

引用格式:Wu Lizong, Wang Liangxu, Nan Zhuotong, *et al.* Scientific Data Publication: An Review 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.]

科学数据出版现状及其体系框架

吴立宗, 王亮绪, 南卓铜, 李红星

(中国科学院寒区旱区环境与工程研究所, 甘肃 兰州 730000)

摘要:科学数据出版是深化数据共享的重要手段,能够激励数据生产者发布和共享数据,又能保护数据的知识产权。回顾了国内外科学数据出版现状,从数据中心和传统出版系统两个角度,提出和完善了科学数据出版体系架构。提出数据出版涉及数据提交、同行审议、数据发布和永久存储、数据引用和影响评价 5 个基本环节,应由出版系统、数据中心和第三方数据评价机构协同实施。重点讨论了科学数据出版面临的关键问题:数据授权、数字资源唯一标识符、数据集成和数据稿源问题。倡议建立我国的科学数据出版体系,科学数据中心在开展数据收集和共享的同时,加强与学术期刊的合作,成为数据出版的永久存储机构;学术期刊积极开展机制创新,推动学术论文相关数据的公开发布;积极开展数据知识产权保护,倡导对科学数据的引用;加强科学数据的影响评价研究,让数据成果成为普遍认同的科研产出成果,激励科学家主动参与科学数据共享。

关键词:数据出版;数据共享;知识产权;数字对象唯一标识符

中图分类号:TP 75 **文献标志码:**A **文章编号:**1004-0323(2013)03-0383-08

1 引 言

随着信息技术的发展,科学数据迅速积累,人类社会近 50 a 的数据积累量已经超过了过去 5 000 a 的数据积累总量。数据在科学研究中的作用越来越重要,数据发现已经成为继试验、理论、计算模拟之后的第四种科学研究范式^[1]。科学数据分散在不同的国家、研究机构、科研项目甚至科学家手中,数据共享是促进科学数据有效利用的重要机制和手段。世界上很多国家都通过政策引导和国家投入建立了国家级数据中心体系^[2],用于数据的长期保存和数据共享。很多国际学术组织(如 GEOSS、IGBP、IPCC、IPY、IODP)也建立了自己的科学数据库^[3-4],致力于关键数据集的制备与数据共享。中国早在 1988 年就加入世界数据中心(WDC)体系^[5],开始探索科学数据共享。2002 年起,中国科技部从国家层面上组织实施了“国家科学数据共享工程”,初步建

立了我国国家级数据中心体系,积累和共享了大量科学数据,也取得了明显的社会效益^[6]。

科学数据共享经历了多个发展阶段,每一个阶段都有特定的服务目标,如早期的数据共享强调协调和交流,要优先保证研究项目从数据单位获得数据。科技部实施的“国家科学数据共享工程”在第一阶段强调数据的收集和整合,宏观掌握我国历史数据的整体情况,为下一阶段的数据共享奠定基础。我国虽已初步建成了国家级的数据共享系统,但在无国家经费支持和政策约束的情况下,很多数据中心将面临无数据可存、无数据可共享的尴尬局面^[7],急需一种机制保证数据共享能够长期可持续发展。在密集型数据和大数据时代,数据的管理、共享、集成和挖掘不是依靠数据中心的有限力量就能完成的,必须广泛接纳科学家参与数据管理和发现,让科学数据共享、管理和发现真正成为科学研究的一部分。

收稿日期:2012-10-22;修订日期:2013-05-13

基金项目:国家自然科学基金项目(91025001)、国家科技基础条件平台“地球系统科学数据共享平台”资助。

作者简介:吴立宗(1975—),男,山东即墨人,博士,副研究员,主要从事数据共享和冰川变化研究。E-mail: wulizong@izb. ac. cn。

数据出版是近两年出版界和数据共享界共同提出和积极探索的新概念,它有望从机制上改变目前数据共享所面临的根本问题,改变数据共享和数据中心的格局。本文对国内外科学数据出版现状进行了回顾,从数据中心和数据出版两种角度总结和凝练科学数据出版的体系框架,进一步讨论科学数据出版对数据共享的深远影响。

