData*、*Geoscience Data Journal* 等期刊仅关注论文使用许可, 未提及或未对数据的使用许可进行要求。要求使用 Open Data Commons Attribution License (CC) 系列开放许可协议的期刊较为普遍, 如 *Scientific Data*、*Earth System Science Data*、《中国科学数据》等。其中, *Data in Brief*、*Biodiversity Data Journal*、*GigaScience*、*Journal of Open Archaeology Data* 等期刊采用了较为严苛的作者弃权许可协议, 如 Open Data Commons Attribution License (ODC-By 1.0)、CC0、Open Data Commons Public Domain Dedication and Licence (PDDL 1.0)。但是, 在实践中, 如 *F1000 Research* 虽然要求数据许可使用 CC0, 实际操作中允许使用 CC-BY 4.0 等协议。

在大多数期刊管理办法中, 重点关注论文的版权保护, 尤其是国外期刊较少提及数据的版权保护相关内容。其中, *Data in Brief* 在版权政策提到了作者可保留无限制使用数据的权利, *Earth System Science Data* 提到作者有权复制数据。国内数据期刊相对对数据版权的重视程度较高, 《中国科学数据》提到版权归作者所有, 作者授权期刊在全世界范围内复制权、发行权、信息网络传播权、翻译权及汇编权等权利, 同时将上述权利再授权给第三方使用的权利; 《全球变化数据学报》的规定数据作者拥有自主知识产权、永久署名权, 将数据的编辑权、不同介质复制权、印刷权、数据散发权、网络传播权、存储权、多语种翻译权和上述产权的转让权以及相关的其他权利与数据出版者分别共有, 双方互不干涉数据出版后的数据管理和共享方式, 数据出版权由数据出版者独有^[22]。而 *IUCrData* 更多从利于数据自由传播、用于研究目的的角度考虑, 认为版权保护不应扩展到科学数据文件, 只要适当地注明数据来源即可 (表 3)。

表 3 调研期刊的数据使用许可和版权

Table 3 Data use license and copyright of journals involved in this study

期刊名称	数据使用许可	数据版权
<i>Data in Brief</i>	未提及	作者保留无限制使用数据的权利
<i>Scientific Data</i>	大多数情况下使用 CC0、CC-BY 或同等许可	存储在数据存储库中的数据遵循存储库相关协议, 此外 Springer Nature 不单独对提交的数据集进行版权声明
<i>IUCrData</i>	未提及	版权保护不扩展到科学数据文件, 只要适当地注明获得这些数据的来源。
<i>Earth System Science Data</i>	在 CC0 或 CC-BY 等非限制性许可下可用	作者有权复制
中国科学数据	推荐使用 CC-BY 4.0	版权归作者所有, 作者授权期刊在全世界范围内复制权、发行权、信息网络传播权、翻译权及汇编权等权利, 同时将上述权利再授权给第三方使用的权利。
<i>Data</i>	要求在文中注明使用许可类型, 但是未作限制。	未特别声明
<i>Biodiversity Data Journal</i>	要求使用 Open Data Commons Attribution License (ODC-By 1.0)、CC0、Open Data Commons Public Domain Dedication and Licence (PDDL	未特别声明

期刊名称	数据使用许可	数据版权
	1.0) 几种数据弃权许可协议。例外情况由编辑根据具体情况批准。	
全球变化数据学报	允许作者根据数据产生资助方合同要求选择将数据面向全社会共享或在限定范围内协议共享。	科学数据作者对以自己的署名出版的数据具有自主知识产权；作者将持有该数据的永久署名权。作者将数据著作权中的编辑权、不同介质复制权、印刷权、数据散发权、网络传播权、存储权、多语种翻译权和上述产权的转让权以及相关的其他权利与数据出版者分别共有，双方互不干涉数据出版后的数据管理和共享方式；数据集（库）出版权由数据出版者独有。
<i>GigaScience</i>	CC0	未特别声明
<i>F1000 Research</i>	要求使用 CC0，实际操作中允许使用 CC-BY 4.0 等许可协议	未特别声明
<i>Geoscience Data Journal</i>	未提及	未特别声明
<i>Journal of Open Archaeology Data</i>	开放许可协议（如 CCO）	未特别声明
<i>Open Data Journal for Agricultural Research</i>	官网无相关说明文件	官网无相关说明文件
地球大数据（ <i>Big Earth Data</i> ）	CC-BY、CC0 或同等的许可	未特别声明
<i>Journal of Chemical and Engineering Data</i>	开放许可协议，如 CC-BY	未特别声明

3.2 数据存储和标识

表 4 调研了数据期刊的存储要求和数据标识符相关规定。不同数据期刊对于数据存储在哪里有不同的要求，为促进开放数据的可发现和数据成果的复用，普遍支持将数据存储在服务安全稳定的公认学科特定主题存储库和公共的通用存储库中，并提供了存储库列表信息。数据存储库信息可以通过 re3data.org 查询，通用存储库一般包括 ZENODO、Figshare、Dryad、Dataverse 等。也有期刊只支持将数据存储在其指定存储库或其认可的存储库列表中，如有例外情况需经编辑部许可。《中国科学数据》要求作者优先将数据存储于科学数据银行（ScienceDB）。《全球变化数据学报》要求作者将数据提交到其自建存储库“全球变化科学研究数据出版系统”中。《GigaScience》规定：除例外申明外，数据都应存储在 GigaScience Database（GigaDB）；因为特定法律限制的数据不能开放共享的

数据，必须提交到持久但受限制的数据库，以便其他研究人员提交使用请求表格。

表 4 调研期刊的数据存储要求和数据标识符

Table 4 Data storage regulations and data identifiers of journals involved in this study

期刊名称	数据存储要求	数据标识符
<i>Data in Brief</i>	未指定存储库，但是要求存储库最低满足两个要求：（1）存储库为数据集分配一个全球唯一的持久标识符。（2）可以下载数据，也可以提供数据使用协议（Data Usage Agreement, DUA）。	全球唯一持久标识符
<i>Scientific Data</i>	期刊不托管数据，但要求作者将数据集提交到适当的公共数据存储库。数据应该尽可能提交到特定学科的、社区认可的存储库。如果不存在合适的特定学科资源库，则应该将数据提交到通用存储库。	稳定标识符（绝大多数情况是 DOI）
<i>IUCrData</i>	数据应在 IUCr 或适当的公共存储库中存档。存储库可以给数据分配全球唯一、永久标识符。	全球唯一持久标识符
<i>Earth System Science Data</i>	数据必须提交到长期存储库。存储库应满足以下基本标准：（1）持久标识符：数据集必须有 DOI。（2）开放获取：数据集必须免费提供，没有任何障碍。（3）自由版权：任何人都可以自由地复制、分发、传输和改编数据集，只要给原作者署名（相当于创作共用署名许可）。（4）长期可用性：存储库必须满足最高标准，以保证数据集的长期可用性和永久访问。	DOI
中国科学数据	数据须优先存储在科学数据银行（ScienceDB）及其他该刊认可的数据库。存储库应至少具备以下条件：（1）可以为数据分配 DOI 并注册；（2）支持数据的免登录、长期稳定访问；（3）可提供数据的私有访问链接供专家审核；（4）支持版本控制以便数据更新；数据发生更新，存储库支持和配合作者对数据进行修改、更新；（5）在数据存储页面标示数据引用信息，以及关联论文信息。	DOI、CSTR
<i>Data</i>	数据集需要在发布之前公开存储。小型数据集也可以作为论文补充材料发表在期刊网站上。	DOI 或其他唯一标识符
<i>Biodiversity Data Journal</i>	数据应保存并永久归档在适当的、可信的、通用的或特定领域的存储库（可在 http://service.re3data.org 查询）。如果不存在特定领域的数据存储库，作者应该将数据集存放在通用存储库中，如 ZENODO、Dryad、Dataverse 或其他存储库。小数据也可以作为研究文章的补充数据文件发布。	登录号、DOI 或其他持久 URL 标识符
全球变化数据学报	数据存储于“全球变化科学研究数据出版系统”中。	DOI、CSTR
<i>GigaScience</i>	除例外申明的情況外，数据都存储在 GigaDB。因为特定法律限制而不能开放共享的数据，必须提交到持久但受限制的数据库，以便其他研究人员提交使用请求表格。	登录号或唯一标识符，如 DOI。存储在 GigaDB 中的使用 DOI。
<i>F1000 Research</i>	强烈鼓励作者将数据存储到学科所属社区认可的存储库。如果不存在社区认可的存储库，将数据提交给通用数据存储库、机构存储库或国家存储库。提供了批准用于托管数据的存储库列表。如果存储库不在列表中，需经期刊评估。	DOI 或登录号

