

《图书馆杂志》的科研数据出版实践与思考

谈洁¹ 胡杰²

(1. 上海图书馆(上海科学技术情报研究所), 上海 200040; 2. 复旦大学图书馆, 上海 200433)

摘要: 在科学研究新范式的驱动下,《图书馆杂志》以提供高质量科研数据为核心,借助自建数据出版管理平台,结合传统学术出版流程,构建全新出版模式及流程,可为我国其他传统学术期刊出版服务模式变革提供借鉴。综合国内数据出版的研究现状及期刊数据出版管理服务模式,提出我国期刊数据出版服务需从建立期刊数据出版管理联盟、提高科研人员数据管理能力、完善数据出版制度、开发智慧型数据出版管理平台等路径出发,积极参与科研数据出版部署。

关键词: 数据出版; 科研数据; 数据管理; 数据共享; 数据服务

中图分类号: G312 **DOI:** 10.3772/j.issn.1673-2286.2021.12.005

引文格式: 谈洁, 胡杰. 《图书馆杂志》的科研数据出版实践与思考[J]. 数字图书馆论坛, 2021 (12): 28-33.

随着信息社会、大数据的不断发展,科研迈入数据驱动型、数据密集型研究范式,科研数据因此变成学术成果的重要组成部分。学术期刊是学术成果发布的载体,肩负着审核科研内容、传播学术成果的责任与义务。在科研数据已成为学术成果重要组成部分的当下,学术期刊有必要将学术论文相关的科研数据纳入出版环节。国际五大出版商Elsevier^[1]、Springer Nature^[2]、Wiley^[3]、SAGE Publishing^[4]和Taylor & Francis^[5]均对旗下期刊制定了相应的数据出版政策,鼓励或者要求作者提交论文的支撑数据。我国自2014年全球变化科学研究数据出版系统(Global Change Research Data Publishing & Repository, GCdataPR)^[6]正式上线,以《中国科学数据》《中国地质》等为代表的主要刊发数据论文的“纯数据”期刊陆续出现。《图书馆杂志》是图书情报类传统学术期刊,为适应科研数据出版的时代需求,将已有的论文投稿系统与新建的数据出版管理平台结合,迈出我国传统学术期刊介入科研数据出版的重要一步。对《图书馆杂志》数据出版实践经验的介绍和深入分析,在一定程度上可以为其他传统学术期刊的变革提供可借鉴的经验,推动我国科研数据出版的发展。

1 科研数据出版的意义

学界对于数据出版有许多定义^[7-9],本文以学术期刊视角讨论的科研数据出版,是指以提供高质量科研数据为核心,按照统一规范的质量管理和控制机制对科研数据(包括科研原始数据集、收集整理加工数据,以及图片、音频、视频等)进行出版,帮助使用者便捷地发现、获取、理解和重用数据,并可在科研论文及相关研究成果中引用数据。简单来说,就是基于数据审核等管理措施,使数据达到可信任、可引用、可追溯的状态。因此,科研数据出版是建立在数据管理的基础上,而不仅仅是简单的数据发布:经过数据审核等规范性的学术操作,提供具有标准化、结构化的数据格式,该数据经发布后,期刊提供唯一标识符供使用者永久访问、下载并再利用。

对学术论文中的数据资源进行科学的管理与出版,是支撑学术研究发挥其科研价值及社会价值的有力举措,也是推动学科健康长远发展的重要手段。期刊通过数据管理来实现学术论文和科研数据的实时关联,可以有效打破信息壁垒、提高论文质量,对扩大期刊的学术影响力、促进国内外学术交流有重要的意义。数据出版也是促进数据共享的有力手段^[10],通过数据

共享可以提高学术研究的真实性和可靠性,并为公众提供更多的信息资源,推动社会进步。我国是期刊大国,但非强国,国内期刊在开放获取理念和对科研数据的重视程度上相对国际期刊有较大差距^[11]。我国学界对科研数据出版与管理的理论研究较多,但以期刊机构为主体对科研数据出版的实践研究依旧是一个较新的领域^[12],传统学术期刊在科研数据出版方面缺乏经验支持,亟须探索与发展。

