

浅谈科学数据出版中的数字对象唯一标识符

吴立宗¹ 涂勇² 王亮绪¹ 南卓铜¹ 李新¹ 彭洁² 潘小多¹ 李红星¹

(1. 中国科学院寒区旱区环境与工程研究所, 甘肃兰州 730000;

2. 中国科学技术信息研究所, 北京 100038)

摘要: 数字对象唯一标识符具有唯一性、永久性和多重链接等特点, 是实现互联网环境下数字对象永久性定位的数据系统。数字对象唯一标识符已经在出版领域得到了广泛的应用, 在数据出版和引用方面也有很好的应用前景。文章介绍了中国西部环境与生态科学数据中心在数据引用方面所做的工作, 详细介绍了如何在中文 DOI 注册中心的帮助下, 利用数字对象唯一标识符对“黑河综合遥感联合试验”产出数据所开展的数据出版和数据引用尝试。文章最后总结了 DOI 在数据出版领域的意义, 并讨论了 DOI 系统对数据出版和引用方面存在的不足, 指出了数据中心在面向数据出版和引用方面存在的挑战和机遇。

关键词: 数字对象唯一标识符; DOI; 科学数据出版; 数据引用; 数据共享; 中国西部环境与生态科学数据中心

中图分类号: G

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2010.05.005

Application of Digital Object Identifier in Scientific Data Publication

Wu Lizong¹, Tu Yong², Wang Liangxu¹, Nan Zhuotong¹, Li Xin¹, Peng Jie², Pan Xiaoduo¹, Li Hongxing¹

(1. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000;

2. Institute of Scientific & Technical Information of China, Beijing 100038)

Abstract: The Digital Object Identifier (DOI) is a system for identifying content objects in the digital environment. DOIs are names assigned to any entity for use on Internet digital networks. Scientific data sets may be identified by DOIs, and several efforts are now underway in this area. This paper outlines the protect actions on data intellectual property rights and introduces the efforts on scientific data citation based on DOIs by Environmental and Ecological Science Data Center for West China. At last, this paper addresses the challenge for Data Centers and DOI System on the application of DOI in publication scientific data.

Keyword: Digital Object Identifier, DOI, scientific data publication, data citation, data sharing, Environmental & Ecological Science Data Center for West China

1 引言

作为科学数据, 有多少数据被正式出版物

所使用是检查数据共享成果的重要指标, 它既体现了数据中心的成果, 也在一定程度上体现了数据的质量和用户的认同度。与数据共享量相比, 论文标注数量并不足以说明实际的数据应用

第一作者简介: 吴立宗(1975-), 男, 副研究员, 主要研究方向: 数据共享和地理信息系统。

基金项目: 中国科学院西部行动计划(二期)项目“黑河流域遥感—地面观测同步试验与综合模拟平台建设”(KZCX2-XB2-09); “中国西部环境与生态科学研究计划”重点项目“中国西部环境与生态科学数据中心”(90502010)。

收稿日期: 2010年9月14日。

效果，因为有相当一部分成果没有标注数据的来源。此外，一般的成果仅标注数据的来源，除了认真阅读文献外，很难通过常规的搜索方法获得哪些数据被使用，这为收集和分析具体数据在使用过程中所存在的问题增加了相当的难度。此外，数据的共享除了要面向用户外，还要对数据贡献者负责，要尽可能地提供数据的引用信息，提高数据共享的积极性。

中国西部环境与生态科学数据中心（简称“西部数据中心”）在促进数据共享的同时，一贯强调对数据知识产权的保护，除了要求用户在使用数据时对数据来源进行声明外，还探索通过参考文献的方式直接对数据进行引用，并与中文 DOI 注册中心合作，共同探讨如何利用数字对象唯一标识符（DOI）促进科学数据的出版和引用。本文将在简要回顾数字对象唯一标识的特征和功能的基础上，介绍西部数据中心将 DOI 技术应用用于“黑河综合遥感联合试验”中的实例，然后探讨产权模糊的其他数据资源如何应用 DOI 技术，最后深入讨论 DOI 系统和数据中心在数据出版和引用中面临的挑战和机遇。

