引 用 格 式:Li Hongxing,Wu Lizong,Nan Zhuotong,*et al*.Collaborative Publishing of Scientific Data:Model and Implementation[J].Remote Sensing Technology and Application,2016,31(4):801-808.[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

科学数据联合出版模式与内容研究

李红星[1],吴立宗[1],南卓铜[2],王亮绪[3],王 建[1],赵彦博[1],张耀南[1],马瀚青[1,4]
(1.中国科学院西北生态环境资源研究院,甘肃 兰州 730000;
2.南京师范大学地理科学学院,江苏 南京 210023;
3.上海师范大学城市发展研究院,上海 200234;
4.中国科学院兰州文献情报中心,甘肃 兰州 730000)

摘要:科学数据出版是推动和深化数据共享的有效手段。本文提出一种科学数据出版模式——数据中心联合传统学术期刊的科学数据出版。该模式通过数据中心与传统学术期刊的合作以数据论文的形式实现数据的正式出版,其中通过数据中心实现数据的永久存储与管理,并开展数据挖掘与共享,数据论文则经过严格的同行评议发表在传统期刊上。该模式不但能够提升数据质量,而且有助于数据成果被现有的科研成果评价体系所接受,激发数据作者出版、共享数据的积极性,有效解决数据共享长期面临的数据质量控制和数据作者权益保护两个关键问题,进而达到整合优质数据资源、深化数据再利用的效果,实现科学数据共享可持续发展。这种出版模式结合了数据中心和学术期刊各自的优势,组织形式灵活,具有易操作可重复的特点,也可供地学以外其他学科借鉴。

关 键 词:科学数据;数据论文;数据中心;学术期刊;联合出版
中图分类号:TP 79 **文献标志码**:A **文章编号**:1004-0323(2016)04-0801-08

1 引 言

有研究表明,仅过去几年全球新增的数据量就已达到人类有史以来全部数据量的总和,而且未来还会以更高的速度激增[1-2]。与此同时,数据的价值也被重新认识,数据已经被认为是一种重要的可再利用资源[3],在科技创新中扮演着重要角色,数据发现也成为继试验、理论、计算模拟之后的第四种科学研究范式[4]。科学数据共享作为数据流通和再利用的关键手段在数据密集型科技时代变得更加必要和迫切。国际社会对数据共享已经普遍重视,近十几年数据共享基础设施和政策法规体系建设发展迅速[5-15],数据中心的共享实践也取得了显著的成效[16-17]。尽管如此,科学研究中的数据瓶颈仍未打破,被共享与再利用的数据仅是产出数据的冰山一角[18-19]。虽然相关的数据政策和法规相继出台,要求受政府经费支持的项目产出数据均开放共享或强制存缴[8-9,20-23],但由于数据成果并不包括在目前的科研评价体系中,数据作者权益得不到很好的体现和保护,使数据共享和存缴缺乏自主动力,数据在数量和质量上都难以得到保证[18-19,24-25]。在这种情况下,科学数据出版的概念被数据共享界和出版界共同提出,希望参照学术期刊的模式通过出版开展数据共享,一方面引入期刊成熟的同行评议体系保证

收稿日期:2016-04-30;**修订日期**:2016-07-24
基金项目:国家科技基础条件平台—国家地球系统科学数据共享平台冰川冻土科学数据中心,国家自然科学基金项目"面向黑河流域生态—水文过程集成研究的数据整理与服务"(91025001),"服务于'面向黑河流域生态—水文过程集成研究'的验证数据库建设"(91325103),中国科学院兰州文献情报中心主任基金项目(Y6AJ083009)资助。
作者简介:李红星(1981—),女,甘肃庆阳人,工程师,主要从事积雪遥感及科学数据共享研究。E-mail:lihongxing@lzb.ac.cn。
通讯作者:南卓铜(1977—),男,浙江温州人,教授,主要从事水文水资源决策支持系统和建模环境、及科学数据共享研究。E-mail:nanzt@njnu.edu.cn。