2 《图书馆杂志》的科研数据出版实践

近年来,随着科学研究规范性的普及,基于科研数据的学术论文投稿已成为期刊收稿常态,由此引发两方面问题亟需思考与解决:一是论文中科研数据的可靠性、真实性、可重复性该如何保障的问题;二是数据生产的成本高,被复制却非常容易,那么数据生产者的劳动成果该如何保护、数据权益该如何保障及作者的创造积极性该如何维护的问题。基于对这些问题的思考,《图书馆杂志》于2016年对学术出版全流程服务模式进行了变革,在维持原有传统学术出版模式的前提下,拓展数据出版新模式,并将两者融合在统一的学术出版流程中。

2.1 新出版流程

通过不断实践,《图书馆杂志》构建了兼容数据出版的全新的出版流程。作者通过《图书馆杂志》官网^[13]的投稿系统在线提交论文,在向导式投稿流程中,嵌入询问作者论文中是否含有数据的步骤,若投稿论文中

涉及科研数据、音频、视频、图片等,则跳转至图书馆杂志数据出版平台^[14]。作者须提供相关信息并上传数据,将数据存储的网址或handle号(数字资源全球唯一标识符)置于论文文末,供数据编辑审核。数据评审环节包括责任编辑初审、数据编辑初审、专家外审、同行评议、责编复审等,通过层层把关,确保发布数据的准确性、可靠性。审核通过后在论文文末添加handle号,通过handle号论文可以直接引证该数据信息。

2.2 新出版工具——数据出版管理平台

2016年《图书馆杂志》联合复旦大学大数据研究院人文社科数据研究所,参考国际元数据标准规范DDI、DataCite的数据引证规范创建了数据出版管理平台。平台基于全球数据管理领域先进的B/S架构,后台使用Java语言,前台使用HTML5、CSS3、ES5/6等语言,以NODEJS作为代理并通过AJAX技术与后台通信。数据库选用POSTGRESQL,有清晰的数据存储层、业务逻辑层、应用层和人机交互层等逻辑结构,具有数据管理、使用、展现、搜索等功能,支持未来接入大数据服务引擎。

数据出版管理平台为科研数据出版提供全生命周期的功能服务,包括数据描述、数据提交、数据评审、数据发布、长期保存、数据引证、影响评价、数据共享、在线分析、版本更新,具体内容见表1。平台支持多种格式数据的导航、上传、下载,数据的基本检索和高级检索,基于元数据的分面检索,并针对数据访问设置了3种不同程度的权限管理(任何人可以访问、授权用户可以访问或者完全保密)。

表1 数据出版管理平台功能描述

序 号	平台功能	详细描述
1	数据描述	采用国际标准的元数据规范,实现数据的长期保存和访问传播
2	数据提交	提供多种数据文件格式的上传
3	数据评审	与论文投稿系统融合,可以进行数据论文和数据集的评审
4	数据发布	数据集被赋予全球唯一标识符,正式发布,用户可以进行关键词检索和分类浏览
5	长期保存	数据能够自动转为适合长期保存的非专有格式存储
6	数据引证	采用全球唯一的句柄系统标识数据,基于数据唯一标识符进行数据出版,使数据能够被公开出版物引用
7	影响评价	用户能够对数据集进行打分和评价
8	数据共享	数据能够通过权限控制,分享给不同层级的用户
9	在线分析	提供多种高级统计分析模型,多层次、多角度展现数据,实现数据可视化
10	版本更新	数据集可以进行在线的版本更新和发布,数据引证的链接会自动导航到最新版本

2.3 新数据出版模式

《图书馆杂志》在综合国内外数据出版研究与实践的基础上,根据自身条件设计了4种数据出版模式。

(1)“数据论文”模式。开设“数据论文”栏目,发表学科中价值高、质量优、可机读、有挖掘价值的数据集,为研究者提供一手的研究数据资料,通过“数据论文”栏目进行“纯数据”出版,充分共享数据,从而发挥数据潜在的能量与价值,促进相关学科领域的理论创新与发展。为了使用者能更好地读懂、理解发布者的数