2 DOI 与科学数据的出版与引用

对科学数据进行发布和引用并非新生事物，在化学、生物和晶体结构领域里，很多期刊都要求在论文正式发表前必须将有关的数据公开发表。如在生物领域里，很多期刊都要求作者在文章正式发表前，将基因序列数据注册到 GenBank 数据库中，同时要求文章在引用相关数据时，提供该数据在 GenBank 数据库中的登录号^[1-2]。

数据共享领域也十分重视数据知识产权的保护，很多数据中心都采取了声明数据来源^[3]、引用数据的在线 Web 地址^[3-6]、引用数据的说明文档^[7]等多种方式来强调数据的知识产权，但越来越多的人倾向于直接对数据本身进行引用^[4-5,8]。有据可查是科技文献的一个基本特征，能够进行引用的文献应该具有可追溯性。对科学数据本身的引用不仅表现在数据的格式或表现形式易变性上，在互联网环境下的科学数据存储地址也经常发生变更，从而很难通过文献中提供的标识信息

追溯数据被引用时的原始状态，因此需要一种技术来对数据的地址进行永久性的保存，只有这样对数据本身的引用才具有参考价值。数字对象唯一标识符的出现为数字对象的永久保存和唯一标识提供了技术支持，可以用来实现对科学数据本身的引用。

2.1 DOI 系统

数字对象唯一标识符（DOI）是为了解决数字对象的跨系统操作，针对网络环境中数字对象具有移动、易变性和多源性的特点，用一种永久不变的数字对象定位方案来保证数据对象的地址永久不变。出版界引进数字对象唯一标识符也是为了应对网络环境下知识产权管理、保护以及产权交易。DOI 是由美国出版协会（AAP）提出并建立的标识系统，通过对 DOI 命名的分配、解析、描述和管理实现在网络环境下对数字对象提供唯一和持久标识并提供信息交换的系统^[9-10]，具有唯一性、永久性和多重链接和互操作等特点。

唯一性是 DOI 的核心价值，DOI 编码遵循 ANSI/NISO Z39.84-2000 的命名标准^[11-12]，保证了数字标识符在全球范围内的唯一性。DOI 编码的基本结构形式是 10.<REG>/<DSS>，其中 10 是 Handle System 最高注册机构将分配给 DOI 最高管理机构国际 DOI 联盟（IDF）的固定代码。<REG> 是 DOI 系统分配给子命名机构的代码，由 IDF 或其授权的注册代理机构（RA）负责分配。DSS 是 DOI 的后缀，是 DOI 子命名机构赋予数字对象的局部名称，由命名机构自主命名，其长度可不受限制。

DOI 是对象的数字化识别符，而不仅仅是数字对象的识别符，理论上能够对所有对象进行唯一标识，DOI 的唯一性和永久性是通过 Handle System 解析系统来实现的，并由国际 DOI 联盟（International DOI Foundation, IDF）来维护和运行。IDF 是全球 DOI 最高管理机构，创建于 1998 年，负责有关 DOI 的政策制定、技术支持、注册管理等业务。IDF 授权不同的 DOI 注册代理机构或注册中心为用户注册 DOI，负责 DOI 元数据登记、维护以及数据处理服务。在我国，中国科学技术信息研究所和万方数据股份有限公司在 2007 年共同申请成为中文信息资源领域 DOI 的注册代理机构，科学出版社则在 CrossRef 框架下建立了

二级 DOI 平台。

2.2 DOI 在科学数据出版中的应用

科学数据作为科研成果的重要表现形式,得到越来越多的重视,期望建立类似学术论文的数据出版和引用机制,进而建立科学数据的评价体系。同时,随着科学数据共享程度的改善,数据中心的工作重点和定位发生了变化。如国际科学理事会(ICSU)就希望对世界数据中心体系进行改革,强烈推荐建立一个有关科学数据和信息出版的战略性框架^[13-15]。IDF 很早就提出可以将 DOI 应用于科学数据的出版^[16],德国科学基金会在 2003 年至 2005 年资助德国 CODATA 启动了科学数据出版和引用项目,通过采用 DOI 来确认和定位以数字形态存在的数据串,并将该标识解析到其存放的有效的 URL,这使得联机出版的科学数据的引用成为可能^[9,17]。德国国家科学技术图书馆(TIB)在 2005 年正式申请成为科学数据领域的 DOI 注册代理。它的注册对象包括各种不同的类型,如晶体结构、地球模型和 3-D 模型等^[18]。科学数据 DOI 应用在国内还处在起步阶段,中文 DOI 中心也开始在科学数据领域推广 DOI 系统。其主要目标是通过与国内外相关机构合作,在中国科学数据领域推广 DOI 标准,使用户能够访问科学数据资源,建立科学数据引用与出版的示范应用,促进数据生产机构、服务机构方便地实现联合共享,实现中文科学数据资源的相互链接,并与国外科学数据机构使用相同的标识、链接和引用标准,实现中西文数字资源的链接与共享。