数据质量,一方面以正式出版促进数据成果被科研评价体系认可,从成果体现和权益保护的角度来提高数据作者的共享积极性,进而提高科学数据的利用率,充分体现其潜在价值。

科学数据出版的内涵是指将数据作为一种重要的科研成果,从科学研究的角度对科学数据进行同行评议和公开公布,创建标准和永久的数据引用信息,供其他研究性文章引证[26]。目前国际上科学数据出版的相关实践工作已经不少[27]。例如全球生物多样性机构(Global Biodiversity Information Facility,GBIF)成立专门工作组针对科学数据出版框架进行了详细讨论[28-32],并面向相关的组织机构(不面向个人)提供生物多样性数据的在线出版服务,通过为元数据注册 DOI 实现数据的规范引用并统计数据使用情况。为了更好地推进科学数据出版,鼓励科学家花更多的精力整理手中的科学数据用于出版和共享,一些出版集团创建了专门的数据期刊,用于刊登数据论文(Data paper),实现科学数据出版,如 2009 年创刊的《Earth System Science Data》、2011 年创刊的《Journal of Open Archaeology Data》、2012 年创刊的《Geoscience Data Journal》,甚至《Nature》也于 2013 年创办了数据子刊《Scientific Data》,2015 年,中国科学院计算机网络信息中心也创办了中国第一家数据期刊《中国科学数据》[33]。此外,许多传统学术期刊为了给研究论文评审提供依据,同时为了促进数据再利用与验证要求论文出版前将相关数据存储在公共数据中心或期刊自己的数据库,并为数据注册 DOI,客观上实现了数据与研究论文的集成出版。

总体来看,较常见的数据出版主要有 3 种形式:①传统研究论文并提供数据下载信息;②基于 DOI 系统的数据网络出版;③通过专门的数据期刊以数据论文的形式出版。其中,传统研究论文对数据自身的成果性体现很弱甚至缺失,发表周期往往较长。网络出版虽然简便快速、周期短,但出版过程中一般缺少专业评审机制,数据质量及可用性不能得到全面的保障,而且在数据引用及成果认可方面被接受程度较低。数据期刊的形式能有效解决前两种出版形式的弊端,但数据期刊作为一种新兴的期刊类型数量非常有限,在中国创办新刊申请严格且投入大、周期长,无法满足数据出版的需求;此外,数据期刊综合性较强,与现在科学研究越来越专业化的趋势不一致,对评审专家库要求很高。

鉴于已有数据出版形式的不足,一种可行的方式是与已有的传统学术期刊结合,充分利用期刊现有资源,将数据论文发表在已有学术期刊上,而数据本身存储在可靠的数据中心。目前这种方式已经有一些尝试,如 GBIF 与 Pensoft Publishing 出版集团建立合作,鼓励数据提供者将元数据转化为数据论文在该出版集团相关学术期刊上发表[34];中国科学院地理科学与资源研究所与多家单位联合建成的"全球变化科学研究数据出版系统",数据论文以长摘要形式发表,第一期数据论文以增刊形式在《地理学报》出版[35]。寒区旱区科学数据中心(原中国西部环境与生态科学数据中心)是世界数据中心成员之一,从 2009 年开始与中国科学技术信息研究所合作,依托国家自然科学基金重大研究计划及大型科学试验在国内首次开展了数据出版试点研究[24,26,36-39],是《Scientific Data》及《PLOS One》等国际刊物的认证数据永久存储机构,同时被汤森路透遴选为国内首个开展 DCI(Data Citation Index)合作的数据中心。该中心目前正在探索联合当地多家地学期刊(5 家中文期刊,1 家英文期刊)开展数据出版,将通过同行评议的数据论文与研究型论文同等对待,发表在传统学术期刊上。本文将结合寒区旱区科学数据中心的实践经验,构建一个数据中心和多期刊联合进行数据出版的模式,并探讨其中的技术步骤、关键技术和相关问题。

2 联合出版模式及流程

科学数据联合出版是指数据中心与传统学术期刊通过合作以数据论文的形式完成数据的正式出版,其中通过数据中心实现数据的永久存储与管理,并开展数据挖掘与共享,数据论文则出版在传统学术期刊上。这种联合出版模式逻辑上可以是多对多的合作关系,即多家数据中心与多家期刊联合,在机构数量及交叉合作关系上没有限制,但一个数据集其数据实体及数据论文只应在唯一的数据中心和期刊上发布与出版。此模式借鉴了传统学术论文出版的方式,并利用了数据中心和学术期刊两种机构各自的优势。

联合出版包含科学数据和数据论文两个内容:出版的数据覆盖所有不涉及国家机密、不存在知识产权争议的原创性科学数据成果;数据论文作为科学数据出版的体现形式,在目标与内容上有别于普通研究型论文,其主要功能是对数据制备及使用的相关信息进行详细的描述,达到通过数据论文可对数据科学性及质量进行评判和指导再利用的水平。

数据和论文出版前都要经过专业评审,而且由于在该模式中数据论文发表在传统学术期刊上,所以要经过与研究型论文同样严格的同行评议。出版涉及的机构有数据中心和学术期刊两方面,其中数据中心负责数据的投稿接收、技术评审、永久存储、管理维护及发布和共享,期刊编辑部负责数据论文的投稿接收、科学评审及出版。出版过程中数据中心为期刊提供评审专用的数据下载通道,同时双方互相反馈数据与论文的评审意见,最终将出版数据与论文关联。联合出版具体流程包括以下 4 个阶段(图 1)。

