

# 数据期刊的出版模式与发展研究

雷秋雨<sup>1,2</sup> 马建玲<sup>2</sup>

(1.中国科学院大学 北京 100049)

(2.中国科学院兰州文献情报中心 甘肃兰州 730000)

**摘要:** 数据出版当前一个新的发展热点是数据期刊。文章分析了当前几种典型的数据期刊出版模式,即以出版短数据文章(Data Papers)为特征的数据期刊、以整合出版为特征的数据期刊、数据以补充资料出版特征的数据期刊。最后简要论述了数据期刊出版当前的发展状况、存在的问题、建议及数据期刊的发展趋势。

**关键词:** 数据出版; 数据期刊; 数据文章; 科学数据; 科学数据共享

中图分类号: G237.5 文献标识码: A DOI: 10.11968/tsygb.1003-6938.2015020

## Research on Development of the Publishing Models of Data Journal

**Abstract** A new hot topic in current development of data publication is the data journals. The article summarizes several typical current data publishing models of data journals, including short data papers, integrated data journals, and data as the supplemental materials. The article also briefly discusses the current development status, existing problems, suggestions, and the trend of data journals

**Key words** data publishing; data journals; data papers; scientific data; scientific data sharing

### 1 引言

数据出版是近几年出版界和数据共享界共同提出和积极探索的一个新概念,它有望从机制上改变目前数据共享所面临的根本问题<sup>[1]</sup>。数据出版强调的是直接出版数据,而不是从中分析或者获得结论,因此数据出版对于保持数据的完整性、防止科学造假、减少重复的时间及资金投入、提高数据发表者或者发表机构在科研领域的知名度和影响力以及减少跨学科研究的障碍、提高科研合作机率等具有重大意义<sup>[2]</sup>。在实证科学领域,传统的科学数据一般都作为学术出版的一部分。然而随着技术的发展,大量数据快速涌现,动态地改变了学术出版的范式。这些变化也提出需要新的途径和方式来提供和使用科学数据。

数据出版也是新型数字化交流正在发展的一个新概念,是一种新型的鼓励科学家共享数据的方式,对数据的质量控制也具有积极影响<sup>[3]</sup>。从根本上说,数据出版有多种形式,数据期刊(Data Journal)是其中一种新型的数据出版形式。所谓数据期刊一般都是联机出版期刊,它不只出版数据,而发布数据文章(Data Papers)是数据期刊的核

心和主要特征,Data Papers是以描述数据集收集、处理、格式、过程等细节为主要内容的文章,其篇幅较短,能使读者快速地明白所描述数据集的所有相关细节信息。数据期刊以此探求数据的科学认证和再利用,提高科研方法和结果的透明度,以支持良好的数据管理并提供一个到达数据集可访问的、永久的、可解析的路径。

专门从事Data Papers出版的数据期刊的兴起,从某种意义上来说激励着科研工作者们投入更多的时间和精力去准备将自己手中的大量科学数据被共享和使用。为何科研工作者应致力于提交Data Papers, Pensoft出版商Lyubomir Penev认为:提交Data Papers有多项益处并且这些益处并不只限于对Data Papers的作者。首先,数据采集者、数据管理者和作者将通过合理的永久性科学记录获得Data Papers的优先注册和引用;其次,与数据集相关联的扩展元数据也将被正确描述和出版,使数据更易于被其他科学工作者使用、复用及共享,数据也会打开与其他科学团体和机构合作的新视角;最后,对原始数据和收集的数据集的再利用将极大程度改善公共投资资金在管理这些数据方面的效率<sup>[4]</sup>。

