·“双清论坛”专题：开放科学的实践与政策·

数据出版的趋势、机制与挑战

孔丽华[1,2,3*]　　习　妍[3]　　张晓林[1]

（1. 中国科学院文献情报中心，北京 100190；
2. 中国科学院大学经济与管理学院图书情报与档案管理系，北京 100049；
3. 中国科学院计算机网络信息中心，北京 100190）

[摘　要] 数据出版是激励数据传播、促进数据共享的重要方式之一。本文通过对国内外科研数据存储库、国内外主要科技出版商、代表性学术研究期刊和专门数据期刊的调研，对当前作为数据文档自存储发布、作为学术论文辅助数据文档发布、作为专门数据论文发表等三种机制进行了分析，提炼梳理了它们的数据出版政策要素，分析了我国数据出版面临的挑战并提出针对性建议。

[关键词] 开放数据；数据出版；论文辅助文档；数据期刊；数据存储库；政策

信息技术与科学研究的交互融合引发了数据量的迅猛增长，促成了数据密集型科学发现的科研第四范式[1]的到来，数据成为科研和创新的基础驱动力。同时，科研结果的开放共享、尤其是科研数据的开放共享，是保证科研结果的可验证、可分享、可重现的基础支撑[2]，也是科研促进技术、行业和社会创新发展的重要手段[3]。

历史上，数据以多种形态出现，例如手写记录、机写纸质记录（例如传统的仪器打印记录）、胶片照片、模拟信号影像（例如医疗 X 照片或田野调查照片或口述历史录音等）等。在纸本和模拟信号时代，这些数据的展示、保存和传递存在巨大困难。即使是大型科研设施产生的数字信息型数据，由于体量巨大，也难以便捷地传送和共享。因此，除了少量数据被整理后发表，大多数数据都深藏在不断被遗忘的办公室角落里或置于不断转移任务与责任的个人手里。即使是那些随论文整理发表的数据，受传统论文篇幅限制，很难发布相对完整的原始数据，也难以链接未整理或未发布的其他数据以及相关的方法、工具、流程。而且，来自不同时间、地点和学科的相同或相近研究形成的相同或相似数据，因为难以共享和集成，不仅造成研究重复，也使得潜藏和融汇其中的价值很难被挖掘出来。

数字化感知、测量、实验仪器的发展、数字信息网络的发展，使得数字信息成为科研数据的基础原生态，数据的实时采集、在线提交、封装标识、关联组织、大规模存储成为数字科研的标配。科研数据不再仅是学术论文的附属物，而是科研的基础产出和“一等公民”（First Class Citizen）[4]，具有独立的身份识别[5]、属性描述[6]、监护机制[7]、溯源流程[8]，通过信息网络可发现、可获取、可互操作和可重用（FAIR 原则[9]），并逐步支持把数据监护和共享纳入科技界认可的学术贡献体系[10]。在整个科研数据生态体系中，建立方便、可靠、可测量、可验证、可支持 FAIR 原则的数据出版机制是重要基础，也是数据共享政策的实施条件之一。

1　科研数据开放共享和数据出版

数据出版（Data Publishing）是指通过一定的公共机制发布科研数据集，使得公众根据一定规则可以发现、获取、评价和应用这些数据集[11,12,13]。需要指出，这里的 Publishing 并不一定具有中文“出版”一词常带有的“正式的经过审批的机制”的含义。总体上，数据出版模式可分为三类：独立的数据出版即在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发布（附属于出版物

收稿日期：2019-02-09；修回日期：2019-03-04

* 通信作者，Email：klh@cnic.cn

的数据发布)、以数据论文形式发布(作为出版物本身的数据出版)[14]。其中,数据论文作为研究要素载体之一,被纳入到新兴的研究要素(Research Elements)出版体系,包括数据论文、材料论文、方法论文、软件论文等[15]。

我国在 2015 年发布的《促进大数据发展的行动纲要》[16]中明确提出"积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享"。2018 年 3 月,国务院办公厅发布的《科学数据管理办法》[17]中提出"主管部门和法人单位应积极推动科学数据出版和传播工作,支持科研人员整理发表产权清晰、准确完整、共享价值高的科学数据",并要求"科学数据使用者应遵守知识产权相关规定,在论文发表、专利申请、专著出版等工作中注明所使用和参考引用的科学数据"。