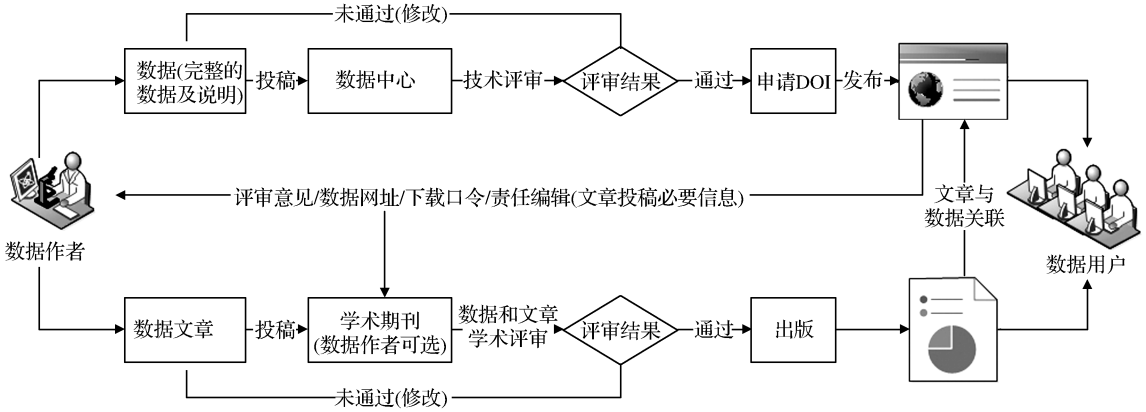


图 1　科学数据联合出版流程

Fig.1　The process of collaborative publishing for scientific data

第一阶段:数据投稿与评审。数据作者将整理好的数据(包括数据说明信息)投稿至数据中心,数据中心负责对数据进行技术评审,并通过评审—修改—再评审机制使数据在形式上达到发布标准,评审通过后数据中心为数据指定临时数字对象标识符(Digital Object Identifier,DOI)并开具“数据接收函”。“数据接收函”包含数据技术评审意见、数据 DOI 信息、数据下载口令及数据责任编辑联系方式等信息,是数据论文投稿的必要信息。在该阶段数据并未正式发布,仅是在数据中心存储,并为下阶段的科学评审提供访问权限。

第二阶段:数据论文投稿与评审。数据作者选择相应的期刊并凭“数据接收函”进行数据论文投稿,期刊根据数据论文内容指定专家(数据作者及数据中心亦可推荐专家)对论文与数据的科学性进行同行评议,数据作者根据评议意见对论文及数据进行修改,最终评议结果不但反馈给数据作者,同时将发送至数据中心。

第三阶段:发布出版。如同行评议最终认为数据具有科学价值,且论文与数据均达到出版水平,则数据中心将技术评审阶段为数据指定的临时 DOI 注册为永久 DOI,数据在数据中心正式发布,数据论文由期刊正式出版,数据与数据论文通过 DOI 建立关联,并分别给出规范的引用方式。如数据的科学价值得到认可,但论文的组织未达到评审要求,数据可先行发布,论文经修改并通过评审后再出版。若数据的科学意义被否定,则论文与数据均被拒稿,数据中心将不再提供数据的存储服务。

第四阶段:开放共享。数据经正式出版后即面向公众开放,进入共享阶段。数据共享方式包括在线共享和离线共享两种,由作者进行指定,其中在线共享是指用户注册登录数据中心网站填写个人及数据用途信息并同意数据共享许可协议后即可获取数据下载权限;而离线共享则需要作者对申请者个人信息和数据的用途信息进行审核,审核通过后用户方可获取数据下载权限。数据中心安排专职人员负责响应用户的数据申请及相关咨询,并保证数据审核及下载过程畅通。在整个共享过程中数据中心将持续记录数据在网站被使用的情况,如数据访问量、下载量及用户和用途等信息,同时对基于数据正式发表的科研成果通过期刊引用、DOI、DCI、主要文献数据库和搜索引擎等方式进行追踪,并根据以上信息对数据共享成效及数据质量进行评价,定期生成评价报告并反馈给作者。

3　关键技术

3.1　数据论文

数据论文是一种新的论文类型,其核心目标是详细描述科学数据并实现数据的正式出版。不同于元数据或数据文档,它不但包含数据的基本说明信息,还要提供数据生产的背景及详实的数据生产方法和制备过程,系统描述数据的质量,同时对数据的

科学意义和适用性,包括潜在用途等信息进行介绍,如有可能还应提供一些相关的应用案例和使用技巧,为数据再利用提供充分的有价值信息。同时,数据论文也不同于传统的研究型论文,它关注于数据本身,而不是利用数据针对某个科学问题进行分析,所以数据论文中不应包含新研究或新发现。也就是说,数据论文与基于相同数据的研究论文是不冲突的,且二者互为补充,其发表顺序也没有任何限制,作者可根据具体情况自行安排。