3 DOI 在“西部数据中心”中的应用

2007 年,“西部数据中心”与中国科学技术信息研究所合作,共同探讨如何利用 DOI 建立科学数据的永久地址,促进科学数据的出版和引用,并对科学数据的使用进行准确的跟踪和统计。中国 DOI 注册中心为“西部数据中心”分配了独立的 DOI 前缀 10.3972。“西部数据中心”首先选择“黑河综合遥感联合试验”的产出数据作为试点开展科学数据 DOI 注册,探讨科学数据的出版和引用,并逐步将 DOI 应用到“西部数据中心”的其

他数据资源上。

3.1 DOI 在黑河综合遥感联合试验中的应用

“黑河综合遥感联合试验”由中国科学院西部行动计划(二期)项目“黑河流域遥感-地面观测同步试验与综合模拟平台建设”与国家重点基础研究发展计划(973 计划)项目“陆表生态环境要素主被动遥感协同反演理论与方法”共同设计并组织实施,是在流域尺度上开展的以水循环及与之密切联系的生态过程为主要研究对象的大型航空、卫星遥感与地面同步观测科学试验。“黑河综合遥感联合试验”在黑河流域的上游寒区水文试验区、森林水文试验区和中游干旱区水文试验区 3 个区域开展了预试验、加密观测试验和常规观测试验 3 个阶段观测^[19-20]。通过积雪和冻融的微波辐射计和高光谱航空遥感与地面同步试验,森林结构参数和森林水文的高光谱、多角度热红外和激光雷达航空遥感试验,干旱区植被生物物理参数、植被结构、土壤水分和蒸散发的高光谱、多角度热红外、激光雷达和微波辐射计的航空遥感与地面同步试验,以及加密气象水文观测网持续观测,总共获得了包括卫星遥感、航空遥感和地面观测在内的科学数据约 2.3TB。为了促进试验产出数据成果的应用,两个项目共同决定在项目执行期间开展数据共享。数据经过质量控制、专家审议后,于 2010 年 7 月 8 日正式宣布对外共享,“西部数据中心”(http://westdc.westgis.ac.cn)和北京师范大学遥感数据中心(http://rsdc.bnu.edu.cn/)负责数据的申请和分发。

“黑河综合遥感联合试验”产出数据的 DOI 注册及其发布过程是:

(1)从数据管理的角度和试验的角度,将具有共同特点(如同一天的星机地同步观测数据,或者若干天的相同观测内容的数据)的一组数据确定为基本的数据单元,用一条元数据对这组数据进行整体描述。该元数据存储于西部数据中心的数据库中,采用 ISO 19115 标准。这些元数据由试验参加者和数据库建设者前后进行了 3 轮撰写和补充,元数据对应的数据实体也经过必要的质量控制,使元数据与数据实体实现了一对多的连接。在数据的正式发布前,召开了由项目负责人和专家共同组成的数据评审组,对元数据和数

据实体进行了评议。这个过程相当于学术论文的撰写、修改和评审过程。

(2) 在中国DOI注册中心为“西部数据中心”申请的独立前缀基础上, 西部数据中心自主地为每组数据集命名了DOI代码。DOI的格式是:

doi:10.3972/water973.xxxx.db

其中, water973代表黑河综合遥感联合试验, xxxx为数据标识码, db表示数据集(database)。

(3) 在西部数据中心元数据的基础上, 抽取必要的要素构建注册DOI所需的元数据。该元数据与DOI编码一起在中文DOI注册中心(<http://www.chinadoi.cn/>)进行注册, 并通过该中心汇总到全球DOI注册中心(<http://www.doi.org>)。西部数据中心将元数据的地址作为实体数据的地址与DOI编码进行关联。通过DOI解析后, 用户通过DOI的Web解析地址(如<http://dx.doi.org/10.3972/water973.0161.db>)就可以直接定位到该数据的元数据地址。当该数据的元数据地址发生改变时, 只要更新DOI中的元数据链接地址, 就可以保证DOI的Web解析地址(如<http://dx.doi.org/10.3972/water973.0161.db>)永久有效, 且不随DOI注册机构的变更而变化。当该数据被其他数据中心收录时, 只需在DOI中添加相应的地址, 就可以实现该数据的多重地址定位。