为了有效支持科研数据的开放共享,为了支持数据出版机制的建设、发展、监测和评价,本文对数据出版机制进行系统分析,并提出推动这些机制建立和发展的建议措施。

2 独立数据出版及其共享策略

独立的数据出版一般是科研人员自主或者按照项目方要求,将数据集存储到指定的数据存储库(Data Repositories)进行发布,并为数据集分配唯一标识符,以便检索和使用。这些存储库的案例包括国家建立的政府数据中心(如美国的 data.gov 等)、国家级的科学数据中心(如美国国家航空航天局数据中心[18],我国科学数据共享平台[19]等),以及各领域的专业数据中心(如世界数据系统 WDS[20]、全球蛋白质数据库 wwPDB[21]等);之外,诸如 Dryad[22]、Figshare[23]、或 ScienceDB[24]等公共存储库,也提供了公共的数据存储与发布服务。

多数国家的科研资助机构都制定了数据管理与共享办法,要求将科研产出的数据存储到指定数据库后开放共享。欧洲核子研究中心(CERN)将项目研究数据存储在分布于全球各地的存储库中,通过 CERN 开放数据门户(http://opendata.cern.ch)提供访问。CERN 研究数据共享政策[25]明确规定,相关研究数据集及其支持文档需尽早提交至指定的数据存储库(如 HEPDATA[26]),并根据项目要求提供开放共享。不同的 CERN 子项目对开放数据访问有不同要求,包括:1 级,与期刊出版物直接相关的数据必须通过适当的存储库提供开放共享;2 级,可根据项目发布政策进行有序的开放访问;3 级,可依据 CERN 数据级别和项目的相关要求,在合理时滞期后以合理方式发布相关数据,最长 10 年;4 级,数据不能提供开放访问,但鼓励使用 Creative Commons 许可和数字对象标识符 DOI。

截止到 2019 年 1 月底,根据 SHERPA Juliet 对研究资助者数据政策的统计[27],28%的资助方明确要求共享数据,15%的资助方鼓励数据共享。欧盟开放科学监控平台(EU open science monitor)对 Re3data(www.re3data.org)注册数据库的监控发现[28],2 414 个注册数据存储库中绝大多数对外开放,只有少数受限访问或禁止访问。在存储库拥有量方面,排名前三位的国家分别是美国(958 个)、德国(319 个)和英国(286 个),中国排名第 10 位(37 个)。学科分类中以生命科学领域的存储库最多。

总体而言,独立数据出版一般是依据资助方或项目方要求而执行,也是对数据作为研究资产进行管理的措施,数据共享多分为若干层级,可允许数据提供者在一定时间内对共享进行合理限制,不一定对存储数据进行 FAIR 化检查或转换,数据检索与利用政策因库而异,即使那些同样采用 Creative Commons 许可的存储库也往往存在差别。

3 数据作为论文辅助资料的出版及其共享策略

科研数据作为支撑论文研究结论的基础,是论文本身的有机部分,也是对论文结论进行验证的必要内容。当前,许多期刊都要求作者在投稿时必须同时提交支撑论文结论的数据集,供同行评议专家和编辑部审稿时核查,如 Science[29]、Nature[30]、BMC[31]等;有些期刊还要求作者将数据集提交到指定的可共享的数据知识库,供其他同行或公众查询和利用;更严格的期刊还将数据开放与共享作为论文发表的前提,如 PLoS[32]。

3.1 期刊数据透明度原则对数据出版的多级要求

"开放科学中心"(Center for Open Science)针对期刊出版提出《期刊透明性与开放性指南》[33](称为 TOP 指南),要求期刊在来源引用、数据、代码、研究材料、研究设计与内容分析、研究预注册和重复验证等方面透明开放。其中,关于数据透明度(Data Transparency),提出了 3 个渐强的等级要求:TOP 一级,声明(Disclosure),期刊所发文章必须声明支撑研究结果的相关数据是否可用,如果可用要提供访问说明;TOP 二级,强制要求(Mandate),期刊所发文章必须在受信任的存储库中共享支撑研究成果

的相关数据，并对数据及其使用技术条件进行准确描述，确因伦理道德或法律约束而无法共享数据则须说明并尽可能给出获取相应数据的其他方法；TOP 三级，验证(Verify)，期刊应在论文发表前利用作者提供的数据和方法来进行重复验证，如果论文结果不能重复，则可能导致论文被拒稿。