期刊名称	数据存储要求	数据标识符
<i>Geoscience Data Journal</i>	需要将数据存储到期刊认可的存储库，存储库至少具备：（1）为数据集分配 DOI 或其他 PID；（2）提供数据长期保存服务。（3）在不同的计算平台上以易于阅读的格式存储数据文件。（4）允许审查人员匿名访问数据及其元数据信息。（5）在数据论文发表后，将数据向公众提供开放访问。	DOI 或者其他 PID
<i>Journal of Open Archaeology Data</i>	鼓励作者将数据存储在其提供的符合期刊同行评审标准的通用和特定主题的存储库列表（ https://oad.simmons.edu/oadwiki/Data_repositories ）中。其存储库应符合以下标准：（1）适合存储所涉及的数据类型；（2）可持续性（即必须有资金和计划支持，以确保数据的长期保存）；（3）允许数据使用开放许可协议；（4）提供持久标识符（例如 DOI、Handle、ARC 等）。	永久标识符（如 DOI、Handle、ARC 等）
<i>Open Data Journal for Agricultural Research</i>	官网无相关说明文件	DOI
地球大数据 (<i>Big Earth Data</i>)	期刊不托管数据，要求作者将数据存入公认的存储库或数据中心。存储地必须能够生成 DOI，确保数据持久免费访问，并允许引用数据集。	DOI
<i>Journal of Chemical and Engineering Data</i>	该刊要求作者将数据存储在其 Figshare。隶属美国化学学会（American Chemical Society, ACS），ACS 旗下期刊的通用数据政策要求数据存储库符合以下标准：确保数据集的长期持久性和保存（至少在发布后 5 年），为存储的数据集提供稳定和持久的标识符，允许无限制地访问数据，在数据登录页面上提供明确的数据使用和数据访问（或许可）条款	DOI

注：“DOI”指数字对象标识符（Digital Object Identifier）；“CSTR”指中国科技资源标识符（Chinese Science and Technology Resource identifier）；“PID”指永久标识符（Permanent identifier）；“Handle”也是唯一持久标识符的一种；“ARC”指考古学资源标识符。

数据期刊对存储库的要求不尽完善，个别期刊未细化对数据存储的要求，如 *Data* 仅数据集需要在发布之前公开存储，小型数据集可以作为论文补充材料发表在期刊网站上。总体来看，数据期刊对存储库的要求和标准无外乎以下几点：（1）唯一标识符：存储库可以为数据集分配一个全球唯一的持久标识符。数据论文出版的一个重要目的是便于数据被更好地发现，使得数据像论文一样被引用，标识符对于定位和搜索数据集至关重要。数据期刊普遍支持使用 DOI 作为数据标识符，也有支持登录号、CSTR、PID、ARC 等其他永久标识符。（2）开放获取：免费提供数据的无障碍访问（包括无 IP、注册登录等限制）。（3）长期可用性：存储库可以提供长期、稳定访问。（4）使用许可协议：存储库可提供期刊要求的数据使用许可协议支持数据开放共享。（5）版本控制：数据发生更新时，存储库支持并配合作者对数据进行修改、更新。（6）数据与数据论文的关联及引用：在数据存储页面标示数据引用信息，以及关联数据论文信息。如《中国科学数据》提到该要求。

3.3 数据论文与数据的集成出版

表 5 中的单向访问指的是通过论文可以找到数据访问和下载地址，通过数据访问页面不能关联到对应的数据论文；双向互链访问是指不仅通过论文可以访问数据，通过数据访问页面也能找到对应的数据论文。数据期刊普遍因存储库的设置而异，通常能确保论文指向数据的单向访问，但是极少期刊能做到完全双向互链访问。ScienceDB、Mendeley Data、PANGAEA、Harvard Dataverse 等存

储库在数据访问页面中提供了数据关联论文的功能，可实现数据论文与数据的双向访问。而在 ZENODO、tDAR 等存储的数据未指向回数据论文。

表 5 调研期刊的出版过程与数据存储库的集成

Table 5 Integration of publishing process and data repository of journals involved in this study

期刊名称	互链访问类型	出版过程与数据存储库的集成要求
<i>Data in Brief</i>	单向或双向	论文 Specifications Table 的 Data accessibility 中提供数据访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Scientific Data</i>	单向或双向	论文 Data Records 章节中提供了数据访问网址或者是可以找到数据的方式。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>IUCrData</i>	单向	论文 Structural data 章节中提供数据访问链接。单个数据被注册 DOI，可以直接下载，数据无元数据信息等页面。单向访问。
<i>Earth System Science Data</i>	单向或双向	认为论文与数据相互连接的信息网络才能保证科学发现的完整性、透明度、可重用性和可重复性，数据应通过 DOI 链接到论文。在论文 Data availability 章节中提供数据访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
中国科学数据	双向	在论文数据集信息简介表中要求作者提供数据在线服务网址，其指定存储库 ScienceDB 可实现论文与数据的双向互链访问。
<i>Data</i>	单向或双向	论文摘要下有 Dataset 字段提供数据 DOI 号或访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Biodiversity Data Journal</i>	单向或双向	论文 Data resources 章节提供数据访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
全球变化数据学报	单向	通过论文的数据可用性声明可以找到数据访问网址。单向访问，全球变化科学研究数据出版系统中的数据未关联数据论文。
<i>GigaScience</i>	单向	论文 Data Availability 章节中提供了找到数据的路径。单向访问，GigaDB 存储的数据未关联数据论文。
<i>F1000 Research</i>	单向或双向	论文 Data Availability 章节中提供了数据访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Geoscience Data Journal</i>	单向或双向	论文 Dataset 部分给出访问链接。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Journal of Open Archaeology Data</i>	单向或双向	论文有 Repository location 可以链接到数据访问网址。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Open Data Journal for Agricultural Research</i>	单向或双向	论文浏览页面有突出的“Dataset”图标可直接访问数据集。因存储库而异，数据访问页面不一定能找到数据论文访问链接。

