中美地理科学数据出版平台研究*

张玲玲[1] 陈媛媛[1,2]
(1. 黑龙江大学信息管理学院，哈尔滨 150080；2. 南开大学商学院，天津 300071)

摘要：数据出版是推动科学数据共享的重要手段，对中美地理科学数据出版平台进行调查研究，可为我国科学数据出版平台建设提出具有借鉴意义的发展建议。本文选取中美8个具有代表性的地理科学数据出版平台作为研究对象，从数据出版的数据提交、同行评审、数据发布和永久存储、数据引用以及影响评价这5个基本环节展开调研，提出我国地理科学数据出版平台发展建议：制定“全流程”政策、建立详细提交流程、完善数据评审体系、构建规范数据发布标准、保障科学数据知识产权、扩大数据质量反馈途径。

关键词：数据出版；科学数据；出版平台
中图分类号：G250.73　　DOI：10.3772/j.issn.1673-2286.2020.10.010
引文格式：张玲玲，陈媛媛. 中美地理科学数据出版平台研究[J]. 数字图书馆论坛，2020 (10)：67-72.

科学数据是科研活动重要内容和主要产物[1]，对科学数据进行出版，在有效避免重复劳动、防止学术造假、提高数据发表者知名度的同时，能够促进数据共享和重用，推动科技创新和经济的发展。数据出版也是解决数据知识产权保护和推动数据广泛共享难题的有效机制[2]。因此，科学数据出版逐渐引起各个国家的高度重视。2018年，一批国家科研资助机构推出cOAlition S计划；2019年5月31日，cOAlition S发布了修订的开放获取Plan S和实施指南，并指出自2021年起，所有由地区、国家、国际研究理事会和资助机构提供的公共或私人资助的学术出版物，必须在开放获取期刊、开放获取平台上出版，或通过开放获取知识库立即提供，而不受封禁[3]。近年来，我国也开始重视科学数据出版平台建设，如2014年我国创建了全球变化科学研究数据出版系统，被科学技术部称赞为“中国科技数据共享新的里程碑”[4]；2017年我国创建了地质科学数据出版中心，它是集元数据、实体数据、论文关联一体化出版的数据平台；2018年国务院办公厅发布了《科学数据管理办法》，其中明确提出要积极推动科学数据出版和传播工作[5]。我国的科学数据出版和传播活动开始逐渐步入正轨。

数据出版模式有很多种，但仍可归为3种模式，即独立数据出版模式、作为论文关联和辅助资料的出版模式以及数据论文出版模式[6]。这些出版模式主要借助于数据仓储、机构库以及期刊自行发表3种方式对外公布，其中以数据仓储和机构库为数据出版形式的数据质量、数据规范性最强。相关学者对数据出版平台进行了研究，段青玉等[7]根据FAIR原则调研了人文社科的平台，总结出数据出版平台需要从5个方面实现数据的利用；王丹丹[8]对数据管理的技术平台进行调研，总结出平台的基本功能和用户体验要求；秦顺等[9]对欧美14个数据出版平台的5个方面展开调研，并针对各方面要素归纳其服务特点。但我国还没有对某一学科领域的数据出版平台建设情况进行针对性研究。随着地理信息系统、全球定位系统的不断发展[10]，关于地理数据的出版平台建设也在不断完善。因此，本文就中美地理科学领域的数据出版平台情况展开调研分析，以期为我国科学数据出版平台服务建设提供启示。

1 调查对象和方法

美国科学数据出版平台服务长期处于领先地位，

*本研究得到黑龙江大学校级研究生创新科研项目“高校图书馆科研数据管理服务体系构建”（编号：YJSCX2020-128HLJU）资助。

实践经验丰富，因此本文以美国创建的地理数据出版平台为对比研究对象。为了使选取的数据出版平台具有代表性，以已在科研数据知识库注册目录系统（re3data.org）注册的平台为样本，以平台创立时间、合作机构、存储数据量、数据更新速度为主要依据，选取以下平台。主要包括美国地质调查局（United States Geological Survey，USGS）、美国国家海洋和大气管理局（National Oceanic and Atmospheric Administration，NOAA）、美国地热数据系统（National Geothermal Data System，NGDS）和美国国家冰雪数据中心（National Snow and Ice Data Center，NSIDC）四大典型性代表平台。国内方面，经过文献阅读和网络调研，选取我国的全球变化科学研究数据出版系统、国家地球系统科学数据共享服务平台、国家气象信息中心及中国南北极数据中心进行调研分析。