数据论文作为科学数据出版的信息载体,一篇论文中可以仅包含一个完整的数据,也可根据应用需求包含多个相关的数据。形式上数据论文与传统研究论文相似,也要投稿给学术期刊并通过同行评议,具体组织与格式遵照相应期刊的具体要求,但其主要内容应围绕数据本身的描述,寒区旱区科学数据中心的出版实践是在参考 Nature 集团数据期刊《Scientific Data》的基础上,要求数据论文涵盖数据生产背景、生产方法、组织形式、质量验证、应用案例(可选)、数据获取与使用说明,此外,论文中应简要说明所有作者在数据生产中各自的贡献,并对数据知识产权进行明确声明。

3.2 评审机制

数据论文及数据评审是出版过程中对数据质量把关的重要环节,不但能有效杜绝数据篡改、伪造等学术造假问题,而且从数据完整性和规范性的角度提升数据质量和可重用性,并确保出版的数据具有科学意义和再利用价值。数据出版过程中的评审分为技术评审和科学评审。数据通过技术评审后方可进入科学评审环节,通过科学评审后方可正式发布与出版。

技术评审主要针对数据及其说明信息的完整程度和数据格式的标准性展开,由数据中心技术人员完成。科学评审则主要针对数据的生产方法与制备流程的科学性和严谨性以及数据的应用价值展开,是评审的核心部分,由学术期刊负责完成。科学评审采用已经非常成熟的同行评议形式,依托数据论文对数据的详细描述并结合对数据实体的具体考察进行。由于科学评审包含论文评审和数据评审两部分,加之数据论文在目标、内容与形式上的特殊性,使评审有别于传统研究论文,故期刊与数据中心将联合共建适用于数据论文评审的专家库。

3.3 引用机制

推动数据引用的规范化是数据出版的重要目标之一。规范的数据引用不但能保护作者权益,还能实现数据与相关论文、相关数据的关联,有助于促进数据挖掘与知识发现,同时便于追踪准确的数据引用绩效,作为数据再利用及影响力的度量[24,40-41]。而科学数据在网络环境发布后,往往面临地址易变、多重链接、数据链接地址不能保证永久有效、通过网络地址的引用方式失去引用追溯作用等问题。针对这些问题,寒区旱区科学数据中心通过为出版数据注册永久性的唯一逻辑标识码 DOI 来构建规范的数据引用机制,取得了较好的效果。DOI 信息中包含数据作者、生产单位及发布机构等完整版权信息,而且可以实现对数据准确、永久定位,以保证引用数据的可追溯性。

寒旱区数据中心目前注册数据达 500 余条,为发布数据自主设计 DOI 代码,具体形式为 10.3972/westdc.001.2015.db(中文网站)和 10.3972/card.001.2015.db(英文网站),其中 10.3972 为中国 DOI 注册中心分配给数据中心的独立 DOI 前缀,为不可变更信息,westdc 和 card 分别为数据中心中文版和英文版网站的代码,对于大型试验或重大科学计划的系列数据,也可根据需要将数据中心代码更改为试验或科学计划代码,001 表示数据的序号,2015 表示注册年份,数据序号与年份由数据具体注册时间决定。

但由于数据引用还未纳入到学术期刊引用标注的标准体系中,目前大部分期刊仍不认可数据的直接引用方式。所以出版数据的引用可采用多种形式供选择:①直接引用数据;②引用数据论文;③引用数据相关研究论文。

3.4 评价机制

科学数据共享后发挥了多少作用、产生了如何的影响是数据作者、数据生产资助方和数据出版发布机构都非常关注的问题,所以有效的数据评价机制是数据出版的必要组成部分。评价的作用不仅在于评价本身,还在于它将成为激励出版和自主共享的原动力。

对于科学数据而言,数据的访问量、下载量及引用量均可作为其共享成效和影响力的评价指标。但相比之下,访问量和下载量不能代表数据的实际引用情况,仅能作为数据潜在应用可能的度量[24],而引用量则体现了数据再利用后成果发表的真实情况,不但能从引用数量上进行评价,还可根据发表成果的质量(如发表期刊的影响因子和论文本身的引用次数等)来评价数据质量的高低,利用数据发表的成果质量越高,则说明数据的质量越可靠,并在以上

数量与质量评价的基础上,给出数据的综合影响力。

基于 DOI 的引用机制为数据引用评价提供了有力的技术保障。通过 DOI 对数据的唯一标识,不论数据在论文中以上一节给出的任何方式进行引用,均可由搜索 DOI 来准确追踪引用数据的论文。

寒区旱区数据中心将为出版数据定期统计访问量、下载量及引用量,评价数据应用成效和影响力,并出具评价报告,同时也根据作者要求随时提供评价服务。

此外,数据用户的直接反馈(如数据使用过程中发现的问题或通过再利用对数据的验证结果等)也是数据评价的重要信息,这部分内容虽然不便于量化评价,但对于提高数据质量具有不可忽略的意义,故数据中心将收集整理用户反馈并及时发送给数据作者。