期刊名称	互链访问类型	出版过程与数据存储库的集成要求
地球大数据 (<i>Big Earth Data</i>)	单向或双向	文中 Data records 或者 Data Availability Statement 提供数据存储相关信息。因存储库而异，数据访问页面不一定能找到数据论文访问链接。
<i>Journal of Chemical and Engineering Data</i>	不详	非 OA 期刊，实际情况不详。

通常数据期刊在论文特定章节中会直接给出数据访问网址，因学科差异或给出数据访问和下载的方式，如 *Data in Brief*、《中国科学数据》论文中有专门的数据集信息简介表 (Specifications Table)，表中需要作者填写数据访问链接；*Scientific Data* 在论文 Data Records (数据记录) 章节中，*Earth System Science Data*、《全球变化数据学报》、*GigaScience*、*F1000 Research* 等在论文 Data Availability (数据可用性声明) 章节，*Biodiversity Data Journal* 在 Data resources 章节，*Journal of Open Archaeology Data* 在 Repository location 章节要求提供数据访问网址或者是可以找到数据的方式；*IUCrData* 论文 Structural data 章节中有数据访问链接，单个数据被注册 DOI，可以直接下载，数据无元数据信息等页面；或者论文访问页面中突出标识数据访问链接，如《中国科学数据》、《全球变化数据学报》、*Geoscience Data Journal*、*Data*、*Open Data Journal for Agricultural Research*。

3.4 数据评审

数据期刊出版中的评审，尤其是数据评审方面的特殊要求，包括要求作者在提交论文的同时提交数据、同行专家参照数据论文对数据进行评审，同行评议透明度，评审意见是否与数据论文同步发表等。

调研的数据期刊一般要求审查数据，并已在期刊官网或者相关文献中发布数据评审标准和要点 (表 6)，如 *Data in Brief*、*Scientific Data*、*Earth System Science Data*、《中国科学数据》、*Data*、*Biodiversity Data Journal*、《全球变化数据学报》、*GigaScience*、*F1000 Research* 等；*IUCrData*、*Geoscience Data Journal* 等期刊提到要求同行专家对数据进行审查，但未在官网发布相关标准；*Journal of Open Archaeology Data*、*Open Data Journal for Agricultural Research*、*Journal of Chemical and Engineering Data* 未提及是否要求对数据进行评审。《地球大数据 (Big Earth Data)》认为作者是数据质量的第一负责人，建议但不强制要求同行专家审查数据。关于数据期刊的评审要点，笔者已发表相关研究^[23-24]，本文不再赘述。

表 6 调研期刊的数据评审要求

Table 6 Data review regulations of journals involved in this study

期刊名称	执行强度	数据评审要求简述
<i>Data in Brief</i>	强	要求同行审查数据，并制定相关评审标准。如为 Elsevier 其他研究期刊发表的研究型论文关联的数据，无需另行审查。
<i>Scientific Data</i>	强	要求同行审查数据，并制定相关评审标准。
<i>IUCrData</i>	较强	要求同行审查数据，未公开评审标准。
<i>Earth System Science Data</i>	强	要求同行审查数据，并发布详细评审标准。
中国科学数据	强	要求同行审查数据，并发布详细评审标准。
<i>Data</i>	强	要求同行审查数据，并发布详细评审标准。
<i>Biodiversity Data Journal</i>	强	要求同行审查数据，并发布详细评审标准。

期刊名称	执行强度	数据评审要求简述
全球变化数据学报	强	要求同行审查数据，并发布相关评审标准。
<i>GigaScience</i>	强	要求同行审查数据，并发布相关评审标准。
<i>F1000 Research</i>	强	要求同行审查数据，并发布相关评审标准。
<i>Geoscience Data Journal</i>	较强	要求同行审查数据，未公开评审标准。
<i>Journal of Open Archaeology Data</i>	不详	官网无相关说明文件
<i>Open Data Journal for Agricultural Research</i>	不详	官网无相关说明文件
地球大数据 (Big Earth Data)	弱	建议审查，但不强制审查，提供审查标准 (未在网站公布)
<i>Journal of Chemical and Engineering Data</i>	不详	官网无相关说明文件

3.5 数据引用

倡导数据引用，鼓励甚至强制要求作者及使用者对数据进行规范引用，对数据期刊而言显得尤为迫切。数据引用包括数据论文中是否规范引用所有来源数据以及该文中描述的数据。强制要求在数据论文的参考文献中对该文关联的数据进行引用，执行强度较高的为 *Scientific Data*、*Earth System Science Data*、*Biodiversity Data Journal*、*GigaScience* 等期刊；*F1000 Research*、《地球大数据 (Big Earth Data)》等期刊虽然要求在数据论文参考文献中引用相关数据，但实际未引用文中描述的数据；*Data in Brief*、《中国科学数据》、*Data* 倡导规范引用数据，推荐了数据的引用格式，但是目前未强制要求作者在数据论文参考文献中引用论文中所描述的数据；*IUCrData* 未规范对数据的引用。其中，*Scientific Data*、《全球变化数据学报》不仅要求在数据论文参考文献中引用本文描述的数据，还要求作者在引用先前发表的数据时，要同时引用相关数据论文及数据本身 (表 7)。

表 7 调研期刊的数据引用要求

Table 7 Data citation regulations of journals involved in this study

期刊名称	执行强度	数据引用要求
<i>Data in Brief</i>	中	鼓励在文中和参考文献中引用数据，提供数据引用格式。
<i>Scientific Data</i>	极强	作者必须在数据论文中正式引用文中描述的数据；对于先前发表的数据集，要同时引用相关数据论文及数据本身；期刊工作人员将检查和落实数据的引用。
<i>IUCrData</i>	弱	未提及
<i>Earth System Science Data</i>	强	在数据论文的参考文献中引用文中描述的数据
中国科学数据	中	引导作者规范引用所有来源数据；文中提及并提供论文描述数据的引用格式。
<i>Data</i>	中	规范了数据引用格式
<i>Biodiversity Data Journal</i>	强	要求作者必须在论文中引用本数据

期刊名称	执行强度	数据引用要求
全球变化数据学报	强	在参考文献或其他适当的位置标明数据的标引；对于用户应用数据的同时参考数据论文的情况，要求用户既要标引数据集，又要标引关联的数据论文。
GigaScience	强	认可并实施 FORCE 11 数据引用原则 ^[25] ，并要求在参考文献中引用数据论文中所有相关公开可用的数据，包括引用数据论文中描述的数据。
F1000 Research	较强	要求在参考文献中正式引用文中提到的所有数据（含文中描述的数据），但实际执行中发现有未引用的情况。
Geoscience Data Journal	强	要求在参考文献中遵循 DataCite 推荐的方式引用数据
Journal of Open Archaeology Data	强	为了更好追踪数据引用情况，要求在数据论文的参考文献中增加对所描述数据的引用。
Open Data Journal for Agricultural Research	不详	官网无相关说明文件
地球大数据（Big Earth Data）	较强	推荐了数据引用格式，要求在文中参考文献中引用相关数据，但实际未引用文中描述的数据。
Journal of Chemical and Engineering Data	强	ACS 下属期刊支持 Joint Declaration of Data Citation Principles（数据引用原则联合声明） ^[25] ，要求在参考文献中引用所有数据，包括文中描述的数据。