2 调查结果分析

数据出版不是简单的数据发布，而是包括数据提交、同行评审、数据发布和永久存储、数据引用、影响评价5个基本环节[11]。下文据此对中美8家地理科学数据出版平台的发展情况展开网络调研。

2.1 数据提交

2.1.1 描述数据

为提高科学数据的利用价值，数据提交者在提交数据时对数据进行规范性、完整性描述。其中，全球变化科学研究数据出版系统和NSIDC做出了很好的规范，数据描述包括标题、关键词、摘要、相关图表、文件格式、资料收集方法等要素，并进行了字符长度规范，如数据集标题不可超过85个ASCII字符等。为使数据使用者了解数据存在的价值，平台要求填写数据的可用性说明，包括数据提供的内容、背景、潜在应用程序和参数等的概述。地理科学学科的数据格式有很多种，如GIF、JPG、TIFF等，但这些格式结构简单，无法存放更多有用信息。NOAA和NSIDC选择HDF作为数据提交格式之一，它具有的异构性、跨平台性、简单分享性等特点，使存储不同类型的图像和数码数据的文件格式，可以在不同类型的机器上传输，同时还有统一处理这种文件格式的函数库[12]。USGS和NOAA选择netCDF（网络通用数据表单）为另一种数据提交格式，该格式具有自我描述性、随身携带性、可扩展性、可追加性、可分享性等特点，支持创建、访问和共享面向阵列的科学数据。而我国地理科学数据出版平台对数据描述内容、数据格式要求较为单一。

2.1.2 数据提交方式

为了保证数据的严谨性，8家地理科学数据出版平台在数据提交之前，都需要完成平台的注册登录。在数据提交过程中，USGS要求提交者根据相关调查表收集元数据内容，或者根据元数据创建工具收集元数据，收集的数据通过USGS元数据解析器或Microsoft XML记事本验证元数据是否创建正确，创建正确方可提交。NOAA对数据集提交的大小有一定规范，数据集小于20GB，要求使用Send2NCEI（S2N）在线工具记录并提交数据；数据集大于20GB，要求提交至用于档案馆藏的高级跟踪和资源工具（Advanced Tracking and Resource Tool for Archive Collections，ATRAC）中。NGDS要求数据提交者根据给定信息交换方案的标准文件填写信息，并在指定网页对文件进行验证方可提交。我国四大地理科学数据出版平台中数据提交方式都是以网页形式填写数据集相关信息并进行提交。

2.2 同行评审

同行评审主要是对数据内容进行评审，科学数据出版中的数据评审主要包括科学性、技术性以及监护性3种评审维度[13]。科学性评审主要体现在对数据内容的准确性、真实性等进行评审，技术性评审主要体现在对数据及元数据的质量进行技术评审，监护性评审主要体现在对数据以外的相关文档进行管理和评估。美国数据出版平台在这3种数据评审维度中有较好的体现。其中NOAA规定了信息质量和传播前的审查标准，主要体现在信息的效用、完整性和客观性。信息质量是传播前审查不可或缺的一部分，它也是NOAA收集信息的组成部分，并已纳入《减少文书工作法》（PRA）要求的审批流程中，以帮助提高NOAA向公众传播信息的质量。为了更加准确收集信息，NOAA采用数据收集和相关培训等方法对数据进行有效监管，并使用一级或二级标准、基础工程和科学方法对仪器进行校准，旨在满足目标用户的要求。为了避免数据存储过程中存