2 国内外科学数据出版现状

2.1 国际科学数据出版现状

2.1.1 学术出版界

一些编目性质的科学数据库很难通过有限项目的集中调查就能完成,需要全体科学家的连续参与。发表在论文中的数据是这类数据库的重要数据来源,很多期刊配合数据库建设,提倡作者在发表论文时,优先将数据注册到相关数据库中。一些著名的数据库,如 GenBank 数据库和全球生物多样性数据库(Global Biodiversity Information Facility, GBIF)就是这样建立的,数据公开已经成为一些科学家的自发行为。但这类数据库的维护也存在很多问题,在缺乏约束的情况下,很多科学家并不对记录进行更新^[8]。因此,也需要建立一种机制鼓励作者主动更新数据。

数据是论文的重要组成部分,决定了论文的质量。为给论文评审提供更多依据,同时促进数据再次利用,很多期刊要求作者在文章正式发表前将相关数据公开,如 Evolution、Molecular Biology and Evolution、Nature、PloS Biology、Proceedings of the National Academy of Sciences USA 和 Science 等著名期刊,这是数据出版的雏形。在实际操作过程中,数据公开程度主要依靠期刊编辑们的提醒和约束,一些作者虽然承诺数据公开,但文章发表后,往往采取种种借口不进行实际的数据公开^[9]。

为了建立类似学术论文的数据出版体系,全球生物多样性机构(GBIF)设立了数据出版框架工作组(Data Publishing Frame Task Group),发表了系列文章讨论了科学数据出版框架^[10-15],并在 Pensoft 出版集团的系列出版物中进行试点^[16]。该数据出版体系的核心内容包括:①学术论文涉及的数据必须在文章正式发表前提交到出版集团的数据库中,这些数据将成为同行审议的重要依据;②为了减少文章篇幅,与文章相关的辅助数据可以以数据论文(Data paper)的形式先行发表,学术论文通过引用建立文章与数据的关系。数据论文(Data paper)的

形式较为灵活,但需要经过同行审议;③为了便于数据的二次利用,数据发布前,编辑部一般要求数据作者签署数据授权协议,明确声明放弃数据的再编译权和发布权;④出版商为数据和 Data paper 创建全球唯一永久标识符(DOI),保证网络环境下的数据地址永久有效。

多数期刊在数据管理、共享和长期存储方面并不具有优势,一些期刊因此联合起来成立专门的数据中心(Data repository),建立期刊与数据中心联合互动的数据管理机制。如 Dryad(<http://www.datadryad.org/>)就是一个由生命和生物领域国际期刊联合发起成立的数据中心,负责各个期刊出版数据的统一注册、管理与共享,已经有 20 多个期刊实现了与 Dryad 的合作。

建立针对科学数据的评价指标是改变学术成果评价体系的一个重要举措,是推动科学家主动参与数据共享和出版的重要动力。GBIF 数据出版工作组提出数据利用指标(Data Usage Index, DUI)^[13,17],希望通过数据访问量、下载量和引用次数等要素建立针对数据的评价指标。著名的文献供应商汤森路透(Thomson Reuters)旗下的知识产权与科技事业部(Web of Knowledge)于 2012 年 10 月推出了数据引文指标(Data Citation Index, DCI),希望通过文献与数据之间的相互引用建立针对数据的评价指标,进而推动数据发现和知识互联^[18]。

2.1.2 科学数据共享界

前世界数据中心(World Data Center)是隶属国际科学联合会(ICSU)的国际性数据共享组织,自 1957 年成立以来已经在全球建立了 51 个学科中心。WDC 在成立初期为国际科学计划的执行起到了重要的数据支撑作用,但在后期遇到了非常严重的可持续发展问题,一些数据中心由于缺乏稳定的数据来源,数据中心越来越不活跃。国际科学联合会在提出对 WDC 进行机构改革的同时^[5],也提出了数据出版的概念,希望改变数据中心的定位,将数据中心作为数据出版的重要组成部分^[19];科学家在发表文章后,相关数据应保存到数据中心,而不是作为废料抛弃。德国是数据共享界开展数据出版探索较早的国家,在国际科学联合会的强烈建议下,国际科学数据委员会(CODATA)和德国国家研究基金(German Research Foundation)联合启动了“科学数据出版与引用”项目,由德国国家科学技术图书馆(TIB)与德国的数据中心合作积极探索科学数据的出版模式。德国国家科学技术图书馆(TIB)申请成