目前，很多出版商也制定了自己的分层化的数据共享政策，鼓励作者将数据存储在合适的存储库中，在论文中引用它，并提供数据可用性声明。例如 Elsevier[34]、Springer Nature[35]、Taylor&Francis[36] 和 Wiley[37]，都提供了相应政策来支持旗下期刊提供透明度。有人对上述四家主要出版社的数据出版政策支持 TOP 指南的程度做了调研[38]，见表 1(其中各出版社的“类型”相应于它们自己政策中的类型)：

在数据引用方面，各出版社通过实施 FORCE 11 数据引用原则联合声明[5]，支持作者像引用其他论文、书籍和网络信息一样引用数据，将数据作为参考文献的一部分来引用。

3.2　国际出版社期刊数据出版案例

Springer Nature 于 2016 年发布 4 类数据政策，提供了逐级增强的数据质量与数据共享要求(表 2)。在对 2017 年度所发表论文进行统计后[39]，它认为新的数据政策促进了数据的引用和共享，将计划对共享数据的论文予以开放数据标记(Open Data Badge)以推进数据共享。

尽管在出版社层面提出了总体的政策，但具体的数据出版与共享措施会在各个期刊层面进一步予以规定，而且不同学科数据共享的需求和进展可能存在差异。为此，本文作者选取了地学、医学、生态学、计算机科学中各学科 2018 年 SCI 来源期刊影响因子排名前 10 期刊，对其网站进行核查，根据其符合 TOP 指南等级程度分类，如表 3 所示。

统计可见，不同学科对数据的定义及共享需求不同；医学类期刊均具有一定的数据出版政策，计算机类期刊却较普遍地缺乏这类政策。生物医学领域在研究数据共享方面总体表现良好，根据 2017 年的关于 318 个生物医学期刊数据共享政策的统计[41]，其中 11.9%的期刊明确指出数据共享是论文发表的必要条件，9.1%的期刊要求数据共享但没有明确是否会影响出版决策，23.3%的期刊鼓励但不强制作者分享数据，9.1%的期刊间接提到数据共享，还有 31.8%的期刊没有提及数据共享。

表 1　国际主要出版社数据透明性程度

	鼓励但不强制要求共享	TOP 一级 数据可用性声明	TOP 二级 强制要求共享	TOP 三级 可重现验证
Elsevier	类型 A、B	类型 C	类型 D、E	
Springer Nature	类型 1、2	类型 3	类型 4	
Taylor&Francis	类型 1	类型 2	类型 3、4、5	
Wiley	类型 1	类型 2	类型 3	类型 4
其他	所有(仅)鼓励数据共享的期刊	• *Psychonomics Society Journals* • *Nature* • *Psychological Science* • *PNAS*	• *Science* • *PLOS* • *Royal Society Journals* • *Cognition*	• *AJPS* • *Biostatistics* • *JEPS* • *JPR* • *Meta-Psychology* • *QJPS*

表 2　Springer Nature 数据共享策略表[40]

共享策略特征	类型 1	类型 2	类型 3	类型 4
数据需存储在可支持的存储库中	强制	强制	强制	强制
允许数据引用	强制	强制	强制	强制
数据存储审查确认	不需要	可选	强制	强制
数据可用性声明	不需要	可选	强制	强制
验证数据的开放和标识(敏感/个人数据除外)	不需要	不需要	可选	强制
对其他数据引用的验证	不需要	不需要	可选	强制
数据经过同行评审	不需要	不需要	可选	强制
出版过程与数据存储库的集成	不需要	不需要	可选	强制

3.3 国内期刊数据出版政策现状

为了解我国期刊的数据出版政策，本文作者针对中国地学领域期刊，选取了 SCI 统计源期刊影响因子前 10 名和 CNKI 年报统计复合影响因子前 10 名中文期刊，对其网站上发布的数据政策相关信息也进行了核查。在 SCI 统计源 10 种期刊，具备 TOP 二级政策的期刊有 5 种，具备 TOP 一级政策的期刊为 3 种，另外 2 种期刊没有明确提及数据出版政策。CNKI 年报统计复合影响因子前 10 名中文期刊中，无一提及数据出版政策。

相较而言，我国期刊在数据出版政策上起步较晚，但也有少数期刊做出相关实践，例如《数据分析与知识发现》[42]在 2016 年就提出《支撑数据提交要求》，明确要求从 2016 年第 2 期起，所有发表论文必须提交支撑论文结论的数据，并发布数据可获得性声明。另外，《中华健康管理学杂志》从 2016 年起要求原始研究类稿件提供相应的原始资料，包括但不限于原始数据、原始结果、量表、干预方法、问卷等。《中华外科杂志》发布《关于投稿人自愿提供稿件支持原始数据的通知》，明确稿件一经录用，支撑数据将在文章发表的同时纳入国家人口与健康科学数据共享平台管理，发布匿名化的全部或部分数据，根据平台规则进行共享。

