

·研究综述·

科学数据出版的基础问题综述与关键问题识别*

涂志芳

(中国科学院文献情报中心 北京 100190)

〔摘要〕在开放存取和数据开放的背景下,数据出版成为推动科学数据存储、开放、共享的重要途径,并且作为近年来相对新兴的议题得到国内外研究者的关注与研究。为清晰地了解国内外对科学数据出版的研究动态与进展,首先对数据出版概念、作用及数据出版模式等三个基础问题进行综述,并以此为依据判断、识别出目前数据出版研究与实践中的关键问题,对其中的数据标识、数据引用、数据评审等问题进行基本阐释,以期为后续的数据出版研究与实践提供方向性的参考和借鉴。

〔关键词〕科学数据 数据出版 数据标识 数据引用 数据评审

〔中图法分类号〕G250

〔引用本文格式〕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(6):86—92,100.

科学数据 (Scientific Data) 或研究数据 (Research Data) 是指产生或收集后有待进一步检查并作为推理、讨论或计算基础的信息,尤其是事实或数字信息,如统计数据、实验结果、测量结果、实地观察记录、调查结果、访谈记录和图像等^[1]。科学数据是学术资源开放存取的重要组成部分,也是大数据时代数据开放共享的重要内容,科学数据开放共享能够为社会带来巨大的利益和回报^[2]。在国内外学术社群对科学数据开放存取的阐释、声明或倡议下^[3-9],在各类组织机构数据开放政策的制定、实施中^[10-12],数据共享已然成为科技界、出版界、图书馆界等的普遍共识和行动,而数据出版 (Data Publishing) 是规范数据共享方式、提高数据共享质量的有效途径。

近年来,国内外关于数据出版的研究不断涌现,已有相当的文献对科学数据出版问题的多个方面进行了思考与探究;但数据出版实践仍有许多问题和困难需要进一步探索和克服。文章旨在对数据出版相关的基础研究问题进行系统性梳理和归纳,并在梳理过程中识别出数据出版现有研究与实践的关键问题与热点议题,了解研究与实践的重点、难点和局限,为未来研究方向提供参考和启示。

1 文献收集方法与过程

为了尽可能全面系统地收集与科学数据出版相关的文献,文章利用了多种工具和方式进行检索并在文献

收集过程中不断进行检索词的调整和细化,主要数据来源与获取途径包括:一是以中国知网 (CNKI)、维普、万方、百度学术等为数据来源,以“数据出版”“数据开放出版”“科学数据出版”“数据期刊”“数据论文”等为主题进行多次检索和人工筛选,获得高度相关的中文文献 30 余篇。二是以 Web of Science、Google Scholar、百度学术等为数据来源,以“data publishing”“scientific data publishing”“research data publishing”“data paper”“data journal”“data repository”“data review”“data citation”等为主题进行多次检索和人工筛选,获得高度相关的英文文献 40 余篇。三是利用 Google 作为搜索引擎,以“data publishing”“scientific data publishing”“research data publishing”等为检索词,补充性地访问了若干数据出版相关组织机构的官网、博客等内容并获得部分政策文件、年度报告等网络资料。四是在特定机构知识库中进行检索,如“中国科学院文献情报中心机构知识库”“北京大学机构知识库”,获得部分相关研究报告、演示报告等资料。

2 数据出版概念与作用

2.1 数据出版概念

在学术传播语境下,科学数据出版通常被认为是与科学研究产生的数据出版有关的活动^[13]。数据出版是科学数据与开放出版领域一个相对新兴的议题,相关研究对科学数据出版概念及其相关概念进行了释义

* 本文系北京市社会科学基金项目“学术图书馆参与数据出版质量控制研究”(项目编号:17XCB007)研究成果之一。

或辨析,主要角度包括:

首先,从出版语义的角度强调数据出版与一般的数据开放、共享的差别。数据“出版”与传统“出版”不同,数据的“publish”(小写字母p)和“Publish”(大写字母P)也有所区别,前者将数据文件上传于网络使数据得以开放,但数据文件没有任何长期存在、不被损坏、相关信息描述等的保障^[14];而后者是正式的数据出版,不仅可揭示数据集的科学质量和重要性,也能为数据生产者带来声誉,同时还意味着对数据长期保存的承诺和面向数据消费者的数据增值。

其次,从出版过程的角度阐释数据出版的功用。数据出版不是单纯将数据传至互联网进行共享,而是将数据上载到网络上使得除作者以外的人和组织也能够下载、分析、重用和引用数据^[15],是学术共同体中的学术期刊、学术机构或学术社群等主体从科学研究的角度对研究人员产生的科学数据及相关信息进行同行评审、编辑加工,使之符合一定规范和标准并能方便获取及再利用和引用的过程^[15-16]。