3.6 数据安全

数据安全涉及的范围较广，从数据出版实践和研究目的角度考虑，本文将数据安全问题的调研重点放在数据期刊对发布数据所涉及内容的安全的审查。尽管数据安全是数据共享的关键问题，较少数据期刊把数据安全作为数据出版的重要议题作明确且全面的规范和说明。调研期刊对于数据安全的规范主要涉及人类隐私、实验伦理道德、社会安全、商业利益、合法合规等方面（表 8）。

表 8 调研期刊的数据安全相关规定

Table 8 Regulations about data security of journals involved in this study

期刊名称	简述	人类隐私	动物伦理	社会安全	商业利益	其他法律法规
Data in Brief	涉及使用人体实验对象，作者应遵守道德守则；对患者或志愿者的研究需要伦理委员会的批准和知情同意；使用来自社交媒体平台（如 Twitter、Facebook 等）的数据，应获得伦理批准、确保相关参与者的知情同意或适当匿名化、有权共享数据；所有动物实验遵循相关指南。	√	√	-	-	-
Scientific Data	所有动物实验按照相关指南和规定进行的。对于涉及人类受试者的实验，须相关委员会批准，确认所有受试者/患者照片的知情同意；极偶然	√	√	√	-	-

期刊名称	简述	人类隐私	动物伦理	社会安全	商业利益	其他法律法规
	情况下,对于危及社会安全数据,通过技术同行评议征求意见。					
<i>IUCrData</i>	未提及	-	-	-	-	-
<i>Earth System Science Data</i>	作者要保证不包含非法、滥用或违反合同、信任或对保密承诺的内容。	未明确说明	-	-	未明确说明	√
中国科学数据	符合国家法律法规,无政治错误,且无数据安全问题(包括但不限于涉及侵犯个人隐私,军事和科学技术等泄密、危害公众安全甚至国家安全,损害商业利益等)。	√	未明确说明	√	√	√
<i>Data</i>	数据不应存在道德、法律或隐私问题;作者应在提交数据时在数据可用性声明中明确说明限制;作者共享数据时应确保获得相关参与者对使用机密数据的许可。	√	未明确说明	-	未明确说明	√
<i>Biodiversity Data Journal</i>	使用个人数据需征得同意,用户对数据有更改权和删除权(即遗忘权)。	√	-	-	-	-
全球变化数据学报	符合国际科学伦理和科学规范、符合中国有关法规、保护个人隐私和遵循相关科学研究项目的相关政策规定。对社会安全、个人隐私、科学伦理及其他国家和国际科学组织认为不宜公开出版的数据作退稿处理。	√	√	√	-	√
<i>GigaScience</i>	涉及人类应符合牛津大学出版伦理声明,相关实验必须得到伦理委员会的批准,获得参与者的知情同意。	√	-	-	-	-
<i>F1000 Research</i>	所有研究必须在适当的伦理框架内进行。对于涉及人类或动物的研究,必须在“方法”部分提供作者所在机构或伦理委员会批准的详细信息,并说明“从参与者/患者/父母/监护人/亲属处获得关于公布参与者/患者详细信息和/或其图像的书面知情同意”。	√	√	-	-	-
<i>Geoscience Data Journal</i>	未提及	-	-	-	-	-

期刊名称	简述	人类隐私	动物伦理	社会安全	商业利益	其他法律法规
<i>Journal of Open Archaeology Data</i>	由于法律原因需要部分编校的数据集将由编辑团队根据具体情况决议。	未明确说明	未明确说明	未明确说明	未明确说明	√
<i>Open Data Journal for Agricultural Research</i>	官网无相关说明文件	-	-	-	-	
地球大数据 (Big Earth Data)	只有符合道德的数据方可共享，且不违反对人类受试者的保护，或其他有效的道德、隐私或安全问题。	√	未明确说明	√	-	-
<i>Journal of Chemical and Engineering Data</i>	在道德和法律上可行的情况下，所有 ACS 期刊都强烈鼓励作者在发表时公开其文章的研究数据。	未明确说明	未明确说明	-	-	√

注：“-”指未找到相关说明；“未明确说明”指提到相关内容，但是不明确。

相较而言，国内期刊对数据安全问题的重视程度更高，相关规定更为全面。例如，《中国科学数据》规定，所有数据应符合国家法律法规，无政治错误，且无数据安全问题（包括但不限于涉及侵犯个人隐私，军事和科学技术等泄密、危害公众安全甚至国家安全，损害商业利益等）；《全球变化数据学报》规定数据应符合国际科学伦理和科学规范、符合中国有关法规、保护个人隐私和遵循相关科学研究项目的相关政策规定，对社会安全、个人隐私、科学伦理及其他国家和国际科学组织认为不宜公开出版的数据做退稿处理。

而国外期刊相较而言更为重视人类隐私和伦理道德等问题，在相关方面做了较为细致的规定，如 *Data in Brief*、*Scientific Data*、*Biodiversity Data Journal*、*GigaScience*、*F1000 Research*。对于涉及人体为实验对象的数据，必须签署患者或受试者知情同意的书面许可协议方可发表。*Data in Brief* 参照 Elsevier 的相关管理办法规定，对于涉及使用人体实验对象，作者应遵守 *The Code of Ethics of the World Medical Association (Declaration of Helsinki)*《世界医学协会关于涉及人体实验的道德守则》（赫尔辛基宣言）。在伦理声明中，作者应该在手稿中包括一项声明，说明在人体实验中获得了知情同意，以及伦理委员会，协议编号。书面同意必须由作者保留，只有在特殊情况下（例如出现法律问题），作者才必须提供书面同意的副本或已获得同意的证据。*Data in Brief* 还规定所有动物实验都应遵守 ARRIVE (Animal Research: Reporting of In Vivo Experiments) 指南 (<https://arriveguidelines.org/>)，并应根据 1986 年英国《动物（科学程序）法》和相关指南、欧盟动物实验指令 2010/63/EU 或美国国立卫生研究院实验动物护理和使用指南 (NIH 出版物第 8023 号，1978 年修订) 进行，作者应在稿件中明确指出已遵循这些指南。此外，*Earth System Science Data*、*Data*、*Journal of Open Archaeology Data*、*Journal of Chemical and Engineering Data* 等期刊提到在遵守相关法律法规的情况下共享数据。*Scientific Data* 还提到发布的数据偶尔会涉及对社会的影响，包括威胁社会安全。在这种情况下，通常会在技术同行评议过程中征求意见，评估和平衡出版面临的风险与收益，但是预计只有极少数情

况下，风险会被认为超过了发布数据的收益。

3.7 数据更新

一些负责任的作者可能根据使用者的反馈或者因为数据处理方法的改善、周期性扩展而需要更新数据，更新的数据在完整性、数据质量和可用性方面有明显改进，而更新的部分又难以支撑起新的数据论文，这时可能会涉及数据及数据论文更新的问题。少数调研期刊已在数据更新方面制定了相关对策，通常理解并支持数据的更新，只有 *Geoscience Data Journal* 表示不支持数据更新，认为数据一经发表不得更改（表 9）。