在丢失，NOAA会对标准操作程序（SOP）进行定期审查和修改。为了保证数据的规范性，NOAA提供了DMP Tool、DMP Editor等工具，对如何获取或收集数据、数据收集的时间以及预算、如何对数据进行质量检查、数据如何存储，以及访问和保护等问题进行详细的管理规划。在同行评审相关政策中，USGS和NOAA皆采用了美国管理和预算局（OMB）所发布的同行评审信息质量公报（Final Information Quality Bulletin for Peer Review）[14]，该公报规定了科学信息何时需要同行评审的最低标准，以及各机构在不同情况下应考虑的同行评审类型，提出同行评审计划应建立一个透明的过程，包括对同行评审计划的网络可访问描述。在决定何种类型的同行审查机制适用于特定的信息产品时，该公报要求机构应考虑个人审查还是小组审查，同行评审时间，审查范围，审稿人的选择，信息披露和归因，公众参与评论，审稿人意见的处理，以及事先的同行审查是否充分等问题，关于审稿人在审稿过程中可能存在的利益冲突进行了详细的规避措施阐述。

我国4家地理科学数据出版平台中，只有全球变化科学研究数据出版系统对同行评审内容进行了阐述，且于2014年起草了《科学数据DOI注册与发表同行评议规定和评议表》[13]，评审的内容包括数据集的产权是否清晰，实体数据内容与数据论文阐述是否一致，实体数据在内容或格式上是否具有智力投入，实体数据的质量是否符合误差小于10%原则，以及引用他人数据记录是否符合小于10%原则等，在诸多方面确保了数据的质量[2]。但在管理员对数据技术性评审和监护性评审方面缺乏相关审核制度。其他3个出版平台都采用内部评估的方式进行同行评审，并未给予对外公布。

2.3 数据发布与永久存储

美国地理科研数据出版平台建设时间较长，平台数据服务较为成熟。其中USGS要求需满足7个要素方可发布数据：①数据管理计划（Data Management Plan，DMP），USGS要求对于每个项目都需要制定数据管理计划，该计划应在开始项目工作之前编写，并在整个项目中进行更新；②科学数据格式确定，通过确保数据提交者的数据采用开放格式，以保证数据使用寿命；③符合FGDC的元数据标准；④USGS数据对象标识符（DOI），在USGS中发布的所有数据都必须具有DOI；⑤数据和元数据审查，USGS要求发布的任何数据都必须经过审查和批准，从而确保数据的完整性、真实性、准确性和有用性；⑥可接受的数据存储库，发布的数据的位置应该是USGS“受信任的数字存储库”的组成部分；⑦数据提交者发布的数据必须通过USGS科学数据目录向公众和科研社区共享。为了使数据利用者快速了解并获取平台信息，8家数据出版平台中有5家选择以API的形式进行科学数据交互（见表1），其中USGS选择REST风格（Representational State Transfer，表述性状态传递）进行数据描述和数据交互。RESTful API具有统一接口URL，能够基于HTTP协议实现多种格式的数据调用，极大地扩展了科学数据出版的覆盖面，有助于科学数据的永久存储[9]。我国的全球变化科学研究数据出版系统、国家地球系统科学数据共享服务平台、国家气象信息中心选择了自主研发框架进行科学数据交互，这为数据交互和共享带来了一定的阻碍，而选取作为统一规范路径的API接口对数据进行标识、关联与交互可有效避免这一障碍。

表1 数据出版平台数据交互、数据标识、许可协议调查表

平　台	数据交互方式	数据标识符	许可协议
全球变化科学研究数据出版系统	-	DOI	CC BY 4.0
国家地球系统科学数据共享服务平台	-	DOI	-
国家气象信息中心	-	-	-
中国南北极数据中心	SOAP	-	CC、CC BY 3.0
USGS	REST、APIS	DOI	Public domain
NOAA	FTP	DOI	Public domain
NGDS	other	URL	CC、ODC、OGL、other
NSIDC	FTP	DOI	CCO、other、Copyrights