(4) “西部数据中心”为每一组数据制定了中英文的建议引用格式, 建议用户在引用该数据时参考该格式。该建议引用格式保存在“西部数据中心”的元数据中, 格式为:

<作者>.<数据标题>.<数据发布单位>.<发布时间>.<DOI>

经过反复的讨论, 为了突出试验参加者和试验参加单位的权益, 数据发布单位采用试验参加者所在的单位, 并且可以设置多个数据的发布单位。如试验中临泽草地站大孔径闪烁仪(LAS)的观测数据的引用格式如下:

中文格式: 刘绍民, 李召良, 徐自为, 唐伯惠, 王维真. 黑河综合遥感联合试验: 临泽草地站大孔径闪烁仪(LAS)数据集. 北京师范大学; 中国科学院地理科学与资源研究所; 中国科学院寒区旱区环境与工程研究所, 2008. doi:10.3972/water973.0161.db

英文格式: Liu Shaomin, Li Zhaoliang, Xu Ziwei, Tang Bohui, Wang Weizhen. WATER: Dataset of LAS (Large Aperture Scintillometer) Observations at the Linze Grassland Station. Beijing Normal University; Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences; Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, 2008. doi:10.3972/water973.0161.db

只要数据用户在参考文献中参照该格式进行了引用, 并标注了该数据的DOI编码, “西部数据中心”就通过文献检索的方式准确地查找到哪些文献使用了“黑河综合遥感试验”的产出数据, 并对数据的使用情况进行跟踪和分析。同时, 因为DOI解析地址永久不变, 在任何时候都可以通过DOI解析地址查找到该文献对应的数据。

(5) “黑河综合遥感联合试验”的数据共享分为两个阶段。一是从宣布数据共享开始到两个项目完全结束, 即数据共享的试运行期, 在此期间广泛听取数据使用者的反馈意见, 对数据实体、元数据以及元数据系统进行改进。二是两个项目结束后, 除了涉及保密的数据外, 所有数据全部开放共享。

3.2 其他数据资源的DOI注册及数据引用格式

“西部数据中心”不是一个数据生产单位, 其核心目标是通过数据共享、加工和集成为科学研究提供能够“直接”使用的数据产品。“西部数据中心”收集和管理的数据库资源来源广泛, 在数据的知识产权保护方面存在以下问题:(1)有部分自主生产的数据产品, 具有完全的知识产权, 可以明确制定数据的引用方法, 但该数据如果被其他数据中心收录, 如何保证数据版权信息的一致;(2)收集和整理了一大批历史数据, 虽然不拥有数据的原始版权, 但数据的格式和表现形式已经发生了变化, 以数据的原始版权信息进行引用, 无法体现出数据表现形式的变化和数字化过程中的质量控制信息;(3)有部分项目汇交的数据, 其知识产权并不十分清晰, 只有部分数据能够确定数据的版权信息, 但不能保证该数据不被其他数据中心收集, 或被数据中心收集后, 数据版权信

息与西部数据中心保持一致;(4)根据集成研究的需要,有些数据是从其他数据集的基础上加工而来,这种集成后的数据如何进行标注。

在采用 DOI 之前,“西部数据中心”根据自己的理解和能力,采用数据使用声明和数据建议引用格式两种方式重视和保护数据的知识产权:

(1)“西部数据中心”明确要求用户在使用数据时须在致谢部分声明数据的来源,格式如下:

中文格式:数据来源于国家自然科学基金委员“中国西部环境与生态科学数据中心”(http://westdc.westgis.ac.cn)

英文格式: This data set is provided by “Environmental & Ecological Science Data Center for West China, National Natural Science Foundation of China”(http://westdc.westgis.ac.cn)

(2)对于有自主版权的数据或能够追查出原始版权的数据,数据中心建议用户通过数据引用的方式在参考文献中对数据进行引用。如黑河流域 1:100 万草场图,该数据来源于仇保铭等 1988 年出版的纸质图,我们建议用户在使用数据时按照如下格式进行引用:

[1] 仇保铭,高前兆.黑河流域草场类型图,西安地图出版社.1988.