3.4 期刊论文支撑数据出版的政策要点

基于前述分析，作者总结了期刊研究论文政策数据的出版策略要点，如表 4 所示。

表 3 2018 SCI 统计源中部分学科 IF 排名前 10 期刊数据政策统计

期刊学科分类		无强制要求或说明	TOP 一级	TOP 二级	TOP 三级
地学	期刊数	3	4	2	1
	案例	*Reviews in Mineralogy & Geochemistry*	*Earth-Science Review*	*Nature Geoscience*	*Earth System Science Data*
医学	期刊数	—	4	6	—
	案例	—	*CA: A Cancer Journal for Clinicians*	*Lancet*	—
生态	期刊数	1	3	6	—
	案例	*Advances in Ecological Research*	*Ecology*	*Ecological Applications*	—
计算机	期刊数	8	2	—	—
	案例	*IEEE Journals*※	*Neural Networks*	—	—

（注：※ IEEE 提供了 IEEE DataPort，目前可为那些需要保留和管理其宝贵研究数据的人员免费上传不超过 2TB 的数据集，允许数据集所有者在论文或报告中引用和链接到其数据集，鼓励引用数据集。）

表 4 论文辅助数据的共享策略要点

策略要点	说明
数据定义	没有在文中体现的，独立验证研究结果所需的最小数据集
数据类型	包含通用数据类型（图、表、数据集、音频、视频等）和特定数据类型（例如生物医学方面的蛋白质序列数据、DNA 和 RNA 序列数据、大分子结构数据、微阵列数据、计算机代码和临床试验数据等）
数据版权	数据所有者（个人、组织或机构）拥有完整的数据版权及所有权
数据存储	特定数据类型应存储到推荐的专业数据存储库或通用型公共存储库中，其他数据也可以根据期刊要求以文件形式直接上传至期刊网站
共享策略	可根据期刊及学科情况按需共享—设置分等级共享策略，一般可分为： • 弱—鼓励作者共享 • 次强—支持作者共享，并提供可用声明 • 强—（强制）要求提交数据、提供数据可用性声明，并需经过同评
作者需确认并提供	确认数据可开放（共同作者许可） 数据准备及说明（注意避免出现隐私数据和其他敏感数据） 提供数据可用性声明
编辑加工	分配 DOI、元数据编辑和审查、生成数据集描述信息、增加项目资助信息、实现数据与论文的关联
数据审核	一般情况下没有强制要求
数据服务费	根据共享策略及服务项目设定收费标准
其他	开放数据标记的使用等

4　数据论文的出版现状与策略

数据论文(Data Paper)是由出版界和数据界共同提出和积极实践的新概念，指经过同行评议的专门对具有科学价值的数据集进行描述的论文，一般采用学术论文式的结构化模式，说明数据集的目的、产生与处理过程、数据内容、以及使用方式等。通过数据论文的出版，实现数据与论文的双向链接(图1)，促进数据更便捷地发现、获取和引用[43]，进而支持数据产权、标识、发现等核心问题的解决，最终可构建包括文献、数据和科学家的知识管理体系[44]。

目前，出版数据论文的期刊分为两大类：一类是纯粹出版数据论文的数据期刊，例如 Springer Nature 出版的 *Scientific Data*[45] 等；另一类期刊发表包括数据论文在内的各种类型论文的期刊，称之为混合型期刊，主要依托于传统的研究期刊，在论文出版的同时接收该论文相关的数据论文[46]。据国外相关研究统计[47]，现有出版数据论文的期刊多为混合型期刊，部分是只发表数据论文的数据期刊。

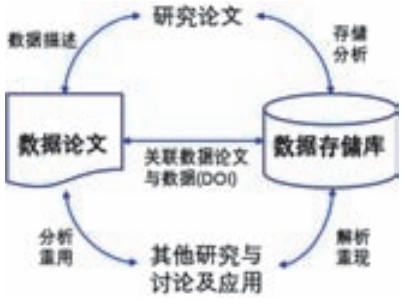

图 1　数据论文出版概念图

4.1　数据期刊现状

数据期刊相比较传统研究期刊，更强调数据发布的完整性、对数据的描述和重用，数据政策方面侧重于对数据的存储、元数据及数据本身及加工处理方法的描述，以及对论文和数据的质量评价。作者调研了 *Earth System Science Data*(ESSD)[48]、*GigoScience*[49]、*Scientific Data*(SD)，以及《中国科学数据》[50]和《全球变化数据学报》[51]等国内外具有代表性的基本数据期刊，总结了数据期刊的基本现状和数据出版政策要点(表5)。