## 2 数据期刊的出版模式

### 2.1 以出版短 Data Papers 为特征的数据期刊

Geoscience Data Journal<sup>[5]</sup>(GDJ)是一个与地球科学领域有关的在线出版期刊,学科覆盖范围有:气象、海洋学、大气和海洋化学、冰冻圈、生物圈、地表地质等。该数据期刊主要出版 Data Papers,为地学领域的读者提供一个科学数据被正式发表的开放获取平台。GDJ 出版 Data Papers 的工作流程为:(1)要求作者提交 Data Papers,并要求 Data Papers 对连同提交的数据集进行详细描述,包括原始数据集的收集、处理方式,以及数据集的格式、大小等细节信息;(2)GDJ 要求作者在提交 Data Papers 之前,为数据集找到合适的数据仓储,并获得 DOI;(3)GDJ 的编委会对数据和提交的 Data Papers 进行严格的同行审议;(4)Data Papers 一经采纳,将文章连同已经获取的 DOI 在线发表于期刊。科研工作者及数据使用者可以通过在线浏览 Data Papers 的同时,还可点击与 Data Papers 描述相对应的 DOI,同步浏览相关原始数据集,以此实现 GDJ 和数据仓储的有机链接。提供一个快速便捷获取科学数据的开放式高效的数据出版平台。

Journal of Open Archaeology<sup>[6]</sup>(JOAC)的特点是对描述具有高潜在复用价值的考古数据集的数据论文进行完全的同行审议<sup>[7]</sup>。JOAC 的 Data Papers 会经过严格的同行审议,以确保其描述数据集的准确性,以及内容是否符合该期刊的标准。JOAC 认为 Data Papers 不是研究性文章的取代形式,更是一种对科学研究文章的补充,它鼓励作者在描述数据集的同时可以提及任何与数据集有关的其他研究性文章,便于读者参考。同时,JOAC 要求作者将数据提交于一个允许无限制访问的开放许可之下的公共数据仓储,建议的存储库清单包括 Dataverse Network<sup>[8]</sup>、Figshare<sup>[9]</sup>、mappa<sup>[10]</sup>等。

具体来说,出版 Data Papers 的数据期刊有以下特征:

(1)篇幅精炼,主要包含对原始数据集环境、特征、过程及利用标准的描述;

(2)不对原始数据集进行分析、假设和推演等研究;

(3)开展面向 data papers 的同行评议,保证 data papers 的质量;

(4)以集成统一的方式对数据(集)进行管理,并与 data paper 形成有效关联,提高数据可发现性与可再利用性。

### 2.2 以整合出版为特征的数据期刊

Biodiversity Data Journal<sup>[11]</sup>(BDJ)是生物多样性领域

的开放获取、综合性在线出版平台,是数据期刊的又一种具有代表性的应用与实践,其特点是对数据和论文进行集成出版。BDJ 要求作者在提交研究论文的时候,统一将其底层支撑性数据公开可用,同时必须在文章中保留单独的“数据资源”部分,在此部分中列出数据集名称以及数据存储地址(DOI 或者其它永久链接标识)。BDJ 的出版特点包括:(1)在研究论文文本中包含数据文件,例如数据目录、数据集、数据资料表、参考文献等;(2)补充资料文件,支撑图表、假设、结果等全部作为“数据”与文章一起发表;(3)与大型国际数据仓储合作,如 GBIF、IPT、Dryad 等,集成出版研究论文和数据,文章以 XML 格式发布,保证机器可读。BDJ 意在尽可能的将每篇研究论文中涉及到的零散数据进行整合,以数据文件或补充性资料的形式进行“打包”,出现在论文中单独的“数据资源”部分,使其也可以上载到该期刊的网页伴同文章一起发表。FP7 项目下的 ViBRANT<sup>[12]</sup>计划,就包含了几个由 Pensoft 出版集团书写工具(PWT)和 BDJ 共同推出的几项创新:(1)PWT 提供大量预先定义的、灵活的生物代码和达尔文核心集标准、论文模版;(2)只需点击按钮就可以在论文文本中导入、导出数据文件;(3)支持在各种平台中(Scratchpads、全球生物多样性信息整合发布工具包(IPT)、作者数据库)自动导入生成的数据文件。

GigaScience<sup>[13]</sup>侧重于从整个生命谱学和生物医药科学领域发布“大数据”研究。其出版模式定位为,与提供数据分析工具与云计算资源的大型数据仓储进行合作,要求作者在提交论文的同时将支撑性数据及源代码等原始数据提交于数据仓储。GigaScience 强调数据的质量和效用,目的是为读者群提供优质的大型研究数据(集)。这些数据(集)包含大量的彩色插图和活动图像,可以直接进行读取,从而便于读者根据自身需求对数据进行操作。在 GigaScience 发表的文章也被 PubMed 和 PubMed Central 收录,也可以全文链接至 SpringerLink 数据库,便于读者多途径浏览文章与数据。期刊自身的数据库 GigaDB,也以 DOI 为桥梁,提供已发布论文和相关支持数据之间的直接链接。