“西部数据中心”对数据进行 DOI 注册的初衷是为了保护数据的知识产权,但很多数据的知识产权不明确,我国现有的法律没有对数据的知识产权作明确的规定,这在很大程度上限制了科学数据的 DOI 注册。“西部数据中心”根据《科技进步法》以及《柏林宣言》和《布达佩斯开放取用倡议》的精神,尝试如何在充分尊重数据原有知识产权的情况下,对科学数据进行 DOI 注册,并制定数据引用格式,促进科学数据的出版和引用。

在数据引用格式方面,西部数据中心总体上参考国家标准《文后参考文献著录规则》中电子文献的引用格式。

<主要责任者>.<数据标题(数据版本)>,(其他数据责任者和数据标题信息,如数字化作者和数字化后的数据标题),<数据发布地>:<数据发布机构>,(数据再分发机构[分发]),<数据发布时间>.(doi 编码)

其中:

主要责任者:是对数据的内容负主要责任的个人或团体,这里强调数据的原始作者,数字化的作者不在这里表现。

数据标题:是数据的正式标题,包括版本号。

数据出版/发布日期:是指数据(或原始数据)的发布日期,而不是数据的生产日期,也不是数据的数字化日期。

数字化者、整理者或汇编者:数字化者是指对纸质数据进行数字化,数据整理者是在原始数据的基础上进行整理产生新的附加值或进行质量控制,数据汇编者不产生新的信息,如数据切割、地图投影等。

数据发布机构:一般指数据中心,但也可以是研究项目或研究机构。

数据分发者或协助发布者,当数据分发者与数据发布者不一致时,可加该字段。

对于“西部数据中心”的不同数据类型,采用不同的方式进行 DOI 注册和数据引用,优先体现原始数据的版权信息,同时体现数据处理者和数据再分发机构的信息。

(1)具有自主知识产权的或获得授权进行数据发布的数据,DOI 由“西部数据中心”负责编码和注册,数据引用格式是:

<数据作者>.<数据标题>,<数据发布地>:<数据发布机构>,<数据发布时间>.<doi 编码>

[1] 晋锐,李新.中国长时间序列地表冻融数据集,兰州:中国西部生态与环境数据中心,2006. doi:10.3972/westdc.xxxx.db

[2] 王世杰等.西南喀斯特地区典型洞穴化学指标检测数据,贵阳:中国科学院地球化学研究所,2009. doi:10.3972/westdc.xxxx.db

(2)由纸质介质数字化后的数据资料,数据的表现形式发生了变化,数据内容没有根本性变化,应该突出数据的原始作者,但应尽量标注数字化者的信息,以示对数据化的质量进行负责,DOI 代码继承原有的 ISBN 编码。数据引用格式是:

<数据原始作者>.<数据原始标题>,(数字化作者[数字化]).<数据发布地>:<数据发布单位>,<原始数据出版时间>.<数据再分发机构[分发]>,(数据数字化时间).(doi 编码)

[1] 李福兴,仇保铭.黑河流域土壤图,李新

等 [数字化]. 西安: 西安地图出版社, 1988. 兰州: 中国西部环境与生态科学数据中心 [分发], 2000. doi:10.3972/ISBN: 9787805450056 /7805450056

(3) 对于二次加工后的数据产品, 数据的内容发生改变, 参照论文集或专著析出文献的引用方式, 将数据加工者和加工后的数据标题列在前面, 数据的原始作者和标题放在后面。如果原始数据有 DOI 代码, 则在原有 DOI 代码的基础上添加必要的标识代码注册新的 DOI 代码, 如果原来的数据没有 DOI 编码, 则由“西部数据中心”重新注册。数据的引用格式是:

<数据加工者 [加工]>.<加工后的数据标题>.<数据原始作者>.<数据原始标题>.<数据发布地>.<数据发布机构>.<数据原始发布时间>.<数据再分发地>.<数据再发布单位 [分发]>.<数据再分发时间>。(DOI 编码)