2.4 科学数据引用

数据引用是数据出版的关键环节，是保障数据作者与管理者数据权益的有效方式[15]。截至目前，应用最多的数字标识符有DOI、URL、OpenURL等。从表1可以看出，大部分数据平台的数据引用方式选择DOI，DOI具有唯一且永久命名、解释链接、多重解析、唯一标识符、点击即链接、元数据和链接地址可更新、元数据/DOI可查询等特点，数据标识贯穿数据出版整个过

程，在数据出版过程中可做到版权保护，对数据资源的可信度和质量做到有效保障，因此，DOI的应用和研究也是最为广泛的[16]。

阻碍数据出版主要体现在数据共享与知识产权保护这一矛盾，建立共享协议有助于解决这一突出问题。由表1可知，除了USGS和NOAA放弃版权将数据完全公布到公共领域（public domain）以外，大部分平台采取CC BY、CCO共享协议，十分重视科学数据在出版过程中的知识产权保护。不同共享协议有不同的优点，若原始数据采用的是CCO和CC BY许可协议，其数据可采用CC家族的任意类型许可协议，其中CC BY4.0版本具有较好的兼容性，这将极大地促进了数据的利用、重用和创造性使用[17]。

2.5 科学数据影响评价

对科学数据进行影响评价具有两种目的：一是让数据使用者直观评判数据质量；二是将评价指标纳入科研成效评价指标中，激励科研人员主动参与数据出版与数据共享过程[11]。NOAA可允许数据访问者对发布的数据进行星级评价，用户可通过星级分数来了解数据发布的质量。值得关注的是，为了使平台建设更加完善，NOAA提供了平台反馈，点击反馈链接可进行填写网站客户满意度调查问卷[18]，内容为对网站整体印象、信息组织、寻找特定信息难易程度、改善网站意见等相关问题进行调查。NOAA、USGS、NGDS、NSIDC 4个平台都与社会媒体进行合作，其中USGS合作的社会媒体较多，如Google、github、Facebook、youtube、TWTTER、FLICHR、instagram等，使科学数据可进行广泛地分享，这不仅可以使更多数据得到重用，还可以加快数据影响评价速度，从而识别优质数据，增强科研人员知名度，进而提高科研人员的共享意识。

我国地理科学数据出版平台除了全球变化科学研究数据出版系统以外，其他出版平台对发布的科学数据可进行星级评价（共分为5个等级，依次为非常满意、满意、一般、不满意、非常不满意），其中，国家气象信息中心和中国南北极数据中心的数据可分享到微信、微博等社交媒体上。4个出版平台都可通过数据访问次数、下载次数、引用次数了解数据质量。国家地球系统科学数据共享服务平台与国家气象信息中心皆建立了微信公众号，不定期地分享最新的地理科学数据，以供数据使用者及时了解相关信息。

3 我国科学数据出版平台发展建议

我国的科学数据出版平台建设较晚，但随着人们的数据共享意识不断提升，该领域的发展引起国家和科研机构的高度重视，并在数据出版领域做出了有益探索。例如，全球变化科学研究数据出版系统在联合国大会中得到了一致肯定，并一致认为其是发展中国家实现科学数据共享可借鉴的实践案例，其通过互联网实现科学传播和公益性共享的机制为科学数据知识产权保护和数据共享这一问题的解决起到了很好的借鉴作用。但从整体而言，我国数据出版平台实践较为薄弱，需要进一步吸取美国等国家数据出版平台建设所积累的经验。

3.1 制定“全流程”政策

数据出版过程面临复杂的知识产权保护问题，知识产权是否能得到有效保护直接影响科研工作者数据共享意识。因此，为使数据出版流程更加顺利，制定贯穿科学数据出版流程的规范化科学化政策尤为重要。《科学数据管理办法》的颁布，打破了我国科学数据无法可依的局面，为我国数据共享的出版和引用提供了方向，但其对具微观的层面还有待加以完善和优化[19]。无论国内、国外，数据出版都处于尚在探索的阶段，建立一个多样互补的数据出版模式，并对数据出版模式流程中数据提交、同行评审、数据发布和永久存储、数据引用及影响评价各个环节制定详细细则，明确规范化、标准化的多方主体合作机制，推动科学技术与服务理念持续为科学数据出版平台服务。