3.5　共享许可及版权转让

科学数据出版面向的对象为不涉及国家机密、不存在知识产权争议的原创性数据成果。数据一经出版即面向公众展开完全或限制共享。在保护作者知识产权的前提下,为了尽可能使科学数据通过共享充分发挥价值,数据中心要求数据作者放弃出版数据的再分发权和再编译权,仅保留署名权和限制商业使用的权利,同时从知识共享许可协议(Creative Commons license,简称 CC)中选取署名(BY)和署名－非商业性使用(BY－NC)两种许可协议供数据作者选择。

数据出版前作者针对数据实体和数据论文分别与数据中心及期刊编辑部签署版权转让授权书。其中数据论文的版权按照期刊已有的模式转让即可。数据实体的授权主要包含:①作者声明数据为原创,且不涉及国家保密及任何知识产权侵权行为;②数据作者署名排序无争议;③作者选定一种共享许可协议(BY 或 BY-NC);④允许数据中心在编辑过程中对数据的格式及文字说明信息等内容进行适当修改和加工;⑤数据再发布权归数据中心所有。

4　讨　论

数据中心与已有学术期刊联合实现科学数据出版是一种非常可行且高效的途径。一方面学术期刊是正式的出版机构,数据论文参考一般科研论文发表的流程进行出版,不仅易操作而且更容易被学术界接受;另一方面,数据中心作为专业的数据管理、发布和共享机构,在数据的永久存储、系统管理和开放共享等方面都具有齐备的软硬件设施和丰富的经验,而且有稳定的专职人员团队维护从数据投稿接收到共享分发全过程的运行畅通。通过这种形式灵活的合作,不但充分发挥了数据中心与学术期刊两者各自的优势,避免了重复建设造成的资源浪费,而且在提升数据质量、保护数据生产者权益和整合优质数据资源等方面具有显著优势:

(1)提升数据质量和可用性。大量的科研产出数据在用于共享时,被指出缺少必要的说明信息,从而导致数据不可用。一些数据经再利用后被发现数据准确率较低甚至错误,这些问题严重阻碍数据的再利用。通过联合出版,数据接受严格的同行评议和技术评审,很大程度上排除了造假数据,数据的科学性及利用价值得到权威认可;其次,经评审后数据及元数据在内容的完整性和格式及组织的规范性上都达到了较高的水平,有效避免了数据文件残缺、组织混乱或缺少必要的说明信息等阻碍数据再利用的问题。对于目前科研产出数据体量巨大且数据质量参差不齐的状况,可以说数据出版是一个对数据去伪存真、优胜劣汰及规范优化的过程,出版后的数据在质量和可用性上都将得到提升。

(2)保护数据生产者权益。通过出版,数据获得唯一和永久性的标识符 DOI,数据的生产者、生产单位及发布机构等版权信息得以正式声明,并形成规范的数据引用方式,数据共享中相关的责权利也给出明确的定义,避免了数据被粗暴使用(未经许可或不引用不标注)、甚至数据成果被窃取等问题,从根本上保护了数据生产者的著作权、署名权等知识产权。同时,数据的论文引用方式也可以显著提升论文的引用率,使数据作者受益。

此外,期刊论文作为学术成果的体现形式已经被科研界普遍认可,是科研人员工作评价的重要指标。与已有的学术期刊联合以论文的形式实现科学数据出版,不但体现了数据是科研成果的属性,而且可以无缝地融入目前的科研成果评价体系,改变以往数据被认为是科研工作的附属产出而非正式科研成果,数据生产者的贡献被忽略,相应的利益无法体现的局面。当出版数据数量及数据影响力成为具体的指标与评价及奖励机制直接挂钩,生产者的积极性将得到有力地激发,他们将愿意出版更多的数据在更大的范围内共享,使数据共享从被动转变为自发。

(3)整合优质数据资源。如上所述联合出版机制能够激发数据生产者积极性,吸引更多科研人员投稿,通过数据评审,大量优质的数据资源被筛选整

合,彻底解决了目前数据资源流通不畅甚至严重闲置和浪费而使用者却拿不到需要的数据的矛盾,以及数据中心数据陈旧缺乏新数据源的问题。所以数据出版无疑是数据共享大环境形成和可持续发展的强劲推动力。

科学数据联合出版具体实施中还存在以下需要重视和解决的问题:

(1)数据永久保存和共享保障。数据出版后数据安全和永久保存与共享是需要保障的首要问题。联合出版的数据中心最好经过权威认证,具有长期运行的经费支持,能保证与学术期刊建立长期稳定的合作关系,永久存储数据并开展共享。在进入联合出版系统时,应要求数据中心制定可持续发展计划,对数据中心可能出现的问题(如被撤销或没有后续支持)及数据如何转移至另一个数据给出预先解决方案,以确保数据安全和持续数据共享。