[1] 车涛, 晋锐, 吴立宗等 [加工]. 中国地区被动微 SMMR 亮度温度数据集. Knowles, K.W., Eni Njoku, R.L. Armstrong, and M.J. Brodzik. Nimbus-7 SMMR Pathfinder Daily EASE-Grid Brightness Temperatures. Boulder, Colorado USA: National Snow and Ice Data Center. Digital Media, 2002. 兰州: 中国西部环境与生态科学数据中心 [分发], 2005. doi:10.3972/westdc.xxxx.db

(4) 对于授权再分发的数据, 仅在原有的数据引用格式的基础上, 添加数据再分发单位信息, 以示数据获取来源的不同, DOI 采用原有的编码, “西部数据中心”不主动注册新 DOI 编码。数据的引用格式是:

<数据原始作者>.<数据原始标题>.<数据出版地>.<数据出版机构>.<数据发布时间>.<数据再分发地>.<数据再发布机构 [分发]>。(DOI 编码)

[1] 徐祥德等, JICA 中日气象灾害合作研究中心项目数据集. 北京: 中国气象科学研究院灾害天气国家重点实验室, 2008. 兰州: 中国西部环境与生态科学数据中心 [分发].

4 问题与讨论

4.1 科学数据出版的意义

科学数据的积累和共享对我国科学研究的

可持续发展有长久支持作用^[21-23], 我国先后建立了若干数据中心, 尤其是在“十一五”期间, 科技部启动了科学数据共享平台, 投入了大量的资金, 建立了若干准国家级的数据中心。但总体而言我国的科学数据共享尚处于培育阶段, 距离在全社会范围内形成良好的数据共享氛围相差甚远。在诸多影响科学数据共享的因素当中, 权益不分明和缺乏统一的管理是阻碍数据共享的重要原因^[24]。科学数据是一种研究资源, 与学术论文和专利处于同等重要的战略位置, 但长期以来, 科学数据仅作为“原材料”来使用, 没有形成完善的评价机制, 极大地影响了科学家和数据生产单位开展数据共享的积极性。将 DOI 引入到科学数据共享中具有重要的意义。

(1) 有利于形成数据的引用和评价体系

科学界长期以来希望科学数据与文献类似, 在成果发表时进行引用, 同时根据科学数据被引用次数来衡量和评价科学数据价值。DOI 已经在出版领域得到了广泛和成功的应用, 将 DOI 引入到科学数据共享中, 能够将科学数据的评价与学术论文的评价“捆绑”在一起, 可以充分地利用学术论文的评价体系, 形成和建立科学数据的评价体系。

(2) 有利于推动同行评议和出版前评议

并不是所有的“原材料”都有被引用的价值, 因此以影响因子的方式来评价数据的价值有失公平, 仅依靠数据的引用来推进数据的共享和出版也存在一定的局限性。但随着国际上学术论文评审机制的日趋完善, 科学数据的质量是决定论文质量的重要因素, 论文的同行评议和出版前评议将是推动数据出版的另一重要的因素。

(3) 有利于促进数据和文献的集成

DOI 系统本身是一个商业系统, 所有的注册机构每年都需要向 IDF 缴纳年费。DOI 之所以能够在出版领域得到广泛的应用, 与学术期刊集成和交叉访问的大趋势有很大的关系。CrossRef 的成功就反映了这一趋势。CrossRef 成立于 2000 年, 致力于实现数据库(集)之间(或数据与文献之间)交叉参考链接, 目前已经初步建立了一个以 DOI 为核心, 基于引用关系、开放的科技文献知识链接体系。学术论文的基本价值表现为被更多

的研究者所重视和引用,这就促使期刊或出版机构自发地遵循 DOI 标准,从而借助 CrossRef 通道实现学术论文在更大范围内传播。

现在,越来越多的研究者关注如何通过文献查找感兴趣的数据。这在一定程度上促进了科学数据的出版。但更重要的是,这种相互链接可能催生更大的学术成果集成系统,一旦科学数据库形成规模和权威,对科学数据出版的吸引力将更加直接和强烈。我国的科技界和数据共享界应该未雨绸缪,及早建立科学数据领域的 CrossRef 系统。