以该出版模式为主的数据期刊的特点可归纳为:

(1)基于功能强大的在线出版平台,实现研究论文与数据的集成出版;

(2)以多种形式实现对生物学领域各类特殊数据的出版;

(3)注重研究论文和数据之间互操作的简便性;

(4) 研究论文和相关数据一一对应, 便于读者对数据进行解读。

### 2.3 数据以补充资料出版的数据期刊

Earth System Science Data<sup>[14]</sup> (ESSD) 是一个国际性、跨学科的期刊, 旨在建立一种全新的出版方式, 即在出版学术论文的同时出版原始数据集。发表于 ESSD 的文章包含着几个板块: 固定长度的文章、短篇补充数据集、注释等。文章中的数据部分一般是关于数据的计划、收集、实验、操作等具体方法和途径; 注释部分包含对数据非常规含义的释义, 以及文章描述方法, 比如统计法、筛选法、标准化等。ESSD 有一个创新式的两级出版过程, 第一阶段是基于地球系统科学数据科学论坛 (ESSDD) 的快速网络出版, ESSDD 是一个地球系统科学领域与科学数据出版有关的学术论坛, 它有以下几个特点: (1) 建立科学讨论; (2) 保证最大效率以及科学质量的透明度; (3) 支持快速发表新的科学成果; (4) 科学出版物开放获取。这一阶段, 文章经过快速的同行评议立即发表于 ESSDD 网站, 然后进行互动式公开讨论, 在此期间, 相关评论和短评以及作者的回复都将发布在 ESSDD; 在第二阶段, 同行评议过程完成, 如果文章被采纳, 最终修订后的文章则被发表于 ESSD。保证作者优先权的同时, 对科学讨论提供永久性记录。ESSDD 和 ESSD 都注册了 ISSN 号, 支持永久存档和完全引用。ESSD 也给出了数据仓储建议清单, 如 BODC (British Oceanographic Data Centre)<sup>[15]</sup>、CDIAC (Carbon Dioxide Information Analysis Center)<sup>[16]</sup> 等。

Ecological Archives<sup>[17]</sup> 包含三种出版物: 附录、副刊和 Data Papers, 这三种出版物通常作为 ESA 系列期刊 *Ecosphere*<sup>[18]</sup>、*Ecology*<sup>[19]</sup>、*Ecological Monographs*<sup>[20]</sup>、*Ecological Applications*<sup>[21]</sup> 的补充性资料出现。附录是包含着主体文章支撑信息的在线文件, 它们基于标准浏览器或是可执行音频及视频) 直接可视并且自成一体, 以及包括所有出现在标题、注脚和相关原创性文章的关键匹配资料; 副刊包含着不太适合印刷出版的文件, 这类文件一般以数字格式出现, 通常包括原始和派生数据集、源代码仿真模型、异常统计分析软件等; Data Papers 是用来呈现包含元数据描述的巨大数据集。在发表这些补充性资料时, 要求作者尽可能的精简自己的文章篇幅, 同时将这些有助于读者理解的数据表、额外图表说明分析、图片和所有可引用实体等数据提交于 ESA 官方数据仓储 [data.esa.org](http://data.esa.org)。

以该出版模式为主的数据期刊的特点可归纳为:

(1) 将数据作为研究论文的补充性资料进行出版;

(2) 从提交到出版整个生命周期注重时效性。

## 3 数据期刊的发展现状与发展趋势

### 3.1 发展现状

就国外数据期刊当前发展状况而言, 数据期刊实质上可以归纳为两大出版模式: (1) 以出版研究性论文为主, 将数据集作为论文的补充性文件并存储于出版商网站。具有代表性的出版商有医学、科学领域的 PLoS One<sup>[22]</sup>、Biodiversity Data Journal (BDJ) 等; (2) 专门出版 Data Papers 的数据期刊, 典型的有本文之前介绍的 Journal of Open Archaeology Data (JOAD), 以及 Wiley 于 2012 年与皇家气象学会合作推出的 Geoscience Data Journal (GDJ)。