为国际上首个科学数据 DOI 代理机构,积极探索科学数据的引用格式,为原 WDC 数据中心的很多数据资源进行了 DOI 注册^[19-20]。在 TIB 工作的基础上,DataCite 于 2009 年在伦敦成立,为数据中心和科学家提供数据出版的标准流程,为期刊出版商提供数据引用格式和有效链接地址,也为科学家提供数据发现的统一门户^[21]。

与出版商面对的科学数据不同,数据中心管理和共享的大部分数据都没有经过期刊“正式出版”,数据引用规范不标准的问题更加突出。对数据的引用包括对数据说明文档的引用、数据相关论文的引用和对数据本身的引用 3 种基本形式^[22],数据中心、出版商和科学家对如何确定数据的引用格式还存在一定的争议,如出版商推荐通过 Data paper 的形式,将数据和文章的引用格式统一化,数据中心更加愿意突出对数据本身的引用,而部分科学家更倾向于对数据相关文章的引用。数据施引情况也存在很多问题,只有少数文章愿意将数据的引用纳入参考文献(Reference)部分,正文中的数据说明方式差异也很大,多数文章不提供数据的唯一标识符或数据存储地址^[23]。

2.2 国内科学数据出版现状

为了体现数据中心在数据共享中的贡献,国内一些数据中心较早时期就要求用户在使用数据时注明数据的来源,如 WDC 冰川(雪冰)冻土学科中心、中国西部环境与生态科学数据中心、中国科学院计算机网络信息中心国际数据镜像网站。中国西部环境与生态科学数据中心(<http://westdc.westgis.ac.cn>)因对使用声明要求较为严格,可利用数据使用声明中的关键词(如“中国西部环境与生态科学数据中心”,“中国西部数据中心”或“<http://westdc.westgis.ac.cn>”),检索到使用该数据中心共享数据发表的文章。截至 2012 年底,中国西部数据中心共检索到明确声明利用该数据中心共享数据发表的论文 455 篇,其中硕/博士论文 89 篇,英文文章 92 篇,中文文章 275 篇^[24]。其他数据中心也试图第三方引用标注分析数据库的影响^[25],但由于关键词代表性差,无法获得理想的检索结果。从数据服务成效角度,数据中心需要一种机制对共享数据的使用情况进行跟踪,以准确评价数据共享效果。

为了保护数据知识产权,国内一些数据中心希望通过传统数据引用的方式体现数据的知识产权,并为此讨论了不同情况下的数据引用方式^[26-27]。如基础科学数据共享网基于唯一标识符概念建立了一

套科学数据引用规范^[28],并在全网推广。DOI 系统引入中国后,中国科技信息研究所借鉴德国国家科学技术图书馆(TIB)的经验,将数据 DOI 注册和数据引用介绍到中国^[29],并给出基于 DOI 的科学数据引用规范建议^[27]。“中国西部环境与生态科学数据中心”从数据跟踪和知识产权保护的角度,与中国科技信息研究所合作以“黑河综合遥感联合试验”为例试点了基于 DOI 的数据引用^[4,26],但只有少数文章在参考文献部分对数据进行了正式引用,很多期刊尚拒绝对数据进行引用。

国内数据共享起步较晚,数据中心在相当长时间内关注历史数据的收集和整理,科学家个人直接发布的数据较少,因此没有形成完整的数据出版框架。“中国西部环境与生态科学数据中心”先后服务于“中国西部环境与生态科学计划”^[3]、“黑河综合遥感联合试验”^[4]和“黑河流域生态—水文过程集成研究”,负责这些研究计划产出成果的汇交和共享。为了鼓励更多的科学家主动地汇交数据,数据中心通过数据引用强调数据的知识产权,通过数据跟踪统计,让数据提供者了解数据共享成效,为此形成了一套数据发布—评审—建议引用—跟踪统计的数据共享体系,初步具备了数据出版的特征:

(1)各研究项目需要通过“中国西部环境与生态科学数据中心”进行数据汇交,实现元数据撰写和数据提交。科学家个人也可以通过该系统发布数据。

(2)元数据撰写完成后,数据中心将元数据发送给相关专家,由他们对元数据的质量进行把关,如果审核合格则发布元数据,否则将反馈给元数据作者进行修改。

(3)在元数据中明确数据的建议引用方式,要求用户在使用数据时必须对数据进行引用。根据数据提供者的意愿,数据引用方式可以对数据本身进行引用,也可以对与数据相关的文章进行引用。

(4)“中国西部环境与生态科学数据中心”保留数据的申请信息,并利用 DOI 和关键词对使用本数据中心共享数据发表的文章进行检索和统计,将数据共享成效信息(包括下载次数、申请次数、使用该数据发表的文章)反馈给数据提供者或生产者。

3 科学数据出版的体系架构

3.1 科学数据出版的内涵

科学数据出版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审议和公开公布,创建标准和永久的数据引用信息,供其

他研究性文章引证。经过同行审议,科学数据出版可以避免数据造假;通过标准化的引用信息,可以对数据的引用情况进行准确的跟踪,从而建立类似学术出版物形式的评价体系,体现数据生产者的科研创造,激励更多的科学家参与数据共享。

3.2 科学数据出版主要环节

数据出版并不是简单的数据发布,而是包括数据提交、同行审议、数据发布和永久存储、数据引用和影响评价 5 个基本环节。根据国内外的数据出版经验,总结和提出的数据出版框架体系见图 1。

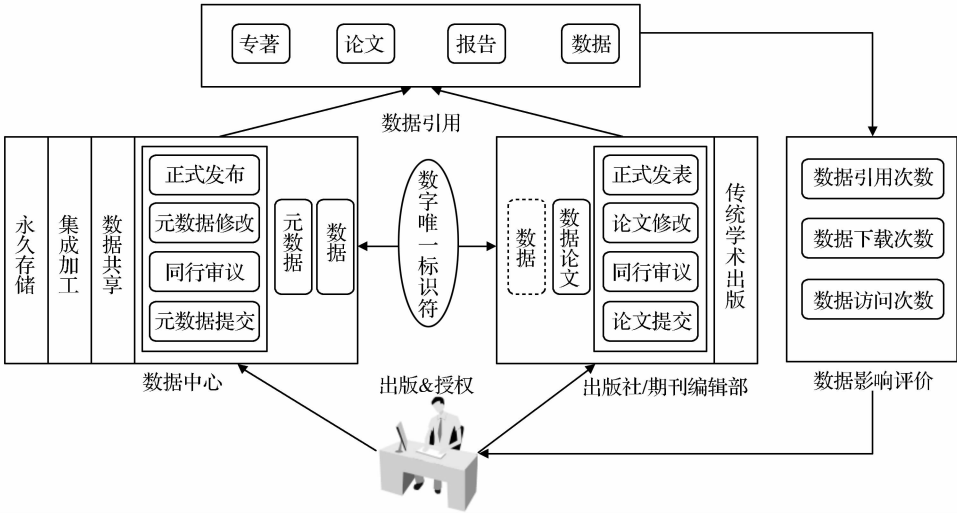


图 1 科学数据出版的体系框架

Fig. 1 Scientific data publication: An review and framework

3.2.1 数据发布

数据发布是指数据拥有者将数据公布在其他用户可访问的公共数据平台的过程,这种公共平台要求有永久的数据保存能力。数据发布可以通过出版商和科学数据中心等两种平台进行,与文章密切相关的数据可优先通过出版商发布,其他数据可通过科学数据中心发布。

3.2.2 同行审议

数据出版的重要目标之一是避免数据篡改和学术造假,但数据质量评估是一个非常复杂的问题^[29],有时并不能通过专家的简单浏览查看就能确认其质量,还需要大量的应用和检验。因此,对于数据出版而言,数据审核主要包括:①数据发布前的同行审议,重点从数据完整性角度审核与数据相关的各种信息是否完整,是否能够让用户正确使用数据;②数据使用后的用户反馈。要有效收集数据的使用信息以及用户的评价信息,为其他用户使用该数据提供更多的信息。

3.2.3 数据出版和永久存储

对科学数据而言,数据出版是指数据达到可引用和追溯的状态,核心内容是为数据引用提供标准的数据引用格式和永久访问地址。出版界倡导的 Data paper 实质是将数据出版利用传统论文形式将