4.2 DOI 体系在促进科学数据出版中存在的不足

DOI 可以应用于科学数据出版中,但就目前而言,DOI 系统本身的优点还仅体现在数据管理层面。DOI 系统要成功应用于科学出版领域就必须建立一套行之有效的管理和运行体系。

(1) 利用 DOI 技术开展数据的出版与引用必须与科学数据共享紧密结合起来,否则将成为空中楼阁。不存在的数据是没有价值的,同样,不共享和不公开的数据也没有多少引用价值。

(2) 科学数据的 DOI 注册不等同于科学数据的出版,DOI 的核心价值是为数据对象提供永久的、唯一的数字标识符。在理论上,可以为任何数据提供数字化标识符,但科学数据的出版只有通过质量检查或同行评议后才有推广应用的价值,两者的目标存在明显的差异。这一问题需要与当前的学术论文评审制度和数据中心转型结合起来加以解决。

(3) 要加强科学数据的引用,在学术论文发表过程中形成约定俗成的习惯,这需要依靠现有数据出版机构的支持。在相当长的时间内,只有通过出版社或编辑部的强制措施,才能形成对科学数据的引用习惯。

(4) 科学数据的出版缺少后评价体系。学术论文的引用之所以得到很好的执行与 SCI (科学引文索引)、EI (工程索引)、ISTP (科技会议录索引) 的收录作用和文章的引用因子密不可分,科学数据的出版也需要一套行之有效的后评价体系。与论文的引用因子不同,对于未来的科学数据后评价体系,权威科学数据库的收录和集成可

能是主要的评价因子和动力。

4.3 科学数据出版带来的挑战

(1) 数据中心的定位

最初,全球范围的数据共享程度低,科学家个人获取研究所需的科学数据比较困难,也容易造成重复劳动,成立科学数据中心的初衷很多都是为了与数据生产单位进行交涉,为大规模的研究团队提供数据支持。随着数据共享程度的提高,数据中心的协调作用就不再显著,一些没有特色的数据中心逐渐被边缘化。我国成立的各种数据中心在相当长的一段时间内都是以历史数据的恢复与整理为主,数据中心不约而同地努力收集数据,数据收录的门槛很低,数据的同质化问题很严重。随着我国科学数据共享水平的提高,这种现象必然会发生改变,现有的数据共享体系必然会重新“洗牌”。数据中心只有坚持精英化和特色化,并能够提供高质量数据服务,才能得到保留。同时,必然有部分数据中心走集成化道路,而制定科学数据和信息出版的战略性发展框架是重要的一个方面。

(2) 数据的知识产权问题

除了正式出版发行和申请专利外,科学数据尚没有一个专门的知识产权认证机构,大量科学数据的知识产权没法得到确认。如何建立适合科学数据特点的知识产权认证体系,值得深入研究和探索。此外,数字化、二次加工和数据集成时的知识产权如何确定一直困扰着数据界,直接影响了数据的再发布。

(3) 数据管理

数据的永久保存与科学数据的最优化管理存在矛盾,一方面现有的数据管理要充分考虑如何对数据的状态进行保留和记录,另一方面数据的出版和引用也需要考虑数据的动态变化。

参考文献

- [1] Lu Hongzhou, Tang Yiwei. How to Register Geneorder in GenBank[J]. Journal of Microbes and Infections, 2009, 4(3): 173. (in Chinese)
〔卢洪洲,汤一苇. 如何在 GenBank 中注册基因序列[J]. 微生物与感染, 2009, 4(3): 173. 〕
- [2] The National Center For Biotechnology Information[EB/