(2)数据评审专家库建设。数据评审与传统研究论文评审相比具有自身特点,审稿人不但应具备相应专业知识背景,最好还有使用或制备同类数据的经验,具备相应的软件使用技能;同时审稿观念也应是全新的、不同于研究论文的,通过评审应将所有具有再利用价值的数据筛选出来,而非仅仅关注数据本身表达的科学含义。所以数据中心与学术期刊共建数据评审专家库十分必要。

(3)数据投稿系统和期刊投稿系统对接。数据投稿由数据中心负责接收,如何与期刊已有的文章投稿系统对接也是需要重点解决的问题之一。对接后应能实现投稿过程中两个系统的切换、数据与文章的关联及评审意见的相互反馈等功能。

5 结 语

科学数据出版能有效解决长期以来数据共享面临的两个核心问题:数据作者权益保护和数据质量控制。寒区旱区科学数据中心提出了联合多家传统学术期刊的数据出版模式,能有效解决以上两个问题,而且与创建数据期刊的出版模式相比较,联合出版模式具有优化利用现有资源、形式灵活、易操作可复制的特点。

科学数据联合出版涉及数据中心和学术期刊两方,主要包括数据论文、数据评审、数据引用和数据评价 4 个重要环节。联合出版能很好地提升数据质量及可用性,结合数据中心成熟的共享体系,使数据变得更易发现、易获取,也更加易用,能很好地促进和深化数据再利用,体现数据价值;同时以数据论文形式体现数据作者的科研成果,通过引用机制达到对数据贡献的持续追踪和评价,有助最大化体现数据作者的权益,进而提高数据作者共享积极性,有助于吸引和整合优质数据资源,使数据共享可持续发展。寒区旱区科学数据中心的实践是以地学专业为例,但该模式可以被其他科学专业复制和借鉴。

目前,我国科学数据联合出版尚处于试点阶段,仍需要更多国家层面的政策和经费支持:首先,要建设、维护权威数据中心的长期业务运行,保障数据安全及永久存储及共享;其次,应鼓励和支持数据中心与学术期刊加强合作,尤其是高水平的期刊,同时,应出台鼓励数据引用和评价的相关政策,将数据引用正式纳入到期刊文献引用的范畴中,以提高数据出版积极性。

参考文献(References):

[1] Hilbert M, Lopez P. The World's Technological Capacity to Store, Communicate, and Compute Information[J]. Science, 2011, 332(6025): 60-65.

[2] Adshead A. Data Set to Grow 10-fold by 2020 as Internet of Things Takes Off[Z]. 2014.

[3] Li Deren, Zhang Liangpei, Xia Guisong. Automatic Analysis and Mining of Remote Sensing Big Data[J]. Acta Geodaetica et Cartographica Sinica, 2014, (12): 1211-1216. [李德仁, 张良培, 夏桂松. 遥感大数据自动分析与数据挖掘[J]. 测绘学报, 2014, (12): 1211-1216.]

[4] Hey T, Tansley S, Tolle K. The Fouth Paradigm: Data-intensive Scientific Discovery[M]. Redmond, Washington: Microsoft Research, 2009.

[5] Organization for Economic Co-operation and Development, OECD Principles and Guidelines for Access to Research Data from Public Funding. http://www.oecd.org/science/sci-tech/38500813.pdf[R]. 2007.

[6] European Research Council, ERC Scientific Council Guidelines for Open Access[R/OL]. https://erc.europa.eu/sites/default/files/document/file/erc_scc_guidelines_open_access.pdf. 2007.

[7] National Institutes of Health, NIH Data Sharing Policy and Implementation Guidance[Z/OL]. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. 2003.

[8] National Science Foundation, NSF ENG Data Management Plan Requirements[Z/OL]. https://www.nsf.gov/eng/general/dmp.jsp. 2013.

[9] Shi Lei, Yuan Wei. Some Thoughts on the Long-term Mechanism Construction of S&T Resources Collection[J]. China Science & Technology Resources Review, 2012, 44(4): 2-5. [石蕾, 袁伟. 建立科技计划资源汇交长效机制的思考[J]. 中国科技资源导刊, 2012, 44(4): 2-5.]

[10] Michel S, Bernard D. Canadian National Consultation on Ac-

cess to Scientific Research Data[J]. Data Science Journal, 2007,6(17):D26-D35.

[11] Open Access to Scientific Data:Promoting Science and Innovation[J].Data Science Journal,2007,17(6):D21-D25.

[12] Scientific Data Sharing Research Group.The General Framework of Scientific Data Sharing Program in China[J].China Basic Science,2003,(1):65-70.[科学数据共享调研组.科学数据共享工程的总体框架[J].中国基础科学,2003,(1):65-70.]

