科学数据出版模式比较研究

□黄国彬　王舒*　屈亚杰

摘要　根据科学数据产生情形，将出版模式归纳为科学数据集成出版与独立出版两种模式，并从数据集描述方式、质量审核方式与出版费用三个方面对归纳的出版模式进行了比较分析。通过对不同出版模式下的典型案例的分析，总结出如下结论：数据集描述方式和对象与出版模式中涉及的出版主体有关；独立出版模式的质量审核更为完善；不同出版模式，其付费标准和费用承担主体不同。对当前科学数据出版模式的比较研究，可为国内科学数据出版实践发展提供借鉴。

关键词　科学数据　出版模式　集成出版　独立出版

分类号　G255.51

DOI　10.16603/j.issn－1002－1027.2018.01.005

1　引言

德国蒂姆(Thieme)出版集团的化学杂志负责人苏珊娜·哈克(Susanne Haak)博士认为："在自然科学领域，获取原始研究数据是研究工作的一个基本条件。"[1]这一观点同样适用于人文社会科学领域。例如人文社会科学领域的定量研究大多基于调查数据、统计数据，其研究开展的关键在于原始数据的获取。而不论是自然科学还是人文社会科学领域，数据的获取一般有以下两种途径：一是通过研究者本人观测、实验、调查或访谈获取数据；二是通过共享其他研究者已产生的数据。在实际开展研究过程中，由于资金、设备、场地、时间等因素限制，部分研究人员难以通过自身努力产生科学数据。基于此，研究者通过数据共享渠道，利用已有的科学数据来开展研究工作显得尤为重要。目前，虽然美国的国家自然科学基金委员会、英国的皇家理事会及其七个研究委员会等科研资助机构，出台了关于科学数据共享的政策，但由于缺乏有效的科学数据引用机制和学术贡献认可机制，大量分散在不同国家、不同地区、不同研究项目及各个科学家手中的科学数据难以主动、有效地共享。如何解决上述问题成为期刊出版界、数据知识库和科研基金机构探讨和研究的主要问题。而在数字环境和开放获取潮流中，科学数据出版被认为是推动科学数据共享的重要手段。但到目前为止，学术界对科学数据出版并没有统一明确的定义。而明确含义是总结和分析科学数据出版模式的前提。因此，笔者结合国内外学者的观点，总结出科学数据出版是正式的科学数据发布，即在数据发布过程中遵循规范的质量管理和控制流程，为数据用户增加数据的价值，如创建元数据和同行评审等，使得科学数据集达到可发现、可获取、可理解和可追溯的状态。

2　国内外研究现状

目前国内外学者对科学数据出版模式的研究，可以归纳为如下三个方面：(1)对某一种出版模式的深入研究。例如，马建玲、苏珊·莱利(Susan Reilly)等人对学术论文与科学数据集成出版进行了研究[2-3]。布瑞泽(Brase)等人以世界数据中心(World Data Centers，WDC)和德国科学技术图书馆为例，对数据知识库独立出版科学数据的模式展开研究[4]。(2)以典型案例为研究对象，探讨该案例的出版模式。例如，李红星等人结合寒区旱区科学数据中心的实践经验，提出数据中心和多期刊联合出版数据的模式，并对该模式深入探讨[5]。高雅探讨了ENCODE项目科学数据出版模式，并总结了该

* 通讯作者：王舒，ORCID：0000-0003-3330-103X，201521260038@mail.bnu.edu.cn。

出版模式在科学数据挖掘和利用方面的创新之处[6]。(3)对某个学科领域的出版模式展开研究。例如埃里克(Eric C. Kansa)等人研究了考古学领域的科学数据出版现状与存在的问题,在此基础上提出"数据共享即出版"的科学数据出版模式[7]。