- OL]. [2010-09-10].<http://www.ncbi.nlm.nih.gov/genbank/GenbankOverview.html>.
- [3] Green T. We Need Publishing Standards for Datasets and Data Tables[R]. USA:OECD, 2009.
- [4] IPY DIS. How to Cite a Data Set[EB/OL]. [2010-09-10].<http://ipydis.org/data/citations.html>.
- [5] SEDAC. SEDAC Citation Guide and Index[EB/OL]. [2010-09-10].<http://sedac.ciesin.columbia.edu/citations/>.
- [6] Altman M, King G. A Proposed Standard for the Scholarly Citation of Quantitative Data[EB/OL]. [2010-09-10].<http://www.dlib.org/dlib/march07/altman/03altman.html>.
- [7] Lehner B, Verdin K, Jarvis A. HydroSHEDS Technical Documentation (v 1.1)[R]. USA:World Wildlife Fund, 2008.
- [8] Global Land Cover Facility. What is this Citation?[EB/OL]. [2010-09-10].<http://glof.umiacs.umd.edu/data/landsat/>.
- [9] Paskin N. Digital Object Identifiers for Scientific Data[J]. Data Science Journal, 2005(4): 12-20.
- [10] Paskin N. Components of DRM Systems: Identification and Metadata[C]//Lecture Notes in Computer Science. Berlin: Springer, 2003.
- [11] ISO. ISO 26324:Information and Documentation – Digital Object Identifier System[S]. Geneva, Switzerland:IDT, 2009.
- [12] ANSI/NISO. ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier[S]. Bethesda, Maryland:NISO, 2000.
- [13] WDC Modernization Task Team. Towards a New World Data Center System: Meeting Global Needs[R]. USA:ICSU WDC Panel, 2003.
- [14] Wang Juanle, Sun Jiulin. Review, Reform and Prospect Analysis of World Data Center[J]. Advances in Earth Science, 2009, 24(6):612-620.(in Chinese)
〔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.〕
- [15] He Defang, Zhang Xu. Study of the Applications of DOI in Scientific and Technical Information Sharing[J]. New Technology of Library and Information Service, 2007(8):26-29.(in Chinese)
〔贺德方, 张旭. 服务于科技信息资源共享的数字对象唯一标识应用研究[J]. 现代图书情报技术, 2007(8): 26-29.〕
- [16] Paskin N. Digital Object Identifiers and Digital Preservation of the Record of Science: Symposium on Digital Preservation of the Record of Science[EB/OL]. [2010-09-10].http://www.doi.org/topics/020210_CSTI.pdf.
- [17] Lautenschlager M, Sens I. Konzept zur Zitierfähigkeit[J]. Wissenschaftlicher Primierdaten Information – Wissenschaft und Praxis, 2003, 54: 463-466.
- [18] Peng Jie, Tu Yong. Research on the Scientific Data Citation[J]. Digital Library Forum, 2008(10):14-18, 45.(in Chinese)
〔彭洁, 涂勇. 科学数据引用的探讨[J]. 数字图书馆论坛, 2008(10):14-18, 45.〕
- [19] Li, X, Li, XW, Li, ZY, et al. Watershed Allied Telemetry Experimental Research[J]. Journal of Geophysical Research. 2009, 114, D22103. doi:10.1029/2008JD011590.
- [20] Li Xin, Ma Mingguo, Wang Jiang, et al. Simultaneous Remote Sensing and Ground-based Experiment in the Heihe River Basin: Scientific Objectives and Experiment Design[J]. Advances in Earth Science, 2008, 23(9):897-914.(in Chinese)
〔李新, 马明国, 王建, 等. 黑河流域遥感—地面观测同步试验: 科学目标与试验方案[J]. 地球科学进展, 2008, 23(9): 897-914.〕
- [21] Qin Dahe. Scientific Data Sharing—The Necessity for Efficient Use of National Resources[J]. China Basin Science. 2003(1):24-29 (in Chinese)
〔秦大河. 科学数据共享: 国家资源有效利用的必然选择——以中国气象局资料共享工作为例[J]. 中国基础科学. 2003(1):24-29.〕
- [22] Sun Shu. Earth Data—An Important Resources for Geoscience Innovation[J]. Advance in Earth Science, 2003, 18(3):334-337.(in Chinese)
〔孙枢. 地球数据是地球科学创新的重要源泉——从地球科学谈科学数据共享[J]. 地球科学进展, 2003, 18(3): 334-337.〕
- [25] Xu Guanghua. Implement Scientific Data Sharing for Strengthening the National Sci-tech Competitive Power[J]. China Basic Science, 2003(1):7-11.(in Chinese)
〔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.〕
- [26] Lu Peng, Miao Liangtian, Li Zhixiong, et al. Investigation on Scientific Data Sharing Situation and Its Analysis[J]. Earthquake, 2007, 27(3):125-130.(in Chinese)
〔路鹏, 苗良田, 李志雄, 等. 我国科学数据共享现状的调查与分析[J]. 地震, 2007, 27(3): 125-130.〕