4.2　中国数据论文出版现状

近年来，我国在数据论文与数据期刊方面呈现丰富实践(表6)，包括前述《中国科学数据》和《全球变化数据学报》等数据期刊，另有部分传统学术期刊专门设置了数据论文专栏。

其中，《中国科学数据》(*China Scientific Data*)是国家网络连续型出版物的首批试点期刊，也是目前中国唯一的面向多学科领域科学数据出版的学术期刊。该刊搭建了由数据出版在线工作平台(http://www.csdata.org)、科学数据储存库(ScienceDB)(http://www.sciencedb.cn)以及其他增值服务系统等组成的数据出版平台。在数据论文同行评议方面，根据数据的不同流程阶段，设置了数据审核员初审、同评专家评审和开放公众评议的多渠道分阶段评审机制，为每个数据集分配 DOI 标识和 PID 标识。《全球变化数据学报》于 2017 年创刊，与全球变化科学研究数据出版系统共同构成元数据、实体数据和数据论文的关联出版。其他期刊设置了数据论文专栏并已在数据出版方面做出了一些尝试，但在数据存储、开放共享、数据质量同行评议、数据引用等的政策方面尚未提供详细说明。

4.3　数据论文出版策略要点

作者对数据论文出版和共享数据的策略要点进行了分析，如表7所示。

4.4　不同数据出版模式的策略比较

作者对以上三种数据出版模式的优劣势和政策要点进行了初步的归纳和比较(表8)。

5　数据出版面临的挑战及建议

科研数据开放共享正逐步成为科研资助的基本要求和科研群体的行为规范，相关政策也在不断完善之中[53]。为有效支持数据共享，需要建立可靠的数据出版基础机制，包括数据存储库、研究论文支撑数据出版和数据期刊出版，尤其是数据期刊，提供了常规化、规范化、来源与标识清晰、质量可信、版权明晰、方便引用与激励的工具。但目前在我国，要形成方便可靠的数据出版，还面临一系列挑战：

(1) 还缺乏可信赖、可开放共享、可持续运行的数据存储库，已有的数据存储库在支持数据 FAIR 化上还较薄弱，权益管理机制还较粗糙，开放共享服务机制还缺乏长期可靠性，未尚未被多数国际学术期刊认可。

(2) 大多数中国学术期刊还没有论文辅助数据的发布与共享政策，已有政策的期刊由于自身平台局限或缺乏公共数据存储库支持，尚难以有效保存和共享作者提交的数据。

(3) 数据密集领域的大多数中国学术期刊还没有将数据论文作为自己学术内容的有机部分，即使少数已发表数据论文的期刊也尚未建立可靠机制来规范地评审和发表数据论文。

表 5 部分数据期刊的数据出版政策要点

	ESSD	GigaScience	SD	中国科学数据	全球变化数据学报
创刊时间	2009	2012	2014	2015	2017
学科	地学	综合	综合	综合	地学
累积发文量(篇)*	303	479	715	177	138
论文处理费(APC)	暂免	$1 056 起	$1 350 起	暂免	按版面收取
数据评审	√	√	√	√	√
数据存储库要求	推荐数据存储库	合作数据存储库 GigaDB 及其他任何数据存储库	合作存储库 Figshare 及其他得到认可的数据存储库	合作数据存储库 ScienceDB 及其他认可数据存储库	合作:全球变化科学研究数据出版系统

(注:该数据统计截止 2018 年 11 月 25 日)

表 6 我国学术期刊出版数据论文情况统计

期刊名称	所属学科	出版单位	栏目名称	创办时间	发表论文数
中国科学数据	综合	中科院计算机网络信息中心	数据论文	2015	177
全球变化数据学报	地球科学	中科院地理科学与资源研究所	数据论文	2017	138
生物多样性	生物科学	中科院生物多样性委员会	编目数据 数据论文	2014	100+
植物生态学报	植物学	中科院植物研究所	资料论文	2013	6
遥感技术与应用	地球科学	中科院遥感联合中心	数据论文	2016	4
图书馆杂志	情报学	上海市图书馆学会	数据论文	2018	2
Data Intelligence	数据科学	中科院文献情报中心	数据论文	2018	0
Big Earth Data	地球科学	国际数字地球学会	数据论文	2017	1
地理学报	地理科学	中科院地理科学与资源研究所	数据论文	2014	21
遥感学报	地球遥感	中科院遥感与数字地球研究所	数据论文	2018	0