据,要求发布者对数据集做一个详细、明确的描述。针对这个要求,《图书馆杂志》制定了数据论文模板规范性文档,即“数据论文”。数据论文模块主要包括数据集基本信息简介、引言、数据采集和处理方法、数据字典和数据样本、数据质量控制、数据价值、数据使用方法和建议、数据及论文引用格式、参考文献、作者信息及分工职责,各模块的具体内容见表2。期刊对已录用数据论文进行出版发行,而数据与数据论文则通过handle号永久互联,形成数据论文出版。

表2 《图书馆杂志》数据论文具体内容

序 号	数据论文模块	模块说明
1	数据集基本信息简介	以表格形式列明数据集中英文名称、作者信息、数据集版本、数据格式、数据体量、全球唯一标识符、数据获取网址、数据集组成和数据使用条款等其他内容
2	引言	说明数据集相关的背景资料及其潜在的利用价值与意义
3	数据采集和处理方法	详细描述数据采集的方法步骤,数据处理过程中涉及的方法、程序、算法与分析等
4	数据字典和数据样本	揭示数据集的典型样本,提供数据字典,描述数据结构、数据字段等,帮助读者理解数据集的内涵
5	数据质量控制	提供资料证明数据的可靠性,包括数据质量控制方法和验证过程,列明误差数据和相关精度。另外,对于连续性数据中的数据缺失、断点等情况进行必要说明
6	数据价值	通过和国内外相关数据集的比较,体现数据来源、数据整理加工过程、质量控制等方面的创新性,从数据覆盖范围、数据加工方法、潜在应用领域等角度说明本数据集的价值
7	数据使用方法和建议	可包含支持数据利用的推荐软件说明,数据的使用方法与步骤等。对于涉及人类隐私或公众安全等问题的敏感数据,应详细描述相关使用限制
8	数据及论文引用格式	作为已发表成果,提供给读者参考的数据集及数据论文可被引用的参考文献格式
9	参考文献	该篇数据论文撰写过程中引用已发表的文献列表
10	作者信息及分工职责	简要描述每个作者信息及对数据产出成果的贡献

(2)“论文+支撑数据”模式。如果文章论述、研究的内容有科学研究数据支撑,如问卷调查信息、软件使用代码、后端输出数据等,期刊编辑在审稿过程中会要求作者将数据上传至数据出版管理平台。该模式下,作者仅需将数据上传后生成的handle号附在文末,便可完成论文与数据的关联。期刊编辑、外审专家等基于论文支撑数据可以更准确地判断论文结构是否合理、实验设计是否恰当、数据分析是否可靠等,达到有理有据地评判论文质量的目的。

(3)“论文+富媒体数据”模式。图书情报学的学科属性决定了其投稿论文中可能涉及一些价值较高的图片、音频、视频等富媒体格式文件,因此在实践中产生了“论文+富媒体数据”模式。作者可以配合上传高清的富媒体文件至数据出版管理平台,以供研究者参考鉴别。在这种模式下,作者需将handle号附在文末,实现

富媒体数据与论文的关联;期刊针对富媒体数据制作放映二维码,在排版中直接与论文标题关联,方便读者扫码读取丰富的数据内容。相较一些“纯数据”期刊单纯发布图片等到数据仓储平台或者传统学术期刊仅出版纸本论文,这种配合学术论文而发布富媒体文件数据的形式更容易引发学术争鸣,促进学术交流与进步。

(4)“仓储数据”模式。鼓励作者将已经发表于他刊或本刊的研究论文中的科研数据、数据集进行回溯,即这些数据可另行投稿至数据出版管理平台,经审核后作为研究数据成果发布,并在数据出版平台上与原论文作相应关联。实践中,遇到作者需要将研究论文投稿至国外期刊,国外期刊需要审核论文中的数据,为了避免核心数据版权外流,作者申请将科研数据发布于《图书馆杂志》数据出版管理平台,生成handle号后提供给国外期刊审稿。综上情况,可归纳为数据出版管理