上述关于科学数据出版模式的研究,多集中于对某一种出版模式或典型科学数据出版案例进行详细介绍和分析,而对于多种模式的比较研究开展得较少。虽然梁子裕和张静蓓对多种科学数据出版模式进行归纳总结和比较分析,但仅笼统地从优缺点的角度进行分析,忽略了科学数据出版过程中的数据集描述、质量审核等关键环节的比较分析[8-9]。此外,现有研究中科学数据出版模式的划分标准模糊不清。因此,文章根据科学数据的产生情形,结合科学数据出版的定义,归纳与总结当前科学数据的出版模式。在此基础上,运用比较分析法,从数据集描述、质量审核和出版费用三个角度对不同模式展开比较分析。

3 科学数据的出版模式及案例分析

首先,根据科学数据的产生情形对科学数据的出版模式进行一次划分。科学数据的产生情形有两种,一是伴随着期刊论文的产生而产生的,通常是科研人员为了研究某个问题,使用实验、调查或模拟的研究方法,并在此过程中产生了原始数据。此类数据通常与某篇学术论文相关,因此通常与学术论文集成出版。第二种情形是通过仪器设备观测自然现象,并以数据的形式将其观测的现象记录下来,例如水文数据、气象数据等;或者是对社会现象的调查、统计,并以数据的形式呈现出来,例如调查数据、统计数据等。此类数据是在特定的时间、地点产生的,因其产生的时空条件具有不可复制的特点,需要被很好地保存。该类科学数据的产生与学术文献没有直接关系,通常以独立学术成果的形式出版。因此,笔者将此出版模式称为科学数据独立出版模式。其次,对集成出版而言,依据科学数据的存储位置,可细分为学术期刊自建设施自行出版和学术期刊与公共数据仓储合作出版两种。对独立出版模式而言,依据出版主体,可细分为数据知识库出版和数据期刊出版两种(如图1所示)。

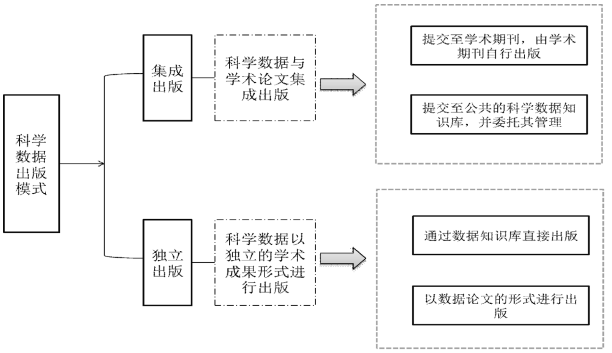


图1 科学数据出版模式框架图

3.1 科学数据集成出版

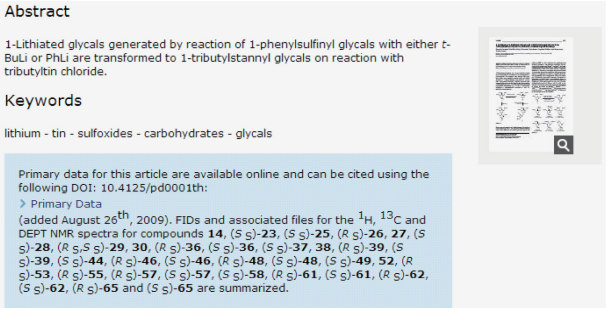
一篇学术文献往往离不开原始数据的支撑。在传统的学术出版体系中,出版商更多地关注学术期刊、科技报告等文本文献的出版。但随着数据驱动研究范式的形成和数字科研环境的影响,支撑学术论文的原始数据在科学研究活动中的地位越来越重要。周波认为科学数据已经由科学研究的起点和基础发展成为科研活动的牵引力之一[10]。越来越多的学术期刊出版商认识到在出版学术论文的同时出版支撑论文的原始数据的重要性。因此,学术论文与科学数据集成出版的模式应运而生。将二者集成出版,一方面为学术论文的评价与评审提供数据支持;另一方面,数据用户通过文献可以更好地理解科学数据,从而推动后续科研人员以出版的科学数据为基础开展更深层次或更广范围的研究,实现科学创新[11]。