表 9 调研期刊的数据更新规定
Table 9 Data update regulations of journals involved in this study

期刊名称	简述	数据 DOI 是否发生变化	关联数据论文是否更新	新旧数据版本之间是否建立关联
<i>Data in Brief</i>	2022 年 10 月新增“Update article（更新文章）”文章类型来支持数据更新。	未提及	是	是
<i>Scientific Data</i>	通过动态版本的并行托管，实现数据更新，并鼓励作者在数据采集方法发生重大变化时发表新数据论文。	未提及	未提及	是
<i>IUCrData</i>	未提及	-	-	-
<i>Earth System Science Data</i>	通过“Living Data（动态数据）”（ESSD LD）流程来支持定期更新或扩展的数据集，以此在保持高标准的数据质量评估的同时，为作者、审稿人和编辑提供便利。	是	是	是
中国科学数据	支持以短论文和论文迭代两种方式来配套更新数据。	不一定。以短论文形式更新产生新 DOI；以论文迭代形式更新时 DOI 保持不变。	是	是
<i>Data</i>	未提及	-	-	-
<i>Biodiversity Data Journal</i>	未提及	-	-	-

期刊名称	简述	数据 DOI 是否发生变化	关联数据论文是否更新	新旧数据版本之间是否建立关联
全球变化数据学报	数据出版后，支持因误或时间序列产生的新数据按照更新版本连续出版。	不一定。仅在作者发生变化时，DOI 相应更新。	未提及	未提及
<i>GigaScience</i>	未提及	-	-	-
<i>F1000 Research</i>	未提及	-	-	-
<i>Geoscience Data Journal</i>	不支持数据更新；提交发表的数据集必须完整且不可更改；包括元数据也可更改，以便对数据进行充分的理解和跟踪。	否	否	否
<i>Journal of Open Archaeology Data</i>	未提及	-	-	-
<i>Open Data Journal for Agricultural Research</i>	官网无相关说明文件	-	-	-
地球大数据 (<i>Big Earth Data</i>)	未提及	-	-	-
<i>Journal of Chemical and Engineering Data</i>	未提及	-	-	-

注：“-”指未找到相关说明。

关于数据更新，数据期刊重点关注的议题包括以下几个方面：（1）新数据的 DOI 是否沿用原数据 DOI。在不产生新的数据论文的情况下，利用存储库的迭代功能新增的数据版本 DOI 通常与原数据保持一致。但 *Earth System Science Data* 要求所有更新的数据集都采用与先前的数据集不同的唯一且独立的永久标识符。《中国科学数据》的做法是以短数据论文形式更新的数据，数据也采用新的 DOI；而以论文迭代形式更新时，数据 DOI 保持不变。《全球变化数据学报》按照 3 种不同情况进行处理：发现错误以后的数据更新，数据 DOI 不变，仅在数据集注释中增加更新日期字样；时间序列数据更

新、不涉及作者变化时,原来的数据文件不动,增加新的数据文件,数据 DOI 不变,仅在注释中增加更新日期字样;时间序列数据更新、作者发生了变化时,采取新数据集出版的方式,DOI 相应变化。(2) 关联数据论文是否更新。*Scientific Data* 认为研究达到新的里程碑或数据采集方法发生重大变化,才需要撰写新的数据论文。而《中国科学数据》要求:只要已发表数据发生更新,就应修改关联数据论文的相关描述,以配套新数据的使用;如果新旧数据采集方法相似,但数据类别不同,宜采用短数据论文的形式对新旧数据进行异同比较和描述,并在文中引用和关联先前发表的数据论文,实践案例可参见 <http://www.csdata.org/p/144/>;如果数据采集和质控等不足以支撑发表新数据论文时,宜采用论文迭代的方式进行,提供版本记录和追溯,并在论文相关评审记录中对更新情况进行说明,实践案例可参见 http://www.csdata.org/p/news_detail/237/。(3) 新旧数据版本之间是否建立关联。开发良好的具有版本控制或更新系统的存储库,如 ScienceDB 可以实现动态版本的并行托管,允许用户追溯数据的不同版本,并查找最新的数据。*Scientific Data*、《中国科学数据》均对存储库提出了相关版本控制方面的要求。ESSD 使用 ESSD LD 流程将在数据不同版本之间建立明确联系,并链接到先前相应的描述中。

此外,更新数据及其关联论文的评审一般不要求重新评估原始数据或原始处理方案,只需要关注数据中更新的部分。*Earth System Science Data* 要求在数据论文中重点描述当前版本中实现的更改,并在可能的情况下记录以前发布的数据的使用和影响。在理想的 ESSD LD 中,寻求同行专家和编辑的最大连续性,但在实践中,可能受到相关人员缺位的影响,更新的频率和效率也会受到用户社区的规模和兴趣的影响。

4 问题与讨论

从上述研究不难看出,期刊普遍在数据使用许可、数据存储和永久标识符、论文与数据的关联访问、数据评审、规范数据引用方面规定较为完善,数据更新实践开展的案例尚少。此外,绝大多数数据期刊存在对数据安全和数据版权重视不足的问题,相关规定较为欠缺,有必要对两者进行重点阐述。

4.1 数据安全的范畴与要点

保证数据安全是数据开放共享的基础和前提^[26],科研人员、出版机构在数据出版过程中应树立数据安全意识。个人信息、国家安全、商业利益是数据安全保障的重中之重,数据出版应避免出现涉及侵犯个人隐私、危害国家安全、损害商业利益的问题。在全球范围内,世界各国近年来本着“科学数据共享交流、技术数据可控流动、重要数据立法保护”的原则出台了相应的法规进行数据治理。

(1) 个人隐私

个人信息的核心定义是已识别或者可识别的自然人信息。在不泄漏个人信息的前提下提高数据的利用率、挖掘数据的价值,是目前数据开放共享的关键问题之一。《中华人民共和国个人信息保护法》^[27]完善了个人信息保护规则,建立了用户信息保护制度,明确了对个人信息采集、使用及保护的要求,并规定了个人对其信息进行更正或删除的权利。欧盟《通用数据保护条例(General Data Protection Regulation, GDPR)》^[28]规定除非数据主体明确同意,否则禁止匿名处理特殊敏感数据,如健康、种族、性取向等个人数据。关于个人信息保护的原则和框架,GDPR 规定了 7 项原则:(1) 合理、合法、透明原则。(2) 目的限制原则。收集个人信息应明确、具体且正当,满足法律规定的涉及公共利益、科学或者历史研究或统计目的的数据除外。(3) 数据最小化原则。仅收集和适当的、相关的、必要的个人信息。(4) 准确性原则。个人数据如不准确,必要时应及时擦除或更正。