的第4个出版模式即“仓储模式”,利用该模式可以实现对科研数据的再挖掘与二次利用,促进本专业领域内论文数据的共享,不断创造数据价值。

2.4 新数据出版成果

一方面是基于数据集的研究性论文的产出。自2016年《图书馆杂志》进行数据出版管理实践开始,截至2021年10月底,共发布数据课题94个、文件688个,数据下载量84 145次。其中数据论文的相关数据下载量62 296次,下载量最高的是“复旦大学师生中文电子期刊资源访问行为数据集”系列,基于该数据集产出了一系列研究论文,例如《基于Session识别的高校图书馆电子资源用户访问行为画像分析》《基于社会网络分析的图书共借网络研究——以复旦大学图书馆为例》等。另一方面是基于数据集的应用开发,即通过数据开放大赛等活动激发数据活力,对数据进行多角度的挖掘,包括辅助经营分析一些数据指标、基于用户画像进行个性化服务、各类分析报告等,促进数据的共享与落地。

3 学术期刊参与科研数据出版的建议

数据出版不论对科学研究还是对社会发展都具有深远意义,值得全社会重视其并积极作出相应部署。虽然国内学者在近几年发表了大量有关数据管理、数据出版的论文,但期刊界很少主动参与数据出版。以图书情报学期刊为例,中国知网《2021中国学术期刊影响因子年报》^[15]共列出47种图情期刊,调研其官方网站的投稿指南,对科研数据管理有要求的期刊有《图书情报知识》《数据分析与知识发现》《图书馆杂志》《知识管理论坛》,占比不到10%。这主要有3个方面的原因。①研究者数据意识薄弱。通过对数据进行出版管理,可以实现数据质量可信、数据知识产权清晰、数据长期保存并持久访问,还可以解决数据引用、数据分析、数据统计、数据共享等多方面的问题。虽然政策和机构不断在呼吁其重要性,但是对于学术界来说,取得科学数据仍然是一个挑战。人文社科研究机构及人员对数据共享的积极性不高,对科研数据管理的意识普遍薄弱。②从业者数据技能薄弱。怎样对大量的科研数据进行管理,如何采用合适的工具从数据中选取全部或是部分数据予以对应的分析与对比,怎样挖掘数据价值等,都是对从业者数据技能的挑战,而现阶段,这些技能相对比较薄

弱。③数据出版误区。期刊参与数据出版的意义在于对数据相关内容进行规范性审查,如通过编辑审核、同行评议等方法,提高数据的准确度、可信度,并通过相关规则(版权协议、数据使用权限等)保护数据机构、作者的权益及隐私。数据出版并非数据发布,数据出版也并非简单地上传数据至仓储平台。

《图书馆杂志》的数据出版实践表明,传统学术期刊参与数据出版管理具有可行性。笔者综合国内数据出版的研究现状及国内期刊数据出版管理服务模式,对学术期刊参与科研数据出版提出以下建议。

(1) 建设期刊数据出版管理联盟。我国期刊整体存在“小”“散”“弱”的特点^[16],单本期刊数据发布量有限,数据体量较小,学术期刊单独创建数据出版管理平台成本较高,实际意义不大。数据出版管理平台与单纯的仓储平台有所不同,涉及审稿、发布等多项出版特性。因此,由期刊数据出版管理联盟共建出版管理平台,不仅有利于节约成本,增加数据量,而且可以扩大资源共享,规范行业数据出版管理流程,真正实现将数据管理融入出版全流程中。在学科大融合的发展趋势下,期刊联盟还可以通过联合开发数据出版管理平台、成立期刊共建共享审稿专家库等措施,来方便利用各学科资源进行论文审稿,促进科研成果传播及数据共享与应用。

(2) 提高科研人员数据管理意识。Alex^[17]在*Nature*上的一篇文章对全球数据管理现状进行了调研,调研数据显示尽管整个社会对科研进行了巨额的投资,但数据取得性以每年17%的速度下降,80%超过20年的存储数据已经消失,而数据的丢失是无法挽回的,重复投入经费对数据进行收集意味着投资浪费。在科研生涯的各个阶段,有效的数据管理始终是科研人员必须共同面对的长期问题,所以需要不断提高科研人员的数据管理意识。①培养数据管理技能。对科研数据进行收集、汇总、处理、有效分析、存储、利用及共享,是数据管理的基本技能,而数据管理技能的成熟不是一蹴而就的,应该加强对研究人员数据管理技能的培训。目前已经有许多高校图书馆开展数据技能培训,如上海外国语大学图书馆开设SPSS、Python等数据管理相关软件培训课程。学术期刊也可以通过科普讲座、新媒体推送等提高研究人员的数据管理技能。②形成良好的数据管理习惯。不断提醒研究人员养成定期整理数据、储存新增数据的习惯,以科学的、有组织的方式对数据进行管理,形成适合自己的分类方法,尽可能地将数据