3.2 建立详细提交流程

规范数据提交格式是实现科学数据出版的前提条件。不同学科的数据存储格式不同，因此在设置在线提交系统中需要设有合适的样例来完成此项工作[20]。在数据提交流程中要十分注重以下环节：①数据的描述，清晰地数据描述有助于数据利用者对数据的理解和重用，因此对数据的描述应详细包括数据的标题、关键词、摘要、相关图表、文件格式、资料收集方法等，对于相关描述信息可进行适当的字符限制以免过于冗长，NSIDC对数据描述做出了很好的规范；②可用性说明，包括介绍数据提供的内容、背景、潜在应用程序和参数

等的概述；③数据提交验证，可通过创建元数据解析器对提交的数据进行验证，数据验证成功方可提交，这样将减少大量人力物力，提高数据出版流程速度。

3.3 完善数据评审体系

我国在数据评审过程中，在学习借鉴全球变化科学研究数据出版系统平台的科学性评审的同时，也应该加强技术性评审及监护性评审。科学数据的同行评审是产生科学性数据的重要保障，一些工具和过程可能有助于快速、便捷地开展数据同行评审[21]，如USGS创建了DMP Tool、DMPEditor、ezDMP、Microsoft Word Templates、Google Forms等工具，极大地简化了同行评审过程[22]。我国科学数据出版平台也应根据数据本身性质创建审核工具，提高评审效率。

3.4 构建规范数据发布标准

国内的数据出版平台对数据如何发布、数据发布要素皆无详细的规范和要求，只是简单的发布与共享，与具质量可信、唯一标识、知识产权清晰等特征的数据出版平台仍具有一些差距[9]。从数据存储层面而言，以API作为数据编程接口，实现数据交互和关联，以DOI作为数据永久标识符，实现数据永久追溯，将有助于数据出版平台有序发展。

3.5 保障科学数据知识产权

通过观察8家数据出版平台服务可知，建立有约束力的共享协议和数据引用标准能够很好地解决数据出版过程中的知识产权保护和数据共享问题。2017年12月，我国印发了《信息技术科学数据引用》（GB/T35294-2017）国家标准，实现了基于OID（Object Identifier）数据标识符的引用格式，这与国家层面科学数据引用规范高度契合[23]。在此基础上，结合我国实际情况，深化DOI、OID数据标识符的应用，保护数据创造者知识产权，与此同时，建立有约束的数据共享协议，实现科学数据的广泛应用。

3.6 扩大数据质量反馈途径

数据在发布之前进行同行评审可对数据进行质量控制，数据发布之后进行影响评价可扩大数据发布者影响力及数据的重用。美国为了使科研数据得到广泛的推广，与各大社会媒体合作，如Google、github、Facebook、youtube等，使数据可随时转发到各大媒体平台，实现数据共享以供更多专业人士参考和评价。为了提高数据出版平台的用户体验，NOAA提供了反馈链接并对用户进行网站满意度调查。我国也应积极与各大媒体合作，实现与百度、微信、微博、QQ等各大媒体平台合作，不断提高数据的影响价值，并将科学数据列入科技成果体系，使科学数据成果列入科研人员对科学贡献的评价体系，提高科研人员数据共享意识，促进更多优质数据得到共享，如此循环，推动科学数据出版持续健康发展。

参考文献

[1] 邱春艳. 国内外科学数据出版理论研究述评 [J]. 中国科技期刊研究，2019，30（3）：271-279.

[2] 刘闯，郭华东，UHLIR P F，等. 发展中国家数据出版基础设施与共享政策研究 [J]. 全球变化数据学报，2017，1（1）：3-11.

[3] The Plan S Principles [EB/OL]. [2020-04-24]. https://www.coalition-s.org/principles-and-implementation/.

[4] 全球变化科学研究数据出版与共享系统获得联合国世界信息峰会奖 [J]. 地理学报，2018，73（5）：987.