数据的引用格式和访问地址固化起来。

网络环境下的数据资源具有地址易变形和多重连接性,为了便于用户访问和追溯,数据的访问地址须具备永久不变性。数据的永久存储并不代表物理地址永久不变,可通过再解析方法使数据的逻辑地址不变。很多数据库都是由项目资助建立的,项目结束后,数据库得不到有效维护和更新,不具备数据永久存储的能力。因此,数据出版需要建立能够长期稳定运行的数据中心。

3.2.4 数据引用

理想的数据引用是指与传统文献引用方式一样,在文章参考文献部分对数据进行引用。但当前阶段,文章正文中的数据标注也应该算是一种数据引用方式。数据引用是数据出版系统中的重要环节,需要数据中心、评审专家和出版商密切合作才能有效推广:出版商首先要允许和鼓励对科学数据的引用,数据中心应提供规范化的数据引用格式,评审专家在审稿过程中应严格要求作者对相关数据进行引用和说明。

3.2.5 数据评价

类似期刊评价系统,数据评价包括对数据中心的评价和数据本身的评价,前者可用来评价数据中心的共享服务成效,后者可用来评价数据的影响程

度。对数据的评价有两种目的:①借助数据影响评价来判断数据的质量,为用户选择数据和使用数据提供帮助;②将数据作为一种重要的评价指标纳入到科研成效评价指标,从根本上激励研究者主动参与科学数据出版和数据共享。

3.3 数据出版涉及的机构

数据出版涉及的单位或组织包括出版商/编辑部、数据中心和评价机构,其作用如下:

3.3.1 出版商和编辑部

目前出版商开展数据出版的形式主要有两种:其一是创建专门的数据期刊(Data Journal),专门刊登与数据有关的文章。如国际科技数据委员会(CODATA)专门创建的 Data Science Journal 和约翰威立国际出版公司(John Wiley & Sons Inc.)联合英国皇家气象学会(Royal Meteorological Society)创建的 Geoscience Data Journal(GDJ)。其二是允许和鼓励作者在传统期刊上发表数据文章(Data paper),如 Pensoft 出版集团和 Dryad 加盟期刊。无论何种形式,出版商最初构想都是期望通过“文章(paper)”的形式来表现数据,利用“传统”方式将数据出版纳入到学术出版体系中。

利用论文出版的框架体系,出版商和编辑部可直接开展数据的注册、同行审议和出版。通过评审专家和期刊编辑的引导或强制要求,既可督促更多的作者公开数据,也可更有效地规范作者在参考文献部分对数据进行引用。

出版商和编辑部参与数据出版也有很多不足:首先,大量的科学数据已经由数据中心管理和发布,以 Data paper 形式的再次数据出版将造成数据的多重链接和浪费。其次,期刊和出版商在数据管理方面能力不足,无法满足后续数据集成和二次加工的需要。出版商和编辑部可以与数据中心加强合作采用联合数据保存的政策来开展数据出版^[30],数据中心负责数据的注册、永久存储和二次加工,由数据中心提供永久唯一标识和固定链接地址,期刊将数据通过 Data paper 的形式发表,提供规范化的引用格式,并由期刊编辑和审稿专家规范作者的数据引用行为。

3.3.2 数据中心

数据中心在数据资源和数据管理方面具有很大优势,元数据也包含了丰富的数据说明信息,拥有数据出版的基本条件。针对数据出版,数据中心存在的不足主要表现为:①很多数据缺乏唯一永久地址,数据存在多重链接和地址易变性,不符合数据追溯

的要求;②多数数据中心并不提供规范化的数据引用格式,数据无法得到正确引用;③因数据管理方式差异,数据粒度不符合数据引用的要求;④缺乏数据评审机制,数据质量参差不齐;⑤缺乏数据投稿机制,科学家个人找不到适合发布数据的网站。

为了满足数据出版的需要,数据中心首先要与唯一标识符系统合作,为数据中心的每一个数据提供唯一永久标识;其次,数据中心要厘清数据版权信息,为每一个数据提供规范化的数据引用格式。再次,数据中心要建立数据发布和评审制度,允许和鼓励科学家个人通过数据中心发布和共享数据,邀请同行专家对(元)数据进行评审,保证数据的质量。