此外, 从数据期刊的出版环境来看, 整个发展周期中主要涉及三大利益相关者: 出版商、数据仓储、科研工作者。出版商方面, 在美国、英国等国家越来越多的出版商开始关注或者参与数据出版, 如 PLoS One 从 2008 年开始就有了自己的数据共享政策, 该期刊规定作者将数据集存储域在一个开放获取的机构仓储中, 常用的数据仓库如 Dryad, 或者将数据集作为已发表论文的补充性材料文件。Wiley 推出的开放获取期刊 Geoscience Data Journal (GDJ) 主要发表篇幅比较短的地球科学数据文章, Wiley 的副主席 Mike Davis 也曾表示, 科学数据越来越被更多的人所需要, 获取和创建大型新数据是大势所趋, 这也是 Wiley 决定创办 GDJ 的一个重要推动力。

数据仓储方面, 2012 年 7 月 17 日, 欧盟提出, 即将实施的 800 亿欧元“地平线 2020” (Horizon 2020) 研究与创新计划中所发表的研究论文, 必须开放出版或在出版之后将其存放到开放知识库<sup>[23]</sup>。2009 年在伦敦成立的 Datacite, 主要由德国和欧洲其他国家的机构组成, 帮助用户发现、访问和再利用数据, Datacite 的所有数据都采用 DOI 标识, 以便能够检索和加以利用<sup>[24]</sup>。Dryad<sup>[25]</sup> 比其他期刊网站提供了更大的存储空间, Dryad 中所有与文章相关的数据文件都被存储在一起, 通过“再利用条款”来确保数据的最大影响。2004 年中国签署了 OECD《获取公共资助科学数据宣言》。2006 年 10 月, 我国政府官员在第 20 届国际科技数据委员会 (CODATA) 会议上表示, 到 2020 年, 80% 以上的公益性、基础性科学数据资源将通过因特网面向全社会共享<sup>[26]</sup>。Thomson Reuters 旗下的知识产权与科技事业部 (WOK) 于 2012 年推出了研究数据引文索引 (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<sup>[27]</sup>。

毕竟共享原始研究数据并不是科学界的惯常做法,所以科研工作者在期刊要求其提交相关研究数据时还有一些犹豫。如果顺应数据出版商的政策——将原始研究数据和论文一并提交,科研工作者也有一些诉求:(1)希望得到对数据的控制,即必须明确数据提交之后其具体动向;(2)学术回报;(3)不规则的数据滥用对科研工作者个人信誉的影响。

因此,各方利益相关者应该照顾彼此的需求,构想出一个相对完善的体系构架,在保证数据共享的同时,也要确保数据提供者的绝对利益,这样才能形成一个行之有效的模式。

### 3.2 发展趋势

作为科学研究基石的数据,其共享与再利用问题已经普遍存在于科学工作者以及广大数据使用者的意识中。数据期刊作为推动科学数据共享进程的新形式,在未来更是会得到长远的发展。数据期刊自身在未来的发展进程中会日臻成熟,包括逐渐完善的出版流程、明确的数据质量控制方法;出版平台功能的加强;数据期刊自身质量的提升,比如逐步提高被各类重要引文库如SCI、DCI、PubMed等的收录;同时保证作者的学术回报,重视作者要求掌握数据的控制权,避免因数据共享政策缺陷所导致的“数据滥用”对作者个人信用的影响等问题。除此之外,以下几方面,也是数据期刊在不断发展、完善过程中的主要趋势:

(1) 数据期刊与数据整合出版平台的联合与集成。数据期刊能达到一种多角度关联、多终端传播的立体出版形态。几乎所有的数据期刊都在寻求和公共数据仓储、图书馆等数据整合出版平台的合作,要求发表数据文章的作者将其数据包存储于定向存储库,以此建立一种研究性论文与原始数据相关联模式,通过数据期刊上发表的文章,以及整合出版平台分配给数据(集)的唯一数字对象标识或者永久URL地址,就能准确跟踪定位到相关的数据所在位置,满足数据使用者对数据开放存取的要求。由此实现一键式的关联数据共享机制。