(注:统计信息截至 2018 年 11 月 19 日)

表 7 数据论文的数据出版及其共享策略分析

数据政策要点	要点说明
数据论文提交	提出数据论文的主要规范
数据存储库推荐	按照学科,提供推荐或认可的数据存储库列表,包含通用存储库和学科专业数据存储库
数据存储库认定	提出可接受的数据存储库的认定标准,包括:在学科领域得到认可,可提供长期稳定的数据存储服务,可为数据集分配唯一标识符,非特殊情况下允许无条件的公开访问,可为作者和审稿人提供在数据发表之前的匿名访问服务等
特定领域数据标准	提供特定领域数据标准,作者应遵守现有的标准来准备和记录数据(FAIRSharing[52] 提供了很多有关特定领域数据标准的信息),以便专家参考审核
同行评议标准	对数据论文的数据质量进行评审的标准
权益规范	关于读者、作者、数据作者、其他(隐私数据的安全发布等)权益
数据标引	可为数据集分配唯一标识符,例如 DOI、Handle 等
数据引用原则	数据可用性声明、数据引用标准,以及是否列入参考文献
数据服务收费	各刊依数据集规格大小及服务项目设定收费标准

表 8　不同模式下数据出版策略比较

数据出版模式	优势	不足	数据政策要点
数据独立出版	基础设施专业，有利于数据的保存、获取、共享及其监测统计；可间接支持数据作者的科研贡献评价。	各学科发展不均衡；学科特性强，缺乏共同解决方案；存储库可持续性容易受影响；很多期刊论文与数据间互引还存在问题。	提供元数据，促进数据可发现和可理解；共享政策可依据资助者和项目管理者要求而定制，可分层级有限制地进行数据共享。
数据作为论文辅助资料出版	数据与论文密切相关，自然成为科研产出的一部分，论文对数据有一定的说明，便于理解和利用。	数据一般不是全部或完整数据集；部分数据的整理和聚合度较高，数据重用可能受限；论文审稿人对数据审查有限。	数据定义与说明；数据可用性说明；数据类型和存储要求；政策实施强度分类及说明；数据引用方法等。
数据论文出版	数据集较完整；数据说明清晰，有助于数据的解释和复用；数据通常存储在可靠稳定的存储库，便于访问及与其他数据集结合使用；数据经过同行评议，质量较高；论文可引用，有利于评价。	需加强期刊与数据库保持持久的双向链接；对大型数据集的同行评议目前还是一项挑战。	数据论文规范；提交数据论文和数据的方法；数据存储库的认证和推荐；期刊对数据论文及数据的评审；数据引用与相关权益人的利益保护等。

（4）中国的数据期刊数量仍很少，数据论文规范还有待完善和统筹协调；数据论文及其数据质量的同行评议仍是难点，多数情况下仍只能由作者自行对数据质量负责。虽有少量实践，但总体说来出版平台还缺乏支持数据评审的工具和流程。

（5）国家、学科和机构还缺乏可操作、可测量、可检验、可问责的科研数据开放共享政策及其实施措施，对科研数据的发布与共享的强制要求和激励措施还都缺失。

要纠正这种状况，当然需要多方协同建立强力推动数据共享的生态系统，例如欧盟委员会 FAIR 数据专家组 2018 年发布的《将 FAIR 数据变为现实：最终报告和行动计划》报告[54]提出的具体方法。但就本文作者而言，聚焦于支持数据共享的数据出版机制，提出以下建议：

（1）国家科技管理部门应加快出台科研项目数据共享的具体政策，要求项目研究数据通过可靠的数据出版机制进行组织、标识、描述、提交、发布和共享。

（2）国家和各学科应支持加快建立一批开放、规范、严格质量控制、支持数据 FAIR 化处理的数据存储库，支持数据永久保存和开放共享；它们应得到公共财政持续支持，接受可信赖性认证，接受对其公共服务的公开检查。

（3）国家科技管理部门和科技期刊管理部门应支持加快建立学术期刊研究论文支撑数据开放共享的强制性政策，建立支撑数据权益归属制度，建立支撑数据标识与引用规范，实行支撑数据在可公共共享的数据存储库进行存缴和共享制度，实行支撑数据可获得性声明制度，并逐步建立支撑数据评审机制与工具。