3.3.3 评价机构

评价数据影响程度的因子很多,如浏览次数、下载次数和引用次数等,但文章对数据的引用次数是评价数据影响程度的重要指标,可以将数据评价与传统文章评价结合起来。由于数据引用还很不规范,只有少数文章将数据引用纳入到参考文献部分,因对参考文献进行分析不能获得准确的数据引用信息。当前,要开展全面公正的数据评价,数据评价机构需要与数据中心全面合作,详细收集数据引用信息。从长远看,数据评价需要期刊编辑和审稿专家对作者的数据引用行为进行规范,数据中心也应尽可能地为每一个数据建立唯一标识符,实现更加准确的数据跟踪。

3.4 数据出版的其他关键问题

3.4.1 数据授权

数据知识产权保护和数据共享是一对矛盾^[31],过分强调数据知识产权保护将妨碍数据共享,但忽视数据知识产权保护将打击数据生产者的积极性,也不利于数据共享。我国为了推动数据共享,在相当长时间内忽略了数据知识产权的保护^[32],这将会在一定程度上损害数据生产者的权益,阻碍科学数据共享。

与数据相关的核心知识产权包括署名权、发布权/再分发权和再编译权。署名权是体现数据作者贡献的基本方式,而数据的分发权和再编译权一定程度上会妨碍数据共享和二次加工,因此国际上一般建议数据作者放弃这两项权利。数据授权是数据生产者在将数据发布到数据中心(或出版社)之前所作的数据版权声明和分割。在数据出版和数据共享领域,还没有提出针对数据的使用协议,可参照知识共享领域的 Open Data Commons 协议来执行。如保留署名权可采用 ODC-By 协议(Open Data Com-

mons Attribution License),如放弃所有权利,到要求使用该数据的其他数据库和成果也必须放弃相关权利时可采用 PDDL(Open Data Commons Public Domain Dedication and Licence)协议。如放弃包括署名权在内的所有权益可采用 CC0(Creative Commons CC-Zero)协议。

3.4.2 数据唯一标识

当前的数据共享和数据出版大多是在网络环境下开展的,数据资源具有数字对象的一般特点,即虚拟性、易移动性、易复制性、易变化性和易组合性^[33],因此如何永久和唯一地对数据进行标识,是科学数据发布后,开展数据查找和影响评价的关键所在。

DOI是唯一标识符的一种实现,已在出版界得到广泛应用。利用DOI进行数据标识可以很容易地建立科学数据与学术文章的联系,有利于开展数据评价。在数据出版领域也广泛应用。如Datacite推荐利用DOI建立数据唯一标识符,中国科技信息研究所正在国内积极倡导利用DOI进行科学数据唯一标识。

DOI系统也是一套再解析系统,数据出版采用DOI唯一编码后,DOI注册中心就天然成为数据检索中心,用户可通过全球DOI解析系统查找数据的引用格式。DOI这一特点是其他标识系统或局部解析系统所不具备的。

3.4.3 数据集成和二次加工

在信息爆炸的时代,数据量越来越多。在大数据时代,数据的价值不仅表现为单个数据的价值,还表现为大样本数据的统计意义和相关关系,数据的集成和二次加工将是继数据共享和出版之后的重要关注点。要迎接大数据时代的挑战,数据中心首先要对数据管理方法进行创新,提供人和机器都能够读懂和自主获取的数据。

数据集成和二次加工必然要对不同来源和版权的数据进行处理,需要涉及处理后数据和原始数据的知识产权关系。目前国际上尚没有得到普遍认同的处理方式,亟待深入研究。

3.4.4 稿源问题

数据出版是数据共享的更高形式,但如果没有科学家主动参与数据投稿,数据出版依然是纸上谈兵。只有将数据成果纳入到科研成果评价体系中,科学家才会主动地参与数据共享和数据出版。

当前,解决数据出版的稿源问题可从3方面着手:①期刊制定严格的数据公开标准,要求文章投稿

前必须将相关数据发布在权威数据中心,这些数据将成为文章质量审核的重要依据。期刊编辑和审稿专家在审稿过程中要对数据进行严格审核,推荐作者将数据正式出版,同时引导作者在参考文献部分对数据进行规范化引用;②数据中心根据数据出版的要求完善数据管理机制,提供规范化的数据引用信息和永久链接地址。也可为期刊合作,推荐相关数据以Data paper的形式发表;③学术界接受Data paper作为论文的一种新形式,接受评价机构对Data paper的评价,鼓励研究人员发表Data paper。