[13] Sun Jiulin,Shi Huizhong.Construction of National Earth System Science Data Sharing Service Network[J].China Basic Science,2003,(1):78-83.[孙九林,施慧中.中国地球系统科学数据共享服务网的构建[J].中国基础科学,2003,(1):78-83.]

[14] Wang Guofu,Li Jiming,Li Deng,*et al*.China Meteorological Data Sharing Service System:Design and Development[J]. Journal of Applied Meteorological Science,2004,15(z1):10-16.[王国复,李集明,邓莉,等.中国气象科学数据共享服务网总体设计与建设[J].应用气象学报,2004,15(z1):10-16.]

[15] Zhang Xian'en.National Scientific Data Sharing Program[J]. Scientific Chinese,2004,(9):10-13.[张先恩.国家科学数据共享工程[J].科学中国人,2004,(9):10-13.]

[16] Wang Juanle,Sun Jiulin.Review,Reform and Prospect Analysis of World Data Center[J]. Advances in Earth Science, 2009,24(6):612-620.[王卷乐,孙九林.世界数据中心(WDC)回顾、变革与展望[J].地球科学进展,2009,24(6):612-620.]

[17] Zhu Yunqiang,Sun Jiulin,Liao Shunhao,*et al*.Earth System Scientific Data Sharing Research and Practice[J].Journal of Geo-information Science,2010,12(1):1-8.[诸云强,孙九林,廖顺宝,等.地球系统科学数据共享研究与实践[J].地球信息科学学报,2010,12(1):1-8.]

[18] Darby R,Lambert S,Matthews B,*et al*.Enabling Scientific Data Sharing and Re-use[C]//2012(IEEE 8th International Conference on E-Science,2012.

[19] Costello M J.Motivating Online Publication of Data[J].Bioscience,2009,59(5):418-427.

[20] Liu Chuang.Scientific Data Sharing Mechanism in US and Its Implications for China[J].China Basic Science,2003,(1):36-41.[刘闯.美国国有科学数据共享管理机制及对我国的启示[J].中国基础科学,2003,(1):36-41.]

[21] Huang Ruhua,Qiu Chunyan.Review of Research of the Scientific Data Sharing in Foreign Countries[J]. Information and Documentation Services,2013,(4):24-30.[黄如花,邱春艳.国外科学数据共享研究综述[J].情报资料工作,2013,(4):24-30.]

[22] National Aeronautics and Space Administration.Data& Information Policy[R/OL]. http://science. nasa. gov/earth-science/earth-science-data/data-information-policy/2011.

[23] Wang Juanle,Yang Yaping,Zhu Yunqiang,*et al*.Data Archiving Progress and Data Types Analysis of National Basic Research Program of China (973 Program) in Resource and Environment Field[J]. Advances in Earth Science,2009,(8): 947-953.[王卷乐,杨雅萍,诸云强,等."973"计划资源环境领域数据汇交进展与数据分析[J].地球科学进展,2009,(8): 947-953.]

[24] Wu Lizong,Wang Liangxu,Nan Zhuotong,*et al*.Application of DOI in Data Citation:Issues and Suggestions[J].Remote Sensing Technology and Application,2013,28(3):377-382.[吴立宗,王亮绪,南卓铜,等.DOI 在数据引用中的应用:问题与建议[J].遥感技术与应用,2013,28(3):377-382.]

[25] Lu Peng,Miao Liangtian,Mo Jihong,*et al*.Present Situation of Scientific Data Sharing in China[J].Recent Developments in World Seismology,2007,(6):26-32.[路鹏,苗良田,莫纪宏,等.我国科学数据共享现状[J].国际地震动态,2007,(6): 26-32.]

[26] Wu Lizong,Wang Liangxu,Nan Zhuotong *et al*.Scientific Data Publication:An Revierw and Framework[J].Remote Sensing Technology and Application,2013,28(3):383-390.[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.]

[27] He Lin,Chang Yingcong.A Review on Development of Scientific Data Publishing[J]. Library and Information Service, 2014,(5):104-110.[何琳,常颖聪.国内外科学数据出版研究进展[J].图书情报工作,2014,(5):104-110.]

[28] Roberts D,Moritz T.A Framework for Publishing Primary Biodiversity Data Introduction[J].BMC Bionformatics,2011, 1215(I1).

[29] Moritz T,Krishnan S,Roberts D,*et al*.Towards Mainstreaming of Biodiversity Data Publishing:Recommendations of the GBIF Data Publishing Framework Task Group[J]. BMC Bioinformatics,2011,1215(S1).

[30] Ingwersen P,Chavan V.Indicators for the Data Usage Index (DUI):An Incentive for Publishing Primary Biodiversity Data Through Global Information Infrastructure[J].BMC Bioinformatics,2011,1215(S3).

[31] Goddard A,Wilson N,Cryer P,*et al*.Data Hosting Infrastructure for Primary Biodiversity Data[J].BMC Bioinformatics, 2011,1215(S5).