（4）国家科技管理部门和科技期刊管理部门应加快建设和支持一批规范运行的数据期刊，逐步覆盖所有适用领域，建立数据论文规范及其评审规范，实行数据论文相关数据文档在可共享数据存储库存缴和共享的制度，并逐步建立数据论文的评审支撑工具和平台。

（5）国家科技管理部门应加快建立对数据出版的保护与激励机制，包括明确数据权益、完善数据使用许可、建立数据隐私保护规范等，提供数据下载引用统计服务，将数据出版结果纳入科研成果体系和科研人员贡献度和影响力评价中。

参　考　文　献

[1] Tony H, Stewart T, Kristin T. The Fourth Paradigm: Data-Intensive Scientific Discovery. Washington: Microsoft Research, 2009.

[2] Royal Society. Science as an open enterprise. https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/. (2012-06-21) [2018-11-04].

[3] Open Science. https://en.wikipedia.org/wiki/Open_science. (2019-2-28) [2018-11-20].

[4] Bolikowski L, Houssos N, Manghi P, et al. Data as "First-class Citizens". *D-Lib Magazin*, 2015, 21(1—2). DOI:10.1045/january2015-guest-editorial. [2018-11-20].

[5] Data Citation Synthesis Group. Joint Declaration of Data Citation Principles. San Diego CA: FORCE11, 2014. https://doi.org/10.25490/a97f-egyk. [2019-01-04].

[6] DCC. Discipline Metadata. http://www.dcc.ac.uk/resources/metadata-standards. [2019-01-04].

[7] Data Curation Center. http://www.dcc.ac.uk/. [2019-01-04].

[8] 沈志宏，张晓林. 语义网环境下数据溯源表达模型研究综述. 现代图书情报技术，2011，27(4)：1—8.

[9] LIBER. Implementing FAIR Data Principles—The Role of Libraries. https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf. [2019-01-05].

[10] Bierer B，Pierce H，Drazen J，et al. Credit for Data Sharing. Berlin：FORCE 2017，October 25—27. https://dataverse. org/files/dataverseorg/files/creditfordata-force2017.pdf. [2019-01-05].

[11] Costello M J. Motivating online publication of data. BioScience，2009，59(5)：418—427. DOI：10.1525/bio.2009.59.5.9.

[12] Smith VS. Data publication：towards a database of everything. BMC Research Notes，2009，2(113). DOI:10.1186/1756-0500-2-113.

[13] Lawrence B，Jones C，Matthews B，et al. citation and peer review of data：moving towards formal data publication. International Journal of Digital Curation，2011，6(2)：4—37. DOI：10.2218/ijdc.v6i2.205.

[14] 张晓林，沈志宏，刘峰. 科学数据与文献的互操作// 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.

[15] Elsevier. Research Elements. https://www.elsevier.com/authors/author-services/research-elements. [2019-01-04].

[16] 国务院. 促进大数据发展行动纲要. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm. (2015-09-05) [2018-11-24].

[17] 国务院. 科学数据管理办法. http://www.gov.cn/home/2018-04/02/content_5279296.htm. (2018-04-02) [2018-11-04].

[18] NASA's Data Portal. http://data.nasa.gov. [2019-02-04].

[19] 国家科技基础条件平台中心. 中国科技基础资源共享网. https://escience.org.cn/. [2019-01-04].

[20] The World Data System (WDS). http://www.icsu-wds.org. [2019-01-04].

[21] The Protein Data Bank archive (PDB). http://www.wwpdb.org. [2019-01-04].

[22] Dryad. http://www.datadryad.org. [2018-11-04].

[23] FigShare. http://figshare.com. [2018-11-04].

[24] ScienceDB. http://www.sciencedb.cn. [2018-11-04].

[25] CERN Open Data Policies. http://opendata.cern.ch/collection/Data-Policies/. [2018-11-04].

[26] An overview of the data held in SHERPA Juliet. http://hepdata.cedar.ac.uk/. [2019-01-04].

[27] Funders Policy Overview. http://v2.sherpa.ac.uk/view/funder_visualisations/1.html. [2019-01-04].

[28] EU open science monitor—open data. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en. [2019-02-04].

[29] American Association for the Advancement of Science. Science：editorial policies. http://www.sciencemag.org/authors/science-editorial-policies. [2019-02-04].