再者,对数据出版及其相关概念进行辨析以突出数据出版的特点。其中,数据发表^[17]是将数据发布在网络或其它媒介上,在遵循版权协议的前提下,供他人引用、下载、分析和重用,与数据出版含义基本相同;数据发布(Data Release)则侧重于将数据公布在互联网上,但缺乏永久保存、同行评审等环节。数据开放存取和数据开放共享强调数据面向一定范围公开并共享;数据开放出版则是在数据开放、共享的基础上,以“出版”的标准对开放数据进行规范处理、质量控制和版权保护,是开放存取模式的数据出版。广义的数据管理包含数据出版,狭义的数据管理是对数据进行收集、存档、保存、维护等的一系列管理性活动,而数据出版是在数据管理的基础上进一步规范形式、提升质量并明确知识产权的传播性活动。

可见,数据出版是使科学数据获得“来源可靠”“质量可信”“公开发布”“公共利用”“唯一标识”“知识产权清晰”“可正式引用”等特征的一系列活动和过程。

2.2 数据出版作用

数据出版的作用是有目共睹的,国内外研究者均高度肯定数据出版的积极作用。数据出版对研究人员、社会公众、数据本身、学术出版及科学事业的发展都能产生作用,具体而言:

一是作为一种全新的数据共享机制,数据出版能够协调数据的知识产权问题,保障科学数据创建者和

管理者相应的学术信誉与权益,从而提高研究人员数据共享的积极性^[16];二是数据出版便于作为数据用户的社会公众发现、获取、利用数据^[18],其中的同行评议环节是面向研究资助者和数据消费者的一种数据质量保证,数据出版还能够彰显科学研究过程的透明度并改善公众问责制度和环境^[18];三是数据出版能够确保数据集上传到受信任的存储库并得到正确存档和监护^[18],而且数据出版物与其他相关资源进行关联后形成完整的价值链也让数据变得更有价值^[19];四是数据出版是一种规范的数据开放共享方式,能够为学术评价、学术规范和学术质量控制提供基础,是科学数据管理及其可持续发展的必然结果和最终环节,是推动科学数据共享的最根本途径^[20]。

可见,数据开放出版不仅拥有数据开放、共享的全部功能,还能在此基础上极大解决数据知识产权的问题,提升研究者的数据共享意愿,保障科学数据的质量,提高数据重用的价值。

3 科学数据出版模式

3.1 数据出版模式

数据出版模式(Data Publishing Model),又被称为数据出版途径、形态、机制等,现有研究主要从出版物客体和出版物主体两个不同视角对数据出版模式进行划分和归纳。

就数据出版客体而言,学者们对数据出版/发表模式的划分有三种^[21]、四种^[22]、五种^[23]等不同说法,但以“三种模式”的划分居多。Lawrence等^[23]较早提出数据出版的“五种模式”,分别为独立数据出版、学术论文辅助数据出版、附录数据出版、期刊数据档案出版、数据论文出版,后来国内外研究者大多以此为基础进行数据出版模式划分的讨论。例如,王丹丹^[24]、顾立平^[25]、张小强和李欣^[13]、何琳和常颖聪^[26]等进一步将数据出版模式归为“独立的数据出版”“作为论文附件的数据出版”“数据论文出版”等三种,其中“数据论文出版”有时又被称为“数据集独立出版”^[27],而“作为论文附件的数据出版”有时又被称为“科学数据合作出版”^[28]。归根到底可总结为:独立数据出版,该模式将数据作为独立的信息对象提交到具有合作关系的开放存取数据知识库;作为论文附件的数据出版,数据往往作为论文的附件、附录而出版或存储到期刊指定的数据知识库并建立论文和数据的关联;数据论文出版,将科学数据作为一种文本文献进行出版,包括专门数据期刊的数据论文出版和综合性期刊的数据论文出版。

就数据出版主体而言,一方面从数据管理与出版动力视角可总结为三种模式:自上而下模式,通常由政府、资助机构等制定政策并提出强制性数据开放要求,相应的机构或人员遵照执行;横纵联合模式,数据在国际、区域、跨部门和部门内部之间进行数据多元合作开放;自下而上模式,多立足于激励措施的非强制性科学数据开放,如部分期刊、数据知识库的数据提交、存储与管理规定^[29]。另一方面,从数据管理与存储主体视角可分为四种模式:出版机构控制模式,即期刊出版单位控制数据和数据论文出版;非出版机构独立控制模式,主要由研究资助机构、教育科研机构等建立专门的数据仓储并进行数据出版;混合模式,由出版机构和数据存储机构分工合作进行数据出版;“二次出版”模式,基于海量数据信息和大规模数据出版而建立数据出版物的监护、索引、发现系统从而提高其可发现性和被利用的可能^[13]。