[32] Chavan V,Penev L.The Data Paper:A Mechanism to Incentivize Data Publishing in Biodiversity Science[J]. BMC Bioinformatics,2011,1215(S2).

[33] Chinese Academy of Sciences.《China Scientific Data》is Aproved to be Mational Unified Serial Publication[Z].2015.[中国科学院.《中国科学数据》获批国内统一连续出版物[Z].2015.]

[34] Lyubomir P,Daniel M,Vishwas C,*et al*.Pensoft Data Publishing Policies and Guidelines for Biodiversity Data[Z].Pensoft Publishers,2011.

[35] Liu Chuang.Global Change Research Data Publishing and Repository[J].Acta Geographica Sinica,2014,69(S1):3-11.[刘闯.论全球变化科学研究数据出版[J].地理学报,2014,69(增刊):3-11.]

[36] Wu Lizong,Qu Yonghua,WangLiangxu *et al*.Data Management and Its Sharing Application of Watershed Allied Telemetry Experimental Research[J].Remote Sensing Technology and Application,2010,25(6):772-781.[吴立宗,屈永华,王亮

绪,等.黑河综合遥感联合试验的数据管理与共享[J].遥感技术与应用,2010,25(6):772-781.]

[37] Wu Lizong, Tu Yong, Wang Liangxu, *et al*. Application of Digital Object Identifier in Scientific Data Publication[J].China Science & Technology Resources Review, 2010, (5): 22-29.[吴立宗,涂勇,王亮绪,等.浅谈科学数据出版中的数字对象唯一标识符[J].中国科技资源导刊,2010,(5):22-29.]

[38] Wang Liangxu, Wu Lizong, Li Hongxing, *et al*. Design and Implementation of Data Collection System for Integrated Research on the Eco-hydrological Process of the Heihe River Basin[J].Remote Sensing Technology and Application, 2013, 28(3):362-369.[王亮绪,吴立宗,李红星,等.面向黑河流域生态水文过程集成研究的科学数据汇交与管理[J].遥感技术与应用,2013,28(3):362-369.]

[39] Wu Lizong, Nan Zhuotong, Wang Liangxu.Data Publishing: A New Mode of Data Sharing[J].China Science & Technology Resources Review, 2014, (5): 72-78.[吴立宗,南卓铜,王亮绪.科学数据出版——促进数据共享的一种新模式[J].中国科技资源导刊,2014,(5):72-78.]

[40] Yao Ge, Wang Shuhua, Wang Hengjun.Significance of DOI in the Big Data Era and Its Current Status in Chinese Scientific Journals[J]. Acta Editologica, 2014, (1): 63-65.[姚戈,王淑华,王亨君.大数据时代 DOI 的应用意义与中国科技期刊应用现状[J].编辑学报,2014,(1):63-65.]

[41] Liu Runda, Zhu Yunqiang, Liu Chuang, *et al*. Analysis on the Current Status, Problems &Countermeasures of Scientific Data DOI in China[J].China Science & Technology Resources Review, 2014, (5): 65-71, 78.[刘润达,诸云强,刘闯,等.我国科学数据 DOI 应用现状、问题与对策[J].中国科技资源导刊,2014,(5):65-71,78.]

Collaborative Publishing of Scientific Data: Model and Implementation

Li Hongxing[1], Wu Lizong[1], Nan Zhuotong[2], Wang Liangxu[3], Wang Jian[1], Zhao Yanbo[1], Zhang Yaonan[1], Ma Hanqing[1,4]

(1, *Northwest Institute of Eco-Environment and Resouces, Chinese Academy of Sciences, Lanzhou* 730000, *China*;
2.*School of Geographic Science, Nanjing Normal University, Nanjing* 210023, *China*;
3.*Institute of Urban Study, Shanghai Normal University, Shanghai* 200234, *China*;
4.*Lanzhou Library, Chinese Academy of Sciences, Lanzhou* 730000, *China*)

Abstract: Scientific data publishing is a new and effective way to promote data sharing. In this paper, a novel collaborative publishing model for scientific data was proposed by incorporating data center and academic journal. In this model, data are published on a traditional academic journal in a form of data paper which has been rigorously peer reviewed, and preserved in a selected data center which is committed to data stewardship, management and sharing. This model takes advantage of the mature framework of peer review in the academic journal system to ensure the quality of published data. The form of data paper makes it more acceptable by the existing academic evaluation system in China, which usually neglects any contributions in form of data. As a result, it helps foster positive atmosphere in data authors of sharing data. The model can therefore effectively address two key issues mostly concerned: controlling of data quality and protection of author rights. The proposed model is easy to operate and can also be learned by disciplines other than geoscience we showcased in this study.

Key words: Scientific data; Data paper; Data center; Academic journal; Collaborative publishing