4 结 语

数据出版是一种全新的数据共享机制,既可以解决数据知识产权保护问题,又可以推动数据的共享和再次利用。数据出版与传统学术出版的结合,有望解决数据成果的评价问题,从根本上推动科学家参与数据共享的积极性。

(1)数据出版是一个新兴事物,还存在各种各样的问题。数据出版的推广需要出版商/编辑部、数据中心、评价机构的共同努力。出版商/期刊应制定严格的数据政策,要求与文章相关的数据必须公开。期刊编辑和审稿专家要对文章相关数据进行严格审核,引导研究者将数据正式出版,并在参考文献部分对数据进行规范化引用。数据中心要协助出版商/期刊进行数据注册和长期保存,同时要完善数据管理机制,为数据资源提供规范化的数据引用信息和永久链接地址。评价结构要结合传统论文评价方法提出针对科学数据的、可被广泛接受的评价指标和方法,学术界要承认数据出版,将其纳入科研成果评价体系中。

(2)科学数据是一种重要的战略资源,而数据出版又符合大数据研究的特征,因此数据出版将是国际数据资源争夺的重要阵地。我国有必要建立自己的数据出版体系,创立数据期刊,鼓励期刊发表数据论文,建立权威科学数据中心,让科学家将数据发布到我国的数据中心,为未来的大数据研究和大科学研究奠定基础。

(3)数据出版是一种新的数据共享机制,可实现我国数据共享跨越式发展。国家应从数据出版的角度,调整数据中心的定位和数据管理机制,更加重视针对数据作者的服务,协调数据中心与期刊/出版社的合作。要重新认识知识产权与数据共享之间的关系,通过数据出版充分体现数据的署名权,通过授权解决数据集成和再次利用的问题。

致谢:感谢中国科技信息研究所彭洁主任、涂勇博士,中国科学院动物研究所黄晓磊博士,中国科学院地理科学与资源研究所诸云强博士和中国极地研究中心张洁博士为本文提供的大量信息和宝贵意见!

参考文献(References):