(2) 数据期刊与学术出版商建立合作机制。学术期刊是学术交流、科研成果传播的支柱,其早就具备了一套成熟的出版发布与利用机制,数据期刊就是巧妙的应用传统学术出版的这种成熟发布模式,将科学数据嫁接进来,使数据得到和研究论文相同的待遇,经过在线联机出版的渠道展现在读者眼前。科学界的顶级学术出版商Nature<sup>[28]</sup>出版集团在2014年推出了一个数据期刊—

Scientific Data<sup>[29]</sup>,其采用Data Descriptor的格式出版作者提交的数据(集),该格式可以与其他外部出版物或检索平台相关的论文和存储在定向存储库里的数据集有机关联,形成一个关联检索过程,以此实现科学数据的引用与再利用。因此,在传统学术出版中加入数据期刊的元素,不仅方便读者对研究论文和数据进行互操作,其注重时效的在线联机出版,也提高了整个出版过程效率,省力又省时。

(3) 学科交叉下的领域数据库合并<sup>[30]</sup>。数据期刊大都是首先出现在某一学科领域的,比如生物多样性、地球科学、化学、生命科学等领域。但随着其自身不断发展以及学科自身的综合性质,数据仓储不可能一直以专注某一特定学科而存在,势必会出现各领域的数据库合并现象。一旦数据库进行合并,各方利益相关者会发掘到更多的潜在合作机会,数据期刊与数据集的关联集成方式、数据期刊的出版形式等也都会相应发生改变。

## 4 存在问题与发展建议

由于数据期刊本身还是一个新兴的正在发展的领域,各领域的数据期刊也都在探索符合自身特点的发展模式,同时出版数据并非是当今学术交流活动的常规做法,因此,数据期刊本身也还存在一些问题:

(1) 作者对数据期刊的认可度还有待提高;

(2) 在对数据进行同行审议的过程中,数据质量的控制方法、检测标准等流程缺乏系统标准;

(3) 由于数据期刊通常都采取在线联机出版形式,优势在于可以基于网络的超链接特点,便捷的链接至各类数据及文档,方便读者在数据和文章之间轻松操作,但随之而来的数据能否得到安全策管,也是值得商榷的问题。

针对这些问题,笔者提出的完善建议有:

(1) 期刊应当重视数据共享者的利益诉求,并且加强对数据共享者回报机制的建设。为了防止数据的“合法滥用”,有些研究人员希望获得已经共享数据的“控制权”,希望期刊能够提供已共享数据被谁使用、使用在何处等信息。除了对数据进行安全维护,期刊还应当适当配合数据共享者的需求,巩固自身数据出版政策。

(2) 健全数据质量控制评议标准。数据期刊侧重对科学数据进行在线公开发布,因此科学数据也作为重要的科研成果被纳入到学术出版体系中,也就应对其质量加以控制。传统的学术期刊都会由专家对作者提交的论文进行严格的同行评议,其自身也有完善的出版流程。数据

期刊也规定对作者提交的数据进行严格的同行评议, 但是数据由谁、如何、以何标准进行评议并没有明确说明。在这种情况下, 数据的可信任度值得探讨。Nature<sup>[28]</sup>出版集团就明确提出会有专门的数据标准审核专家对作者提交的数据进行评议。

(3) 提高数据策管能力。在网络出版环境下, 读者虽然可以便捷地在数据和论文之间进行互操作, 但数据的安全也存在隐患, 期刊应当强制要求作者将数据提交于有专业管理数据能力的公共数据仓储, 而非建议作者可选择性的将数据链接至个人或机构主页。这样可以有效的降低数据丢失的可能性, 保证对数据的安全策管与链接地址的稳定性。