[30] Springer Nature. Availability of data，material and methods. http://www.nature.com/authors/policies/availability.html. [2018-11-04].

[31] BioMed Central. Open Data. http://www.biomedcentral.com/about/policies/open-data. [2019-01-04].

[32] PLOS. Data Availability. http://journals.plos.org/plosone/s/data-availability. [2019-01-04].

[33] Brian Nosek，George Alter，George Banks，et al. Transparency and Openness Promotion (TOP) Guidelines. DOI:10.1126/science.aab2374. https://osf.io/vj54c/. (2018-07-02) [2019-01-04].

[34] Elsevier's Research Data Guidelines for Journals. https://www.elsevier.com/authors/author-services/research-data/data-guidelines. [2019-01-04].

[35] Springer Nature's Research Data Policy Types. https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096. [2019-01-04].

[36] Taylor&Francis' data sharing policies. https://authorservices.taylorandfrancis.com/understanding-our-data-sharing-policies/. [2019-01-04].

[37] Wiley's Data Sharing Policies. https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html. [2019-01-04].

[38] Mellor D. The Landscape of Open Data Policies. https://cos.io/blog/landscape-open-data-policies/. (2018-11-07) [2019-01-04].

[39] Data availability report for EXAMPLE UNIVERSITY 2017. https://www.springernature.com/gp/open-research/institutions/research-data-services/data-availability-reporting. [2018-11-04].

[40] Springer Nature Research Data Policies FAQs. https://www.springernature.com/gp/authors/research-data-policy/faqs/12327154. [2018-11-04].

[41] Vasilevsky N A, Minnier J, Haendel M A, et al. Reproducible and reusable research: Are journal data sharing policies meeting the mark? PeerJ, 2017, 5(10):e3208. https://doi.org/10.7717/peerj.3208.

[42]《现代图书情报技术》支撑数据提交要求. http://www.infotech.org/Jwk_infotech_wk3/fileup/1003-3513/NEWS/20160408165409.pdf. [2018-11-20].

[43] 何洪林，杨萍，于贵瑞.《中国生态系统研究网络(CERN)专刊》卷首语. 中国科学数据，2017，2(1). DOI: 10.11922/csdata.0.2017.0136.

[44] 吴立宗，王亮绪，南卓铜，等. 科学数据出版—促进科学数据共享的一种新模式. 中国科技资源导刊，2014 (3): 72—78.

[45] Scientific Data. https://www.nature.com/sdata/. [2019-01-04].

[46] 孔丽华，邵明玥. 科学数据出版内容与案例分析. 科研信息化技术与应用，2018，9 (6): 39—46.

[47] Candela L, Castelli D, Manghi P, et al. Data Journals: A Survey. Journal of the Association for Information Science and Technology, 2015. https://doi.org/10.1002/asi.23358. (2015-01-30) [2019-01-04].

[48] Earth System Science Data. https://www.earth-system-science-data.net/. [2018-11-04].

[49] GigaScience. https://academic.oup.com/gigascience. [2018-11-04].

[50] 中国科学数据. http://csdata.org. [2018-11-20].

[51] 全球变化数据学报. http://www.geodoi.ac.cn/WebCn/Default.aspx. [2018-11-20].

[52] FAIRSharing. https://fairsharing.org/. [2018-11-04].

[53] 张晓林. 实施公共资助科研项目研究数据开放共享的政策建议. 中国科学基金，2019，33(1): 79—87.

[54] EU Publications. Turning FAIR data into reality: Final report and action plan from the European Commission expert group on FAIR data. https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283. (2018-11-26) [2019-01-04].

Trends and challenges in research data publishing

Kong Lihua[1,2,3]　Xi Yan[3]　Zhang Xiaolin[1]

(1. *National Science Library, Chinese Academy of Sciences*, 100190 *Beijing*;
2. *Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences*, 100409 *Beijing*;
3. *Computer Network Information Center, Chinese Academy of Sciences*, 100190 *Beijing*)

Abstract　Data publishing is one of the key means to stimulate data dissemination and promote data sharing. Based on surveys of data repositories, major STM publishers, representative scholarly journals and data journals, this paper analyzes the current mechanisms of data publishing, such as sharing by depositing into data repositories, publishing as supplementary data files of journal articles, and publishing as data papers in research journals and data journals. The paper also summarizes the key elements of the policies related to those mechanisms, points out the challenges to data publishing in China, and recommends some course of actions to develop data publishing for the benefit of data sharing.

Key words　open data; data publishing; supplementary data file; data journal; data repository; policy