综上所述,主客体两种视角的划分具有相互映射的关系,同一视角下的划分有不同的说法,反映了当前人们对数据出版的认识还处于探索阶段,也说明数据出版实践具有复杂性。

3.2 模式一:独立的数据出版

“独立数据出版”模式是将数据提交、存储至开放存取数据知识库并提供公共利用,数据知识库(Data Repository)又被称为数据仓储、数据中心、数据平台等,是为有研究价值的数字对象提供长期管理的存档服务^[30],也指专门用于存储、管理、出版、利用开放数据的平台。科学、技术、医学等自然科学领域和社会学、经济学等人文社会科学领域的数据出版情况存在较大差别,数据知识库也通常被划分为通用型和学科型,研究中常用数据知识库案例包括 Dryad、Figshare、Dataverse、Zenodo、GenBank 等。

目前,学者们主要通过对国内外开放数据知识库进行调查分析的方式研究数据知识库的平台建设与数据出版。例如,Aalbersberg 等^[31]以国际知名出版机构 Elsevier 为例,介绍了作者贴标签、自动链接和数据链接应用程序等三种链接学术论文与对应数据知识库的方法;Assante 等^[32]选取 Zenodo、Dryad、CSIRODAP 等五个数据知识库为调研对象,从数据出版的数据集格式、文档、许可、出版成本、验证、可用性、发现和获取、引用等八个要素研究基于数据知识库的数据出版,并认为数据知识库还不能完全满足数据集资源的管理和使用需求。黄永文等^[33]选择 Figshare、Dryad、

PANGAEA 三个数据知识库进行学科领域、数据内容类型、数据唯一标识符、数据质量管理等方面的比较分析;朱玲等^[34]从需求与定位分析、平台建设实践、平台应用效果三方面剖析北京大学开放研究数据平台的建设过程与经验。

3.3 模式二:作为论文附件的数据出版

期刊出版商通过制定数据共享政策,从数据提交、数据审查、数据存储、数据服务、数据权益管理等方面对作为论文附件的数据出版加以控制和规范,以实现论文和数据关联出版,其中影响范围和影响力度较大的包括美国科学促进会(AAAS)、施普林格·自然出版集团(Springer Nature)、美国科学公共图书馆(PLOS)等的政策。国内外学者大多通过调查一种或多种学科领域期刊的数据政策内容以了解作为论文附件的数据出版情况:

一方面,对特定学科领域期刊的数据政策进行研究。例如,Zenk-Möltgen 等^[35]选取 140 种社会学期刊作为调研对象,调查发现较少期刊有明确的数据政策,大多数期刊参考其所属出版商对科学数据的相关规定;陈秀娟、吴鸣^[36]以美国化学学会出版的 49 种期刊为例,其中 47 种将科研数据作为支撑信息并对数据类型、数据存储、数据获取等加以规定;雷秋雨、马建玲^[37]从数据存储方式、数据格式、存储位置、数据质量控制、开放性与权益等角度对 JCR 收录的 46 种进化生物学核心期刊的数据政策进行分析。

另一方面,对多种学科领域的期刊数据政策进行综合研究。例如,Sturges 等^[38]调查分析了 371 种覆盖自然科学与人文社会科学领域的期刊数据政策,从存储内容、存储地点、存储时间、共享期望四个方面总结数据政策的主要内容,并为期刊数据政策设计了包含 21 个元素的框架模型;傅天珍、陈妙贞^[39]调查发现我国学术期刊数据出版政策数量较少且以鼓励性政策为主,政策学科分布不均衡并且政策内容各异;吴蓉等^[40]从数据提交政策、审查政策、权益政策三方面对 PLOS 的数据政策进行调研分析。

3.4 模式三:数据论文出版

数据论文(Data Paper)是一种新兴的出版物形态,与传统学术论文最大的不同在于数据论文重点描述科学数据本身而不是描述基于科学假设和科学问题的研究结果^[27],通常发表在专门的数据期刊(Data Journal)或综合期刊的相关板块。对数据论文的研究覆盖其概念、要素、作用、类型、形式、结构、同行评审、案例等议

题及其综述:

(1)数据论文概念与作用:数据论文是按照学术出版标准、以机构化且人可读的形式描述一个或一组特定数据集的元数据文档^[41]。作为可引用的期刊出版物,数据论文不仅揭示了数据的内容、价值、功用等关键信息,还能为数据作者、出版商带来相应的学术认可与声誉,同时在广泛传播的过程中进一步引起学术界对科学数据的关注。完整的数据论文出版包括数据类型、发布时间、共享规范、作者声誉、同行评审、错误纠正等要素^[27]。