(5) 限制存储原则。存储时间不能超过实际处理所需的必要时间。(6) 数据的完整性与保密性原则。

(7) 可问责性原则。

(2) 国家安全

国家安全包括政治安全、国土安全、军事安全、经济安全、文化安全、社会安全、科技安全、网络安全、生态安全、资源安全、核安全。数据是基础性战略资源、新型生产要素，是科技创新、经济发展、生命健康和社会进步的重要资源，是国际博弈和对抗的重要元素，与国家安全密切相关。违反国家安全的数据泄漏，可能损害国家在政治、经济、国防、外交等领域的安全和利益。国家重要军事设施、关键基础设施、重要交通要道等的地理空间高精度数据、核科学数据，与大规模杀伤性武器、群体杀伤技术、生物恐怖主义相关的反恐科技数据^[29]，敏感的化学、医学数据等，海洋资源、环境数据、特殊材料和设备数据等，与国防安全密切相关。一些环境、灾害真实数据可能会带来不必要的社会影响^[26]。对于数据出版，应首先遵守国家相关法律条款，在作者与期刊出版机构签署协定时，应明确不涉及国家秘密、国家核心数据与国家重要数据。

(3) 商业利益

商业机密是指不为公众所知悉、具有商业价值并经权利人采取相应保密措施的技术信息、经营信息等商业信息。大部分非公开的数据集合都可以满足商业秘密的要件，如电子商务平台的用户操作和交易数据、城市交通数据等。

(4) 其他

A) 规模致敏。单个数据敏感性可能微不足道，但是当数据累积到一定量级后，通过单个数据获得的细节与实质性敏感内容显现出价值，敏感性才暴露出来。在大数据背景下，数据通过主动汇交或被动汇集，形成具有新的数据集合，数据形态和结构发生变化，从而具备单一的数据个体本身并不具备的价值和功能，即“规模致敏”。这些零散数据如果被有意或恶意收集和整理，可能给社会公共利益带来隐患，甚至危及国家安全^[26]。

B) 安全风险的不确定性。随着数据挖掘、规则模型构建、人工智能、知识关联与发现等关键技术^[30]的发展，数据安全风险的辨识、评估、推演难度加大。对这些可能在特定条件发生敏感性变化的数据，有必要明确核心数据范围和弹性空间，定期评估和预测未来技术发展带来的数据敏感性和风险扩张的可能性^[26]。

C) 数据安全的技术与服务支持。在软硬件支持和数据管理层面，应建立有效的数据灾备机制，严格记录数据的存取日志，采用稳定性高的软硬件产品，确保数据安全、长期稳定访问和下载。既不影响数据开放，也要保障数据安全。

4.2 数据版权保护

知识产权保护既是共享数据的动力又是障碍，版权保护的目標和难点在于平衡数据共享传播与利益相关者权益之间的矛盾。数据版权不仅包括数据的所有权、使用权，还包括转让权、收益权等。缺乏对研究数据的版权和使用许可证的理解是研究人员不共享数据的一个常见原因^[31]。但是限制数据，尤其是公共资金资助产生的数据的收集和复用，无疑给使用者增加了障碍和侵权风险，因此《关于科学数据的北京宣言》倡议科学数据免受版权或其他知识产权立法保护^[18]。

仅仅作事实(facts)的数据往往难以受到版权法的保护，而具备创新性(creativity)和原创性(originality)的数据可以被视为创造力的智力成果被保护，数据也可能通过版权法以外的法律获得保护。相关知识产权国际协定表明，版权保护的對象包括文学、科学和艺术作品，无论表达方式或形式如何(《伯尔尼公约(Berne Convention)》第2条)，版权保护只适用于表达，而不适用于思想、程序、操作方法或数学概念(《与贸易有关的知识产权协定》(Agreement on Trade-Related Aspects of Intellectual Property Rights)第9(2)条和《WIPO 版权条约》(WIPO Copyright Treaty)第2条)^[32]。中国新版《著作权法》(2021年发布)也未明确将数据列为受保护的智力成果，数据资源确权、开放、流通、交易等相关数据版权保护制度还在完善中。据美国法律，数据也被视为“facts”，没有版权，原因是数据是被发现的，而不是原创作品^[33]。欧盟《数据库指令》(Database Directive)(1996年)

也指出,版权保护的对象不能扩展到“仅仅是事实或数据”,确定数据库是否应受版权保护的标准,应界定为选择或安排数据库内容是作者自己的智力创造,鉴于这种保护应包括数据库的结构^[34]。欧盟《特殊权利保护》(Sui Generis Right)规定禁止提取或重新利用任何在获取、验证或呈现数据内容方面已投入大量资金的数据库(不需要创造性或原创性),这一权利自非创造性数据库创建之日起持续15年。但是,数据作者可以在数据汇编过程中拥有版权,数据的创造性安排、注释或选择可以受到版权法的保护。此外,可以通过其他途径,如专利法、竞争法、保密法等实现对数据及相关工作的法律保护。如果数据收集带来了新的实用的发明,如机器、工艺、制造或改进,专利法可能适用。如果配方、工艺、设计或方法具有商业优势,数据可能会受到商业机密的保护。一些合同或授权附带保密协议或其他要求保密的条件,以及如涉及国家安全、个人隐私等必须保密的数据也都会受到保护。

出版界主流观点认为,数据版权的主体是数据作者。国际科学技术和医学出版商协会(STM协会)和国际学术出版者学会(Association of Learned and Professional Society Publishers, ALPSP)在2006年发表了一份联合声明,宣布他们不会获得研究数据的版权^[35]。数据版权应该保留给作者,而不应该转移给出版机构已经成为国际共识。

但是,在数字时代,数据出版的载体是互联网。这就导致在互联网中侵权存在零成本复制、隐蔽性强、传播迅速、全球性、复杂性、裂变性、罪证难以收集且易于销毁、难监测等特点,而维权却需要付出很高时间和经济成本,甚至可能维权失败。此外,在数据出版的实践中,关于侵权与被侵权的争议日渐凸显,迫切需要提出合理的解决方案,完善相应法律法规。例如:通过爬虫等大数据与人工智能技术抓取的数据版权以及是否侵权如何界定?将仍在版权保护期内的临时编委会编写的经典纸质书籍扫描录入并形成数据库的信息化行为对于珍贵资料的保存、继承和传播极为重要,是否可视作合理使用,又可按照什么标准来确定是否危害原作者的合法利益?

5 小结与展望

学术期刊的发展已历经超过350年,而数据期刊的萌芽和发展不到30年,数据出版已成为期刊界的重点发展方向和新兴领域之一。数据期刊的出版不是单纯的数据论文的出版,而是与数据的关联出版,数据的重要性和对质量的要求远高于数据论文本身,因而涉及的因素远比传统期刊要复杂得多、出版的难度也要大得多,种种因素共同造就了数据期刊的兴与衰。在开展数据出版的过程中,期刊需要不断学习借鉴先行做法和成败经验、反思总结自身实践经验,不断梳理和完善出版规程,形成一套适合本学科和自身特点的数据出版经验,实现可持续的数据出版。

基于调研分析,提出以下建议:(1)只有充分认可、保障数据生产者、数据加工者的工作与权益,才会有更多的人员愿意从事数据生产、加工和出版等工作。因此,期刊数据版权制度的确立,对于数据期刊的长期发展至关重要,应引起充分重视。(2)在合作数据存储平台的选择上,数据期刊应该具有国际视野,按照国际规则积极与国际先进数据平台开展深度合作,积极融入国际科技数据资源体系,以便数据被更好地发现、保存和维护。(3)数据期刊目前基本实现了从论文到数据的链接指引,而忽略了从数据到论文的反向链接,从而使得数据论文的能效大打折扣,有必要注意规范并实现数据与论文的双向互链。(4)数据出版作为独特的新事物,学术界需要引入新的评价方式来对数据的引用进行科学的评价,促进科学数据的充分发布、共享与再利用。目前每年数据论文出版量还非常小,远不及普通学术论文的出版量,对于数据论文的引用习惯、生态还未完全建立。建