- [1] Hey Tony, Tansley Stewart, Tolle Kristin. The Fourth Paradigm: Data-Intensive Scientific Discovery [M]. Redmond, Washington: Microsoft Research, 2009.
- [2] Xu Guanhua. Implement Scientific Data Sharing for Strengthening the National Scitech Competitive Power[J]. China Basic Science, 2003, (1): 7-11. [徐冠华. 实施科学数据共享增强国家科技竞争力[J]. 中国基础科学, 2003, (1): 7-11.]
- [3] Li Xin, Nan Zhuotong, Wu Lizong, *et al.* Environmental and Ecological Science Data Center for West China: Integration and Sharing of Environmental and Ecological Data[J]. Advances in Earth Science, 2008, 23(6): 628-637. [李新, 南卓铜, 吴立宗, 等. 中国西部环境与生态科学数据中心: 面向西部环境与生态科学的数据集成与共享[J]. 地球科学研究进展, 2008, 23(6): 628-637.]
- [4] Wu Lizong, Qu Yonghua, Wang Liangxu, *et al.* Data Management and Its Sharing Application of Watershed Allied Telemetry Experimental Research[J]. Remote Sensing Technology and Application, 2010, 25(2): 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(2): 772-781.]
- [5] Wang Juanle, Sun Jiulin. Review, Reform and Prospect Analysis of World Data Center[J]. Advances in Earth Sciences, 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.]
- [6] Zhang Xian'en. National Science Data Sharing Project[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.]
- [7] Nelson B. Data Sharing: Empty Archives[J]. Nature, 2009, 461(7261): 160-163. doi: 10. 1038/461160a.
- [8] Noor M A F, Zimmerman K J, Teeter K C. Data Sharing: How Much Doesn't Get Submitted to GenBank? [J]. PLoS Biology, 2006, 4(7): e228. doi: 10. 1371/journal. pbio. 0040228.
- [9] Savage Caroline J, Vickers Andrew J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals[J]. PLoS one, 2009, 4(9): e7078.
- [10] Roberts D, Moritz T. A Framework for Publishing Primary Biodiversity Data[J]. BMC Bioinformatics, 2011, 12(Sup. 15): 11.
- [11] Moritz T, Krishnan S, Roberts D, *et al.* Towards Mainstreaming of Biodiversity Data Publishing Recommendation of the GBIF Data Publishing Framework Task Group[J]. BMC Bioinformatics, 2011, 12(Sup. 15): S1. doi: 10. 1186/1471-2105-12-S15-S1.
- [12] Chavan V, Penev L. Data Paper: Mechanism to Incentivise Discovery of Biodiversity Data Resources[J]. BMC Bioinformatics, 2011, 12(Sup. 15): S2. doi: 10. 1186/1471-2015-12-S15-S2.
- [13] Ingwersen Peter, Chavan Vishwas. Indicators for the Data Usage Index (DUI): An Incentive for Publishing Primary Biodiversity Data Through Global Information Infrastructure[J]. BMC Bioinformatics, 2011, 12: 1.
- [14] Ariño A H, Chavan V, King N. Biodiversity Informatics Potential Index (BIPIndex)[J]. BMC Bioinformatics, 2011, 12(Sup. 15): S4.
- [15] Goddard A, Wilson N, Cryer P, *et al.* Data Hosting Infrastructure for Primary Biodiversity Data[J]. BMC Bioinformatics, 2011, 12(Sup. 15): S5.
- [16] Penev Lyubomir, Mietchen Daniel, Chavan Vishwas, *et al.* Pensoft Data Publishing Policies and Guidelines for Biodiversity Data[R]. Pensoft Publishers, 2011.
- [17] Chavan V S, Ingwersen P. Towards a Data Publishing Framework for Primary Biodiversity Data: Challenges and Potentials for the Biodiversity Informatics Community[J]. BMC Bioinformatics, 2009, 10(Sup. 14): S2.
- [18] Thomson Reuters. Data Citation Index [EB/OL]. 2012, http://wokinfo.com/products_tools/multidisciplinary/dci/.
- [19] Klump Jens, Bertelmann Roland, Brase Jan, *et al.* Data Publication In the Open Access Initiative[J]. Data Science Journal, 2006, 5(15): 79-83.
- [20] Peng Jie, Tu Yong. The Application of DOI in Scientific Data [J]. Qianyan Tansuo, 2007, (2): 27-28. [彭洁, 涂勇. DOI 在科学数据领域的应用研究[J]. 前沿·探索, 2007(2): 27-28.]
- [21] German National Library of Science and Technology. DataCite [EB/OL]. 2010, <http://datacite.org/>.
- [22] CODATA Task Group on Digital Data Citation. CODATA Task Group on Digital Data Citation: Best Practices Research and Analysis Results[R]. CODATA, 2011.
- [23] Hou Jingchuan, Fang Jingyi. Review on Data Citation in the Context of Big Data[J]. Journal of Library Science in China, 2013, (1): 112-118. [侯经川, 方静怡. 数据引证研究: 进展与展望[J]. 中国图书馆学报, 2013(1): 112-118.]
- [24] Li Hongxing, Wang Liangxu, Wu Lizong, *et al.* The Philosophy and Achievements about Data Services of Environmental and Ecological Science Data Center for West China[J]. Remote Sensing Technology and Application, 2013, 28(2): 370-376. [李红星, 王亮绪, 吴立宗, 等. 中国西部环境与生态科学数据中心的数据服务理念与成效[J]. 遥感技术与应用, 2013, 28(2): 370-376.]
- [25] Zhu Yanhua, Hu Lianglin. Analyzing the Influence of Scientific Database based on the Third-Party Cited Marks[J]. China Science & Technology Resources Review, 2012, 44(6): 17-22. [朱艳华, 胡良霖. 基于第三方引用标注的科学数据库影响力分析[J]. 中国科技资源导刊, 2012, 44(6): 17-22.]
- [26] Wu Lizong, Tu Yong, Wang Liangxu, *et al.* Application of Digital Object Identifier in Scientific Data Publication[J]. Chi-