(2)数据论文形式与内容:数据论文形式与名称多样,如数据文章、数据集论文、数据原创文章、数据库论文、数据注释^[42]、简要数据、数据说明^[43]等;数据论文内容应尽可能详细地描述数据集信息,即“是什么”“在哪里”“为什么”“怎么样”“谁”(“4W1H”)等问题,具体包括数据来源、处理方法、覆盖范围、数据格式、数据内容、数据质量、数据重用、数据许可、作者贡献等要素^[14]。

(3)数据论文出版实践案例与综合议题:关于数据论文出版案例,如英国全国环境研究委员会(NERC)环境科学数据中心将对数据集的描述作为数据论文进行出版,并为数据论文分配数字对象唯一标识符(DOI)^[14];关于数据论文出版综合议题,如包括数据论文的概念及相关概念、出版流程、标准要求,研究者、出版商、研究与资助机构等利益相关者及权益机制,同行评审方式与内容等^[43-44]。

4 数据出版的关键问题与热点议题

在文献收集与阅读的基础上,在对数据出版概念、作用及模式等基础问题进行综述的过程中发现,完整的数据出版具有复杂性,需要政策、标准、技术、平台、工具、人员、资金等多个要素的协调与配合。如前所述,较之于一般的数据开放共享,数据出版有着“来源可靠”“质量可信”“唯一标识”“知识产权清晰”“可正式引用”等特点,能够在数据开放共享的基础上进一步规范数据引用、明确数据知识产权、减少数据共享障碍,从而提高研究者的数据共享意愿和数据重用价值。因此,从数据出版生命周期考虑,笔者认为数据标识、数据引用(数据知识产权)、数据评审(数据质量控制)等是使数据出版区别于一般数据共享的几个关键问题,也是关乎数据出版可持续发展的热点议题。

4.1 数据标识

长久以来,数据存储基础设施的缺乏使得独立出版和引用数据非常困难,因此科学数据也一直未曾得到和科学论文同等的学术认可^[45],数据标识符在这样

的需求下应运而生。数字对象标识符(Digital Object Identifier, DOI)、统一资源名称(URN)、开放链接(OpenURL)、句柄系统(Handles)等是目前应用较多的数字标识符,其中尤以DOI的应用和研究最为广泛。

DOI是用于识别数字环境下对象的知识产权的字符串^[46],自1998年成立且于2012年成为国际标准化组织(International Standard Organization, ISO)“信息与文献”领域的一项标准以来,广泛应用于数字化图书、期刊、数据等内容类型的学术出版。DOI具有唯一且永久标识、永久定位、出版源头注册、点击即链接、动态更新、版本更迭^[47]等特点,在科学数据出版中具有跟踪、引用、集成、关联等的多重价值,便于实现对数据出版物的原文获取、引文链接、数字版权管理及永久标识等功能,解决数据多重链接和知识产权问题^[48]。

数据出版所需DOI由注册代理机构及其成员单位负责分配,目前共有覆盖欧盟、中国、日本、韩国等地域范围的10个代理机构(截至2017年8月16日),如欧盟出版物办公室(Publications Office of the European Union)、中国知网(CNKI)、韩国科技信息研究所(KISTI)、Crossref、DataCite^[49]等。现有文献大多聚焦于对DataCite进行研究,如DataCite的发起者与发展历史^[50]、DataCite成员单位及其地域覆盖^[51]、DataCite数据出版元数据方案^[52]等。

4.2 数据引用

数据引用是数据出版的关键环节之一,是保障数据作者与管理者数据权益的一种有效方式,引起了广泛的研究与关注。完整的数据引用涉及“如何引用”“引用什么”“何时引用”^[25]等环节,包含对数据引用标准、数据引用技术、数据引用行为等关键内容的讨论。

目前,国内外研究数据引用尚无一致的标准,不同的数据知识库、数据论文通常有不同的数据引用标准并相应推荐不同的数据引用格式。其中,哈佛大学定量社会研究中心的Micah Altman和Gary King提出的“定量数据学术引用的建议标准”(2007年)^[53]、经济合作与发展组织(OECD)发布的白皮书“我们需要数据集和数据表的出版标准”(2009年)^[54]、FORCE11(The Future of Research Communications and e-Scholarship)联合其他多个数据引用工作组发布的“数据引用原则联合声明”(2014年)^[55]、DataCite发布的“研究数据出版与引用元数据方案(3.1版本)”(2015年)^[56]等是认可度相对广泛的标准和原则,在数据引用实践中起到重要的参考和指导作用。