议期刊倡导数据引用文化,对于必要的引用实行数据和数据论文双引用的强制措施。(5) 数据安全问题在数据出版范围和出版量不断扩大的过程中逐渐凸显,安全的数据方可共享,数据期刊应强化数据安全意识。(6) 数据更新也是数据出版实践活动中面临的问题,提前考虑此类问题的解决方案有助于实现数据出版的连续性。数据出版的目的是更好地实现数据的共享与复用,对同种数据采集方法下更新的数据集延续使用原有 DOI,并相应更新关联数据论文,有利于用户对数据的获取和使用,以及提升数据完整性和质量,不断累积数据影响力。(7) 充分利用集群化的、领先的出版集团在数据政策制定和实践流程方面的发展优势,有助于克服单刊发展势弱问题,助力数据期刊的发展。

致 谢

感谢国家社会科学基金项目“开放科学视域下科技期刊论文数据汇交模式研究”(22BTQ107),中国科学院学部咨询评议项目“开放科学的趋势与影响研究”专题二“开放数据与我国开放数据基础设施研究”、专题三“全球开放科学治理发展态势及我国开放创新生态构建路径研究”(2022-X01-B-008)资助。

作者分工职责

习妍(1984—),中级编辑,研究方向为数据出版、网络出版、新媒体出版。主要承担的工作:论文架构,数据统计,资料调研,论文撰写。

孔丽华(1978—),高级工程师,研究方向为科学数据管理政策、数据开放共享政策与实践,以及数据出版。主要承担的工作:思路指导和调研,论文部分内容的撰写和修改。

汪洋(1984—),高级工程师,研究方向为大数据分析、态势感知系统、信息化发展战略研究。主要承担的工作:论文指导,论文部分内容的撰写和修改。

陈枢舒(1986—),中级工程师,研究方向为文献信息检索与情报分析。主要承担的工作:论文校对和修改。

参考文献

- [1] 郭华东. 问渠哪得清如许,为有源头活水来——《中国科学数据》发刊词[J/OL].中国科学数据, 2016, 1(1). DOI:10.11922/csdata.0.2016.0014.[Guo H D. “Pool mirroring the sky and clouds. From the source run fresh currents”—Introducing China Scientific Data[J/OL]. China Scientific Data, 2016, 1(1). DOI: 10.11922/csdata.0.2016.0014.]
- [2] 张晓林, 沈志宏, 刘峰. 科学数据与文献的互操作[M]// CODATA 中国全国委员会编著.大数据时代的科研活动. 北京: 科学出版社, 2014: 149-158. [ZHANG X L, SHEN Z H, LIU F. Interoperability of Scientific Data and Literature [M]// CODATA, China National Committee. Research activities in the era of big data. Beijing: Science Press, 2014: 149-158.]
- [3] CANDELA L, CASTELLI D, MANGHI P, et al. Data journals: A survey[J]. Journal of the Association for Information Science and Technology, 2015, 66(9): 1747-1762. DOI:10.1002/asi.23358.
- [4] 刘凤红, 彭琳. 国际数据期刊的发展现状调查与分析[J]. 中国科技期刊研究, 2019, 30(11): 1129-1134. DOI: 10.11946/cjstp.201904280323. [LIU F H, PENG L. An investigation and analysis of development status of international data journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2019, 30(11): 1129-1134. DOI: 10.11946/cjstp.201904280323.]

- [5] WALTERS W H. Data journals: Incentivizing data access and documentation within the scholarly communication system[J]. *Insights*, 2020, 33(1): 1-20. DOI: 10.1629/uksg.510.
- [6] 王卫军, 李成赞, 郑晓欢, 等. 全球科学数据出版发展态势分析——基于 Web of Science 数据库的调研[J/OL]. *中国科学数据*, 2021, 6(3). (2021-09-29). DOI: 10.11922/csdata.2021.0019.zh. [Wang W J, Li C Z, Zheng X H, et al. The analysis of the development trend of global scientific data publishing: research based on Web of Science database[J/OL]. *China Scientific Data*, 2021, 6(3). (2021-09-29). DOI: 10.11922/csdata.2021.0019.zh.]
- [7] The FAIR Data Principles[EB/OL].<https://force11.org/info/the-fair-data-principles/>.
- [8] NOSEK B A, ALTER G, BANKS G, et al. Transparency and openness promotion (TOP) guidelines [EB/OL]. 2016. <https://doi.org/10.1126/science.aab2374>.
- [9] CARROLL S R, GARBA I, FIGUEROA-RODRÍGUEZ O L, et al. The CARE principles for indigenous data governance[J]. 2020,19(1):43. DOI: 10.5334/dsj-2020-043.
- [10] CARE Principles for Indigenous Data Governance[EB/OL]. <https://static1.squarespace.com/static/5d3799de845604000199cd24/t/5f0621e031811d08f9c4cd27/1594237427481/Principios+CREA.pdf>.
- [11] 雷雪. FAIR 原则在科技期刊数据出版中的应用实践与推进建议[J]. *中国科技期刊研究*, 2022,33(08):1075–1080.DOI: 10.11946/cjstp.202201100015. [LEI X. Application of FAIR Principles in data publishing of scientific journals and suggestions[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2022, 33(8): 1075–1080. DOI: 10.11946/cjstp.202201100015.]
- [12] 关琳琳, 金鑫, 马瀚青, 等. 国际数据期刊出版质量控制机制研究[J]. *中国科技期刊研究*, 2021,32(10):1269–1277. DOI: 10.11946/cjstp.202103260267. [GUAN L L, JIN X, MA H Q, et al. Quality control mechanism of international data journals: a policy-based study[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2021, 32(10): 1269–1277. DOI: 10.11946/cjstp.202103260267.]
- [13] 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. DOI: 10.3969/j.issn.1002-1965.2022.03.025. [FANG H, SONG L R. Research on scientific data publishing policies of domestic and foreign journals[J]. *Journal of Intelligence*, 2022, 41(3): 181–190. DOI: 10.3969/j.issn.1002-1965.2022.03.025.]
- [14] 李新月, 青秀玲, 韩露, 等. 生命科学领域数据出版实践研究[J]. *中国科技期刊研究*, 2021,32(7): 935–943. DOI: 10.11946/cjstp.202012221041. [LI X Y, QING X L, HAN L, et al. Data publishing practice in the field of life science[J]. *Chinese Journal of Scientific and Technical Periodicals*, 2021, 32(7): 935–943. DOI: 10.11946/cjstp.202012221041.]
- [15] 章紫桐. 数据期刊出版机构运营策略研究 [D]. 南京大学, 2021. DOI:10.27235/d.cnki.gnjiu.2021.001419. [ZHANG Z T. Data journal publishing organization operating strategy research [D]. Nanjing university, 2021. DOI:10.27235/d.cnki.gnjiu.2021.001419.]
- [16] 李红星, 吴立宗, 南卓铜, 等. 科学数据联合出版模式与内容研究[J]. *遥感技术与应用*, 2016, 31(4): 801–808. DOI: 10.11873/j.issn.1004-0323.2016.4.0801. [LI H X, WU L Z, NAN Z T, et al. Collaborative publishing of scientific data: model and implementation[J]. *Remote Sensing Technology and Application*, 2016, 31(4): 801–808. DOI: 10.11873/j.issn.1004-0323.2016.4.0801.]
- [17] Principles for Open Data in Science[EB/OL].<http://pantonprinciples.org/>.
- [18] CODATA, Committee on Data of the International Science Council; CODATA International Data Policy Committee; CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice; et al. The Beijing Declaration on Research Data[DS/OL]. Zenodo, 2019. [2023-01-17]. <https://doi.org/10.5281/zenodo.3552330>.
- [19] PENEV L, MIETCHEN D, CHAVAN V, et al. Pensoft data publishing policies and guidelines for biodiversity data[J]. *Pensoft Publ.*, 2011(1).
- [20] Mendeley Data for journals[EB/OL].<https://www.elsevier.com/authors/tools-and-resources/research-data/mendeley-data-for-journals>.
- [21] Science Data Bank [EB/OL]. <https://www.scidb.cn/faq#item7>.