存到云端或是数据管理系统。期刊要求研究人员在投稿论文时提交数据,潜移默化地培养研究人员形成良好的数据管理习惯,以实际行动为保障数据安全提供有力支撑。

(3) 完善数据出版制度。Alice^[18]对全球大约9万名研究人员有关数据共享的看法进行了调研,调研结果显示,造成研究人员不愿共享数据的原因有担心数据被误用、数据公开会让其他人取得先机、数据所有权的流失、涉及隐私及道德问题等,而67%的受访者愿意将数据提交给期刊,期刊出版单位占数据共享机构排行榜之首。可见,获得研究人员的信任,很大程度上取决于数据出版制度规范建设。期刊应该从数据利用者、平台维护者、资金提供者以及数据所有者(作者)等不同角色出发,不断完善满足不同角色需求的数据出版管理制度,如数据共享、数据付费、数据保密、数据版权、风险分担、数据安全、同行评议、价值分配、开放获取等责权内容,消除各方在数据出版管理方面的顾虑,促进数据共享和科技进步。

(4) 开发智慧型数据出版管理平台。智慧型数据出版管理平台应该以用户需求为核心,集学习、社交、分享、互动等多元化需求于一体,提供数据检索、数据分析、知识发现、数据关联、数据可视化等便捷服务。国内现有的数据管理平台大多只是一个数据仓储平台,《图书馆杂志》在探索实践中赋予其出版的含义,即通过对出版流程作出相应的调整,将投稿系统与数据出版管理平台进行融合,但融合程度不足,后期应根据需求不断开发完善平台功能。平台可以参考一些优秀的学术论坛如小木虫、ResearchGate等,建设以数据出版管理为核心分享的学术论坛,用户可以在论坛上畅所欲言地分享自身的数据管理经验、探讨数据管理规则或者遇到的问题,期刊编辑可以不定期地发布一些数据相关的“干货”,形成一个互帮互助的数据学术圈。平台需不断向互联网科技公司学习,如淘宝、抖音等,利用“机器算法+内容”的方法,打造基于用户与内容的数据出版管理信息服务平台,向有特定需求的用户推荐数据集或数据管理经验、数据管理策略等,不断满足用户个性化、智能化、目的化、互动化的需求。通过对用户行为的分析,定向投放用户感兴趣的学术内容,不仅有助于满足用户学术发展的需求,还能激发用户参与互动的热情,形成一个良性的数据学术生态圈。

4 结语

学术期刊对规范科研数据内容、整合科研数据资源、推动科研数据开放共享等有着天然的优势。《图书馆杂志》积极探索科研数据出版的发展道路,建立数据出版管理平台,根据自身期刊论文结构,推出“数据论文”“论文+支撑数据”“论文+富媒体数据”“仓储数据”4种数据出版模式,在实践中构建了全新的论文出版流程,并取得科研数据出版相关成果。创新科研数据出版管理服务模式是机遇也是挑战,存在各种各样的问题、难题需要攻克和解决,对此,要协作成立期刊联盟,共建、共创、共用数据出版管理平台,通过培养科研人员的数据管理技能、良好的数据管理习惯等措施来提高科研人员数据管理意识,还应该从用户角度出发不断完善数据出版制度,开发智慧型数据出版管理平台,满足研究人员对科研数据服务的需求。