数据引用技术是实现数据规范引用的技术基础,前文所述的 DOI 便是目前实现数据引用而广泛采用的一种技术,另外数据指纹(Universal Numeric Fingerprints, UNF)也是数据引用的一种有效技术。其中,DOI 所具有的唯一标识、永久定位、动态更新的特点,一方面使得数据引用能够实现数据版本的动态更迭、数据地址的动态更新,不必频繁更改地址便能实现“从一而终”的资源指示与数据引用;另一方面,DOI 有助于维护数据创建者作为一项必不可少的元数据元素,保障作者数据署名权并承认作者的贡献。UNF 是一种加密的数字签名,可用于数据集及其子集(版本)的唯一标识,数据本身的元素不同 UNF 也不同,UNF 可验证不同文献是否引用了同一个数据,也可与 DOI 一起使用形成对版本化数据集的持续引用^[57]。

国内外图书情报机构或数据中心已经开始了数据引用的实践,而学术界也开始了对数据引用行为的研究。对施引文献及其作者而言,引用数据是为了说明数据来源并达到一定的学术规范,数据引用情况取决于数据用途和数据来源,数据用途即数据是否能够解释问题、数据是否有效支撑论文结果或结论、数据是否作为论文补充材料等,而数据来源指数据是否作为论文附件、数据是否存储在数字仓储、数据是否以数据论形式出版等。

从文献研究现状看,数据引用标准、数据引用技术和数据引用行为等话题是当下数据出版领域的热点问题,也是未来研究和实践需要进一步突破的难点问题。

4.3 数据评审

数据评审是一种重要而有效的数据出版质量控制方法,主要指采取必要措施对数据及文档、代码等相关文件进行评估,确保文件可被独立地理解以便重用的有效过程^[58]。与学术论文出版过程中的同行评审环节相似,数据出版也将数据评审作为数据质量控制的重要一环;数据评审至今尚未有相对统一的程序,甚至数据能否被“评审”以及“数据评审”这一术语本身也存在疑问^[59],但数据评审仍在研究和实践中取得了一定的进展^[58]。完整的数据评审包含“评审什么”“怎么评审”“谁来评审”等组成部分,涉及对数据评审内容与维度、数据评审方法与程序、数据评审主体等话题及其综合议题的探讨。

结合对数据评审含义的理解,数据评审包含较为广泛的内容范畴及评估流程,是一项复杂的活动,如从数据计算与分析结果的准确性的双盲评审、数据文件

与文档的充分性的同行评审到元数据的准确性的技术性审查^[60]。数据评审以数据质量及其标准为依据,数据评审内容包括研究计划评审、数据管理计划评审、元数据评审、数据附件评审、数据知识库数据评审、数据论文评审等,这些内容可以归纳为“数据内容”“数据形式与元数据”“语境信息和/或其他相关文档”三个类别^[14],分别对应科学性评审(Scientific Review)、技术性评审(Technical Review)和监护性评审(Curatorial Review)三个维度^[60],但在数据出版实践中数据评审通常融合了三个维度的内容与方法,并对数据进行综合的评审和全面的质量控制。

其中,科学性评审是对数据内容的各类科学性特征进行评审,以保障数据的科学性质量(Scientific Quality),表现为完整性、真实性、准确性、可靠性、有用性等特征,通常由同行专家或高级编辑进行评审。技术性评审是对数据形式及元数据相关的各类技术性特点与细节进行评审,以保证数据的技术性质量(Technical Quality),表现为数据格式的标准化、兼容性及元数据的完整性、准确性、真实性等特征,一般而言论文中的数据由编辑或同行专家进行评审,而存入到数据库的数据(集)由学科领域内的数据专家或专门数据管理人员进行评审。监护性评审是对数据内容、数据格式、元数据以外的上下文/语境信息、数据相关文档、代码等进行管理性的检查和审核,评审内容及其标准包括但不限于文档的相关性、充分性,数据来源、数据环境、数据收集/处理/分析方法等文档的完备性,文档版本控制及相关文档信息更迭的及时性,数据与其他文档关系的协调性(如对相关政策、标准的遵循)等。

当然,数据标识、数据引用、数据评审领域的研究与实践都已经进行,但作为数据出版的关键问题与热点议题,对数据标识、数据引用、数据评审的探索研究在可预见的未来还将继续深入而广泛地进行。

5 科学数据出版研究特点与未来展望

5.1 数据出版困难与挑战

尽管数据出版研究已经较为全面地展开,而数据出版实践也已经在国内外大范围内进行,但作为学术交流、传播与出版领域相对新兴的议题,数据出版在标准、规范、技术、政策、机制等方面仍存在不少困难和问题需要克服。例如,Hense 和 Quadt^[61]认为数据出版面临的挑战存在于格式、内容、科学质量控制、格式质量控制、技术质量控制、浏览与检索、存储等多方面;张

小强、李欣^[13]认为当前数据出版的主要障碍是伦理规范与标准的缺失,包括数据出版体系中奖惩功能的缺失、数据出版复杂性导致的短期内标准规范的缺失,数据出版用稿规范、数据引用规范、元数据标准等的不完善;何琳、常颖聪^[26]认为科学数据的元数据、隐私保护、组织技术、知识产权保护、出版激励机制是数据出版待解决的关键问题;张丽丽^[62]认为,相较于传统学术出版,数据出版的学术地位和行业认可度、运营模式的可持续性、数据开放与版权保护等方面仍有待发展。