#### 参考文献:

- [1] 吴宗立, 王亮绪, 南卓铜. 科学数据出版现状及其体系框架[J]. 遥感与技术应用, 2013(6): 383-389.
- [2] 何琳, 常颖聪. 国外科学数据出版研究进展[J]. 图书情报工作, 2013(3): 104-109.
- [3] Data Publishing 2020: Proposal for a Coordinated Approach [EB/OL]. [2014-12-07]. <https://rd-alliance.org/sites/default/files/Data%20Publishing%202020-Proposal%20for%20a%20Coordinated%20Approach.pdf>.
- [4] The Arrival of Data Journals: an interview with Lyubomir Penev of Pensoft publishers [EB/OL]. [2014-12-09]. <http://biofreshblog.com/2012/06/19/the-arrival-of-data-journals-an-interview-with-lyubomir-penev-of-pensoft-publishers/>.
- [5] Geoscience Data Journal [EB/OL]. [2014-12-06]. [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)2049-6060](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049-6060).
- [6] Journal of Open Archaeology [EB/OL]. [2014-12-10]. <http://openarchaeologydata.metajnl.com/>.
- [7] Archiving [EB/OL]. [2014-11-10]. <http://openarchaeologydata.metajnl.com/about/editorialPolicies#custom-0>.
- [8] Dataverse Network [EB/OL]. [2014-11-10]. <http://the-data.harvard.edu/dvn/dv/JOAD>.
- [9] Figshare [EB/OL]. [2014-12-10]. <http://figshare.com/>.
- [10] mappa [EB/OL]. [2014-11-10]. <http://mappa.com>.
- [11] Biodiversity Data Journal [EB/OL]. [2014-11-17]. <http://biodiversitydatajournal.com/>.
- [12] VIBRANT [EB/OL]. [2014-11-09]. <http://www.fp7-vibrant.eu>.

- [13] Gigascience [EB/OL]. [2014-11-17]. <http://www.gigasciencejournal.com/>.
- [14] Earth System Science Data [EB/OL]. [2014-12-06]. <http://www.earth-system-science-data.net/>.
- [15] British Oceanographic Data Centre [EB/OL]. [2014-11-17]. [http://www.bodc.ac.uk/about/what\\_is\\_bodc/](http://www.bodc.ac.uk/about/what_is_bodc/).
- [16] Carbon Dioxide Information Analysis Center [EB/OL]. [2014-11-09]. <http://cdiac.ornl.gov/>.
- [17] Ecological Archives [EB/OL]. [2014-12-06]. <http://esapubs.org/archive/default.htm>.
- [18] Ecosphere [EB/OL]. [2014-11-27]. <http://www.esajournals.org/toc/ecsp/current>.
- [19] Ecology [EB/OL]. [2014-11-27]. <http://www.esajournals.org/toc/ecol/current>.
- [20] Ecological Monographs [EB/OL]. [2014-12-15]. <http://www.esajournals.org/toc/emon/current>.
- [21] Ecological Applications [EB/OL]. [2014-12-15]. <http://www.esajournals.org/toc/ecap/current>.
- [22] PlosOne [EB/OL]. [2014-11-10]. <http://www.plosone.org/>.
- [23] Horizon2020 [EB/OL]. [2014-11-10]. [http://www.bbc.co.uk/zhongwen/simp/science/2014/01/140131\\_horizon\\_2020\\_uk.shtml](http://www.bbc.co.uk/zhongwen/simp/science/2014/01/140131_horizon_2020_uk.shtml).
- [24] What do we do [EB/OL]. [2014-10-10]. <http://www.datacite.org/whatdowedo>.
- [25] Dryad [EB/OL]. [2014-12-11]. <http://dryad.stanford.edu/>.
- [26] 开放数据调研/开放资源建设 [EB/OL]. [2014-12-10]. <http://open-resources.las.ac.cn/drupal/?q=node/3064>.
- [27] Thomson Reuters. Data Citation Index [EB/OL]. [2014-12-11]. <http://wokinfo.com/products-tools/multidisciplinary/dci/>.
- [28] Nature [EB/OL]. [2014-10-21]. <http://www.nature.com>.
- [29] Scientific Data [EB/OL]. [2014-10-21]. <http://www.scientificdata.com>.
- [30] Susan Reilly a, Wouter Schallier, Sabine Schrempf, Eefke Smit, Max Wilkinson. REPORT EPORT ON INTEGRATION OF DATA AND PUBLICATIONS [EB/OL]. [2014-10-21]. <http://www.ode-project.eu>.

作者简介: 雷秋雨 (1989-), 女, 中国科学院兰州文献情报中心硕士研究生; 马建玲 (1969-), 女, 中国科学院兰州文献情报中心研究员。