参考文献

- [1] Elsevier. Research data [EB/OL]. [2021-11-12]. <https://www.elsevier.com/about/policies/research-data>.
- [2] Springer Nature. Research data policies [EB/OL]. [2021-11-12]. <https://www.springernature.com/gp/authors/research-data-policy>.
- [3] Wiley. Data Sharing and Citation Policies FAQs [EB/OL]. [2021-11-12]. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/faqs.html>.
- [4] SAGE Publishing. Research Data Sharing Policies [EB/OL]. [2021-11-12]. <https://uk.sagepub.com/en-gb/asi/research-data-sharing-policies>.
- [5] Taylor & Francis. Understanding our data sharing policies [EB/OL]. [2021-11-12]. <https://authorservices.taylorandfrancis.com/data-sharing-policies/>.
- [6] 全球变化科学研究数据出版系统 [EB/OL]. [2021-11-12]. <http://www.geodoi.ac.cn>.
- [7] Data Publishing [EB/OL]. [2021-11-12]. https://en.wikipedia.org/wiki/Data_publishing.
- [8] PENEV L, CHAVAN V, HAGEDORN G, et al. Pensoft data publishing policies and guidelines for Biodiversity Data [EB/OL]. [2021-11-12]. https://www.researchgate.net/publication/265422943_Pensoft_Data_Publishing_Policies_

- and_Guidelines_for_Biodiversity_Data.
- [9] 殷沈琴. 面向数据出版的高校社会科学数据资源库的建设[J]. 科技与出版, 2019(4): 29-35.
- [10] KLUMP J, BERTELMANN R, BRASE J, et al. Data publication in the open access initiative[J]. Data Science Journal, 2006, 5(15): 79-83.
- [11] 方恒, 宋立荣. 国内外期刊机构科学数据出版政策研究[J/OL]. 情报杂志: 1-10 [2021-12-10]. <https://kns.cnki.net/kcms/detail/61.1167.G3.20210929.1715.043.html>.
- [12] 许洁, 唐文辉, 夏心悦. 面向实践的数据出版现状分析与对策讨论[J]. 中国科技期刊研究, 2020, 31(11): 1331-1337.
- [13] 图书馆杂志[EB/OL]. [2021-11-12]. <http://www.libraryjournal.com.cn>.
- [14] 图书馆杂志数据出版平台[EB/OL]. [2021-11-12]. <http://data.libraryjournal.com.cn>.
- [15] 中国科学文献计量评价研究中心. 中国学术期刊影响因子年报(人文社会科学)[R]. 北京:《中国学术期刊(光盘版)》电子杂志社有限公司, 2021.
- [16] 陈艳芬, 李东辉. 学术期刊改革: 认知与未来[J]. 中国科技期刊研究, 2013, 24(6): 1060-1062.
- [17] ALEX J. Nine Worrying Stats on the Effect of Poor Scientific Data Management[EB/OL]. [2021-11-12]. <https://blogs.nature.com/ofschemasandmemes/2014/02/03/nine-worrying-stats-on-the-effect-of-poor-scientific-data-management>.
- [18] ALICE M. To Share or not to Share? That is the (Research Data) Question[EB/OL]. [2021-11-12]. <https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/>.

作者简介

谈洁, 女, 1989年, 硕士, 馆员, 研究方向: 数据管理、数据出版、数据共享, E-mail: tanjie.731@163.com。

胡杰, 女, 1992年, 硕士, 馆员, 研究方向: 科学数据管理、数据统计分析。

Practice and Thinking on Scientific Research Data Publication of *Library Journal*

TAN Jie¹ HU Jie²

(1. Shanghai Library (Institute of Scientific and Technical Information of Shanghai), Shanghai 200040, P. R. China; 2. Fudan University Library, Shanghai 200433, P. R. China)

Abstract: Driven by the new paradigm of scientific research, *Library Journal* takes providing high-quality scientific research data as the core, with the help of its data publishing management platform and combined with the traditional academic publishing process, constructs a new publishing mode and process. The practice on scientific research data of *Library Journal* can provide reference for publishing service mode of other traditional academic journals in China. Based on the research status of domestic data publishing and the periodical data publishing management service mode, China's periodical data publishing service needs to participate scientific research data publishing by establishing the periodical data publishing management alliance, improving the data management ability of scientific researchers, perfecting the data publishing system and developing an intelligent data publishing management platform.

Keywords: Data Publishing; Scientific Data; Data Management; Data Sharing; Data Service

(收稿日期: 2021-11-12)