5.2 数据出版现有特点总结与未来研究展望

综合而言,国外数据出版理论与实践工作齐头并进,国内相关研究经历了翻译推介、案例分析、原创性思考探索等渐进过程,研究先于实践而进行但目前已经开始了数据出版的实践探索。具体而言,近年来国内外科学数据开放出版研究呈现以下特点:

首先,概念理解与辨析是基本问题,对数据出版概念及其相关概念的理解与辨析是研究数据出版的首要基本问题,也是突出数据出版作用的必要支撑条件。其次,不同出版模式的划分、对比是研究重点,大部分研究都对数据出版模式进行划分和比较以便尽可能全面而清晰地呈现数据出版的真实形态,同时也便于分门别类地对数据出版相关的其他问题进行阐述。最后,对数据创建者、管理者、使用者对数据出版的认知、态度与需求等进行调查有其必要性,也可作为数据出版实践方向的一种事实依据和导向。再者,数据出版相关标准、规范的缺失或不完善是急需研究和亟待解决的问题:数据出版需尽可能达到数据格式、数据类型、数据伦理等标准规范的统一或兼容,标准、规范是目前较为薄弱的环节,也是需要克服的问题与困难。

综上所述,笔者认为未来可能的研究趋势与展望包括:第一,由于目前数据出版领域存在的标准、规范、技术等的不完善问题,数据出版标准、规范及其技术实现将成为未来研究与实践的一大重点;第二,在大数据时代背景下,随着科学数据开放广度与深度的发展,数据数量呈倍数增长使得数据质量控制的重要性日益凸显,数据质量控制是未来可能的又一研究重点;第三,目前研究与实践中存在的局限也意味着国内外图书情报机构在数据出版标准、规范的制定与实施及数据质量控制等方面有较大的发挥空间和发展机会。

(来稿时间:2017年9月)

参考文献:

1. European Commission. Horizon 2020[EB/OL].[2017-01-

01].http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

2. OPEN RESEARCH DATA IN HORIZON 2020[EB/OL].[2017-01-01]. http://ec.europa.eu/research/press/2016/pdf/opa-data-infographic_072016.pdf#view=fit&pagemode=none.

3. Open Access at the Max Planck Society. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities[EB/OL].[2017-06-20].<https://openaccess.mpg.de/Berlin-Declaration>.

4. IFLA. IFLA Statement on Open Access to Scholarly Literature and Research Documentation[EB/OL].[2017-06-20]. <http://www.ifla.org/publications/ifla-statement-on-open-access-to-scholarly-literature-and-research-documentation>.

5. IFLA. IFLA Statement on open access—clarifying IFLA's position and strategy[EB/OL].[2017-06-20].<http://www.ifla.org/files/assets/hq/news/documents/ifla-statement-on-open-access.pdf>.

6. UNESCO. Policy Guidelines for the Development and Promotion of Open Access[EB/OL].[2017-06-20].<http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>.

7. European Commission. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020[EB/OL].[2017-06-21].http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

8. NIH. Final NIH Statement on Sharing Research Data[EB/OL].[2017-06-21]. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

9. OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding[EB/OL].[2017-06-21].<http://www.oecd.org/science/sci-tech/38500813.pdf>.

10. NSF. NSF'S Public Access Plan: Today's Data, Tomorrow's Discoveries[EB/OL].[2017-06-22]. <https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>.

11. University of Cambridge. Research Data Management Policy Framework[EB/OL].[2017-06-22].<http://www.data.cam.ac.uk/university-policy>.

12. The University of Hong Kong. Research Data and Records Management[EB/OL].[2017-06-22].<http://www.rss.hku.hk/integrity/research-data-records-management>.

13. 张小强,李欣.数据出版理论与实践关键问题[J].中国科技期刊研究,2015,26(8):813-821.

14. Callaghan S, Donegan S, Pepler S, et al. Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres[J].International Journal of Digital Curation, 2012, 7(1):107-113.

15. Penev L, Chavan V, Hagedorn G, et al. Pensoft Data Publishing Policies and Guidelines for Biodiversity Data[EB/OL].[2017-08-02].http://pensoft.net/J_FILES/Pensoft_Data_

Publishing_Policies_and_Guidelines.pdf.