[5] 国务院办公厅. 关于印发科学数据管理办法的通知 [EB/OL]. [2019-12-30]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

[6] 刘兹恒，涂志芳. 数据出版及其质量控制研究综述 [J]. 图书馆论坛，2020，40（10）：99-107.

[7] 段青玉，王晓光. 人文社科数据出版平台FAIR原则应用调查研究 [J]. 科技与出版，2019（4）：6-11.

[8] 王丹丹. 科学数据出版平台的用户测试研究 [J]. 情报资料工作，2017（6）：58-63.

[9] 秦顺，汪全莉，邢文明. 欧美科学数据开放存取出版平台服务调研及启示 [J]. 图书情报工作，2019（13）：129-136.

[10] LIU Y，GUO Q H，TIAN Y. A software framework for classification models of geographical data [J]. Computers & Geosciences，2012，42（9）：47-56.

[11] 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.

[12] HDF [EB/OL]. [2020-04-24]. https://baike.baidu.com/item/HDF/1256312?fr=aladdin.

[13] GEODOI [EB/OL]. [2020-04-24]. http://www.geodoi.ac.cn/WebCn/DocList.aspx.

[14] Issuance of OMB's "Final Information Quality Bulletin for Peer Review" [EB/OL]. [2020-04-24]. https://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf.

[15] 涂志芳. 科学数据出版的基础问题综述与关键问题识别 [J]. 图书馆，2018，285（6）：90-96，104.

[16] 涂志芳. 科学数据出版生态系统与质量控制体系构建 [J]. 图书与情报，2019，185（1）：131-140.

[17] 黄如花，李楠. 国外政府数据开放许可协议采用情况的调查与分析 [J]. 图书情报工作，2016（13）：5-12.

[18] NOAA Website Customer Satisfaction Survey [EB/OL]. [2020-04-24]. https://docs.google.com/forms/d/e/1FAIpQLScM7iupnUIPRBcjKSe7l7elk1roiNtv1hmiPxOKo1afqX1OMg/viewform.

[19] 邢文明，洪程. 开放为常态，不开放为例外——解读《科学数据管理办法》中的科学数据共享与利用 [J]. 图书馆论坛，2019，39（1）：117-124.

[20] 江洪，刘敬仪. 国外期刊科学数据管理调查与分析 [J]. 图书情报工作，2019，63（9）：127-134.

[21] 屈宝强，王凯. 数据出版视角下的科学数据同行评议 [J]. 图书馆杂志，2017，36（10）：71-77.

[22] USGS. tools [EB/OL]. [2020-04-24]. https://www.usgs.gov/products/data-and-tools/data-management/data-management-plans#Tools.

[23] 史雅莉. 科学数据引用标准实施的关键问题探析 [J]. 现代情报，2019，39（4）：35-42.

作者简介

张玲玲，女，1995年生，硕士研究生，通信作者，研究方向：信息分析与情报服务，E-mai：kh30zero@163.com。
陈媛媛，女，1982年生，博士，副教授，研究方向：信息分析与情报服务。

The Research on Geographic Science Data Publishing Platform in China and the United States

ZHANG LingLing[1] CHEN YuanYuan[1,2]
(1. School of Information Management Heilongjiang University, Harbin 150080, China;
2. School of Business Nankai University, Tianjin 300071, China)

Abstract: Data publishing is an important means to promote scientific data sharing. The investigation and research on the geosciences data publishing platforms in China and the United States can provide some development suggestions for the construction of scientific data publishing platforms in China. This paper selects eight representative geographic science data publishing platforms in China and the United States as the research objects, and conducts research from five basic links, including data submission, peer review, data release and permanent storage, data reference and impact evaluation. Suggestions for the development of geospatial data publishing platform in China are put forward as follows: formulating the "whole process" policy, establishing detailed submission process, improving data review system, establishing standard data release standards, safeguarding scientific data intellectual property rights, and expanding data quality feedback channels.

Keywords: Data Publishing; Scientific Data; Publishing Platform

（收稿日期：2020-09-28）