- [22] 《全球变化数据学报》(中英文)编辑部. 全球变化科学研究数据出版与共享指南[J]. 全球变化数据学报, 2017, 1(3): 253-261. DOI:10.3974/geodp.2017.03.01. [EDITORIAL OFFICE OF JOURNAL OF GLOBAL CHANGE DATA & DISCOVERY. Guidelines of global change research data publishing & repository[J]. Journal of Global Change Data & Discovery, 2017, 1(3): 253-261. DOI:10.3974/geodp.2017.03.01.]
- [23] 孔丽华. 数据出版中的数据质量评议机制及其实践研究[D].北京: 中国科学院大学.2019. [KONG L H. Research on data quality Review Mechanism and Practice in data publishing [D]. Beijing: University of Chinese Academy of Sciences.2019.]
- [24] 孔丽华, 习妍, 郎杨琴, 等. 数据期刊中科学数据的同行评议方法研究[J]. 编辑学报, 2019, 31(3): 262-266. DOI: 10.16811/j.cnki.1001-4314.2019.03.007. [KONG L H, XI Y, LANG Y Q, et al. Study of peer review criteria for scientific data in data journals[J]. Acta Editologica, 2019, 31(3): 262-266. DOI: 10.16811/j.cnki.1001-4314.2019.03.007.]
- [25] Data Citation Synthesis Group. Joint Declaration of Data Citation Principles[EB/OL]. Martone M. (ed.) San Diego CA: FORCE11, 2014. [2023-01-17]. <https://doi.org/10.25490/a97f-egykh>.
- [26] 李宜展, 刘细文, 李泽霞, 等. 科学数据安全边界概念模型研究: 基于利益相关者视角[J]. 中国科学基金, 2022, 36(2): 339-347. DOI: 10.16262/j.cnki.1000-8217.2022.02.047. [LI Y Z, LIU X W, LI Z X, et al. Study on conceptual analysis model of scientific data security boundary: from the perspective of stakeholders[J]. Bulletin of National Natural Science Foundation of China, 2022, 36(2): 339-347. DOI: 10.16262/j.cnki.1000-8217.2022.02.047.]
- [27] 中华人民共和国个人信息保护法[S]. (2021-8-20). <http://www.npc.gov.cn/npc/c30834/202108/a8c4e3672c74491a80b53a172bb753fe.shtml>. [Personal information protection law of the People's Republic of China [S]. (2021-8-20). <http://www.npc.gov.cn/npc/c30834/202108/a8c4e3672c74491a80b53a172bb753fe.shtml>.]
- [28] Regulation P. General data protection regulation[J/OL]. Intouch, 2018, 25.<https://gdpr-info.eu/>.
- [29] 钟灿涛. 开放与保密: 科技信息传播控制及其对创新的影响: 以美国科技信息传播控制机制为例[J]. 科学学研究, 2013, 31(3): 335-343. DOI: 10.16192/j.cnki.1003-2053.2013.03.003. [ZHONG C T. Openness and closeness: dissemination control of scientific and technical information and its impact to innovation with dissemination control mechanism of the United States as example[J]. Studies in Science of Science, 2013, 31(3): 335-343. DOI: 10.16192/j.cnki.1003-2053.2013.03.003.]
- [30] 周毅, 刘峥, 张建勇. 关联数据研究的主题结构和研究进展解析[J]. 农业图书情报, 2019, 31(3): 13-24. DOI: 10.13998/j.cnki.issn1002-1248.2019.03.19-0367. [ZHOU Y, LIU Z, ZHANG J Y. Research on subject structure and research progress of linked data[J]. Agricultural Library and Information, 2019, 31(3): 13-24. DOI: 10.13998/j.cnki.issn1002-1248.2019.03.19-0367.]
- [31] KESSLER R. Whitepaper: Practical challenges for researchers in data sharing[J]. Learned Publishing, 2018, 31(4).
- [32] MLECZAK P, MARCINOSKA-BOULANGÉ L. Data and copyright[EB/OL]. (2020-12-10). <https://litigationportal.com/en/p/data-and-copyright>.
- [33] Data Management: Intellectual Property And Copyright [EB/OL]. <https://libguides.library.kent.edu/data-management/copyright#s-lg-box-14471309>.
- [34] Database Directive, Directive 96/9/Ec Of The European Parliament And Of The Council [EB/OL]. <http://www.columbia.edu/~mr2651/e-commerce3/2nd/statutes/DatabaseDirective.pdf>.
- [35] STMA. Databases, data sets, and data accessibility—views and practices of scholarly publishers[EB/OL]. 2006. https://www.stm-assoc.org/2006_06_01_STM_ALPSP_Data_Statement.pdf.

论文引用格式

习妍, 孔丽华, 汪洋, 等. 国内外数据期刊实践与进展研究[J/OL]. 中国科学数据, 2023, 8(1). (2023-03-29). DOI: 10.11922/11-6035.ncdc.2023.0001.zh.

Research on the practice and exploration progress of data journals

XI Yan^{1*}, KONG Lihua¹, WANG Yang¹, CHEN Shushu¹

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, P.R. China

*Email: xiyan@cnic.cn

Abstract: With the booming trend of data publishing, data journals have carried out much extensive practice for the exploration of open data sharing. This study focuses on the publishing practice of data journals, which took representative, relatively mature and different disciplinary scopes of data journals as the research objects. The publishing volume and proportion of data papers in data journals are calculated. Based on the author's practical experience in data publishing for several years, this study explores the formulation rules and implementation of data publishing policies. The study results show that data journals generally have well-developed regulations on data use license, data storage and permanent identifier, associated access from paper to data, data review, standardized data citation, but with few practical cases of data update and less attention paid to the problem of data security and data copyright. To facilitate the further development of data journals, we proposed in this study that more operational and reference details of data publishing should be considered in the practice of data publishing in journals.

Keywords: data journal; data publishing; data copyright; data storage; data reference; data security; data update