16. 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.
17. 刘凤红, 崔金钟, 韩芳桥, 等. 数据论文: 大数据时代新兴学术论文出版类型探讨[J]. 中国科技期刊研究, 2014, 25(12):1451-1456.
18. Murphy F. Transforming Research Communication: Knowledge Management and Data Publishing[EB/OL].[2017-01-09]. <https://ams.confex.com/ams/93Annual/webprogram/Paper217112.html>.
19. Borgman C L. Data, disciplines, and scholarly publishing[J]. Learned Publishing, 2008, 21(1):29-38.
20. 袁曦临. E-science 环境下学术规范的新领域: 科学数据[J]. 甘肃社会科学, 2014(3):85-88.
21. 孔丽华. 科学数据质量同行评议现状[R/OL]/2015 中国开放获取推介周.[2017-08-14]. <http://ir.las.ac.cn/handle/12502/7968>.
22. 张静蓓, 任树怀. 科研数据出版模式、流程及引用策略研究[J]. 图书情报工作, 2015(9):21-27.
23. Lawrence B, Jones C, Matthews B, et al. Citation and Peer Review of Data: Moving Towards Formal Data Publication[J]. International Journal of Digital Curation, 2011, 6(2):4-37.
24. 王丹丹. 科学数据出版过程中的数据质量控制[J]. 图书情报工作, 2015(23):124-129.
25. 顾立平. 科学数据权益分析的基本框架[J]. 图书情报知识, 2014(1):34-51.
26. 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, 58(5):104-110.
27. 王丹丹. 数据论文: 数据集独立出版与共享模式研究[J]. 情报资料工作, 2015, 36(5):94-97.
28. 梁子裕. 国外科研数据出版模式研究[J]. 出版发行研究, 2017(3):82-85.
29. 张丽丽, 黎建辉. 科研数据的开放: 进展、模式与新探索[J]. 大数据, 2016, 2(6):25-33.
30. Research Data Canada. Data repository[EB/OL].[2017-08-14]. <https://www.rdc-drc.ca/glossary/original-rdc-glossary/>.
31. Aalbersberg I J, Dunham J, Koers H. Connecting Scientific Articles with Research Data: New Directions in Online Scholarly Publishing[J]. Data Science Journal, 2013(12):235-242.
32. Candela L, Assante M, Tani A, et al. Are Scientific Data Repositories Coping with Research Data Publishing?[J]. Data Science Journal, 2016, 15(6):1-24.
33. 黄永文, 张建勇, 黄金霞, 等. 国外开放科学数据研究综述[J]. 现代图书情报技术, 2013(5):21-27.
34. 朱玲, 聂华, 崔海媛, 等. 北京大学开放研究数据平台建设: 探索与实践[J]. 图书情报工作, 2016, 60(4):44-51.
35. Zenk-Möltgen W, Lepthien G. Data sharing in sociology journals[J]. Online Information Review, 2014, 38(6):709-722.
36. 陈秀娟, 吴鸣. 学科领域期刊科研数据发表政策剖析——以美国化学学会期刊为例[J]. 中国科技期刊研究, 2015, 26(8):800-807.
37. 雷秋雨, 马建玲. 学术期刊数据出版政策研究综述——以 JCR 中进化生物学领域期刊为例[J]. 图书馆理论与实践, 2016(1):30-34.
38. Sturges P, Bamkin M, Anders, Jane H.S, et al. Research data sharing: Developing a stakeholder-driven model for journal policies[J]. Journal of the Association for Information Science & Technology, 2016, 66(12):2445-2455.
39. 傅天珍, 陈妙贞. 我国学术期刊数据出版政策分析及建议[J]. 中国出版, 2014(23):31-34.
40. 吴蓉, 顾立平, 刘晶晶. 国外学术期刊数据政策的调研与分析[J]. 图书情报工作, 2015, 59(7):99-105.
41. Chavan V, Penev L. The data paper: a mechanism to incentivize data publishing in biodiversity science[J]. BMC Bioinformatics, 2011, 12(15):2399-2405.
42. 屈宝强, 王凯. 数据论文的出现与发展[J]. 图书与情报, 2015(5):1-8.
43. 刘传玺. 数据论文概念辨析及其同行评审研究[J]. 图书馆杂志, 2016(9):76-80, 93.
44. 伏安娜, 张计龙, 殷沈琴. 数据论文国内外发展研究综述[J]. 图书情报工作, 2015(24):131-138.
45. Neumann J, Brase J. DataCite and DOI names for research data[J]. Journal of Computer-Aided Molecular Design, 2014, 28(10):1035-1041.
46. Chandrakar R. Digital object identifier system: an overview[J]. The Electronic Library, 2006, 24(4):445-452.
47. 涂勇, 彭洁. 数字对象唯一标识在中国科学数据领域中的应用研究[J]. 数字图书馆论坛, 2013(8):31-36.
48. 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.
49. DOI Registration Agencies[EB/OL].[2017-08-16]. http://www.doi.org/registration_agencies.html.
50. Brase J, Lautenschlager M, Sens I. The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite[J]. D-Lib Magazine, 2015.
51. Loesch F M. DataCite[J]. Technical Services Quarterly, 2016(33):91-92.
52. Starr J, Ammann N, Ashton J, et al. DataCite Metadata Schema for the Publication and Citation of Research Data[EB/OL].[2017-08-16]. http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf.
53. Micah Altman, Gary King. A Proposed Standard for the Scholarly Citation of Quantitative Data [EB/OL].[2017-09-18]. <http://www.dlib.org/dlib/march07/altman/03altman.html>.

(下转 100 页)

49. Mun Y. Yi, Yujung Hwang. Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model [J]. *Human-Computer Studies*, 2003, 59: 431-449.

50. 赵杨, 高婷. 移动图书馆 APP 用户持续使用影响因素实证研究 [J]. *情报科学*, 2015 (6): 95-100.

51. 陈明红, 漆贤军, 刘莹. 移动图书馆持续使用意向及习惯的调节作用 [J]. *情报科学*, 2016 (6): 125-132.

Review on Factors Affecting of the Adoption Behavior of Mobile Library Users Based on the TAM

Zou Kai Gao Kai Liu Zhao

(School of Public Administration, Xiangtan University)

〔Abstract〕 This paper summarizes the literature of the adoption behavioral in domestic and foreign mobile libraries, focusing on the empirical research design, the expansion of TAM model, the core variables, the pre-external variables, the adjustment variables and so on. and puts forward the following five suggestion: subsequent studies must be dynamically compared to study user adoption behavior, data obtainment diversified, increased sample size, studies according to the characteristics of the sample and model expansion, so as to carry out the research of facing mobile library users from different angles and multiple levels.

〔Keywords〕 Mobile library User adoption Technology acceptance Influencing factors

〔作者简介〕 邹凯 (1965—), 男, 博士, 湘潭大学公共管理学院教授、博士生导师, 研究方向: 信息用户与服务; 高凯 (1992—), 男, 湘潭大学公共管理学院 2016 级情报学硕士研究生, 研究方向: 数字图书馆; 刘钊 (1980—), 男, 硕士, 湘潭大学公共管理学院讲师, 研究方向: 数字图书馆。

(上接 92 页)

54. Green T. We Need Publishing Standards for Datasets and Data Tables [EB/OL]. [2017-09-13]. <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2010/wp.8.e.pdf>.

55. Data Citation Synthesis Group. Joint Declaration of Data Citation Principles [EB/OL]. [2017-04-18]. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>.

56. DataCite Metadata Working Group. DataCite Metadata Schema for the Publication and Citation of Research Data [EB/OL]. [2017-09-18]. http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf.

57. Thomas Leeper, Micah Altman. UNF: Tools for Creating Universal Numeric Fingerprints for data [EB/OL]. [2017-09-20]. <https://cran.r-project.org/web/packages/UNF/index.html>.

58. Peer L, Green A. Research Data Review is Gaining Ground [EB/OL]. [2017-08-12]. <http://isps.yale.edu/news/blog/2015/03/research-data-review-is-gaining-ground>.

59. Shaklee P, Cousijn H. Can data be peer-reviewed? [EB/OL]. [2017-08-12]. <https://www.elsevier.com/connect/can-data-be-peer-reviewed>.

60. Austin C C, Bloom T, Dallmeier-Tiessen S, et al. Key components of data publishing: using current best practices to develop a reference model for data publishing [J]. *International Journal on Digital Libraries*, 2010: 1-16.

61. Hense A, Quadt F. Acquiring High Quality Research Data [J]. *D-Lib Magazine*, 2011, 17 (1-2): 1-7.

62. 张丽丽. 科学数据共享治理: 模式选择与情景分析 [J]. *中国图书馆学报*, 2017, 43 (2): 54-65.

A Review of Fundamental Research and Identification of Key Issues on Scientific Data Publishing

Tu Zhifang

(National Science Library, Chinese Academy of Sciences)

〔Abstract〕 Under the background of open access and open data, open access publishing is becoming a significant way of scientific data storage, opening and sharing, and as a relatively new issue in recent years it gets considerable focus and research from researchers at home and abroad. In order to understand the research progress and trends of scientific data publishing at home and abroad, this paper makes a review on three fundamental problems, they're data publishing concepts, data publishing functions and data publishing models. Accordingly, this paper evaluates and identifies several key issues about data publishing research and practice, they're data identification, data citation and data review. Eventually, this paper hopes to provide indicative and valuable references for further studies.

〔Keywords〕 Scientific data Data publishing Data identification Data citation Data review

〔作者简介〕 涂志芳 (1993—), 女, 硕士, 中国科学院文献情报中心助理馆员, 研究方向: 信息资源建设、图书馆管理。