通过该模式出版科学数据,流程通常是,在学术期刊的要求下,作者在向期刊提交学术论文的同时提交支撑其论文的科学数据,有的期刊选择自行管理科学数据,而有的期刊选择将科学数据提交至公共数据仓储并委托其管理。根据科学数据管理机构和存储位置的不同,可将集成出版模式分为以下两种。

3.1.1 提交至学术期刊,由学术期刊自行出版

该模式以学术期刊为科学数据的出版主体,学术期刊在出版学术论文的同时将支撑学术论文核心论点的原始数据进行出版。该模式下,科学数据是由学术期刊自建设施自行管理,典型案例为德国蒂姆(Thieme)集团出版的化学期刊,如 *Synthesis*。2010年,Thieme意识到,由于化学领域没有科学数据中央存储库,大量的数据分散在科学家自己的电脑上,导致数据不能有效地发布和共享[12]。因此,

它与德国科学技术图书馆合作建立科学数据存储平台，成为第一家将原始数据与学术论文同时出版并在世界范围内共享的出版社。其出版的科学数据主要是实验中收集的原始数据，该数据帮助化学领域的学者了解在化学反应的过程中发生了什么，以便更好地理解论文。原始数据以压缩文件的形式出版，用户可无限制地下载至本地。数据集与论文集成在同一界面，其出版信息包括数据集出版时间、摘要和 DOI 号（如图 2 所示）。


Abstract

1-Lithiated glycals generated by reaction of 1-phenylsulfinyl glycals with either *t*-BuLi or PhLi are transformed to 1-tributylstannyl glycals on reaction with tributyltin chloride.

Keywords

lithium - tin - sulfoxides - carbohydrates - glycals

Primary data for this article are available online and can be cited using the following DOI: 10.4125/pd0001th:

Primary Data

(added August 26th, 2009). FIDs and associated files for the ^{1}H, ^{13}C and DEPT NMR spectra for compounds **14**, (*S* S)-**23**, (*S* S)-**25**, (*R* S)-**26**, **27**, (*S* S)-**28**, (*R* S,*S* S)-**29**, **30**, (*R* S)-**36**, (*S* S)-**36**, (*S* S)-**37**, **38**, (*R* S)-**39**, (*S* S)-**39**, (*S* S)-**44**, (*R* S)-**46**, (*S* S)-**46**, (*R* S)-**48**, (*S* S)-**48**, (*S* S)-**49**, **52**, (*R* S)-**53**, (*R* S)-**55**, (*R* S)-**57**, (*S* S)-**57**, (*S* S)-**58**, (*R* S)-**61**, (*S* S)-**61**, (*R* S)-**62**, (*S* S)-**62**, (*R* S)-**65** and (*S* S)-**65** are summarized.

图 2　Thieme 科学数据出版信息页[13]

Thieme 科学数据的出版流程为：(1)在提交论文的同时，将支持论文的原始数据提交至 Thieme；(2)Thieme 将科学数据贮存在自建的数据中心；(3)在德国科学技术图书馆注册并由其向原始数据分配 DOI 号；(4)原始数据通过添加在 Thieme e-Journal 系统中实现出版[14]。

3.1.2　提交至公共的科学数据知识库，并委托其管理

在该模式下，科学数据的出版主体为学术期刊和科学数据知识库。当作者向期刊提交论文时，被要求将支撑论文核心论点的科学数据提交至科学数据知识库，由科学数据知识库向数据集分配 DOI 号，之后，作者需在待发表的论文中引用数据集，同时科学数据知识库在数据出版页面上插入论文基本信息。该模式的典型案例为 Dryad 数字仓储（Dryad Digital Repository，Dryad）提供的集成出版服务。Dryad 的创建与发展起源于联合数据存储政策提议（Joint Archiving Policy，JDAP）[15]。JDAP 是由生物进化领域的许多期刊于 2011 年联合发表的一个政策模式倡议。其核心内容是：出版学术论文的前提是将支撑论文结论的数据存储至合适的数据知识库中公开出版，并委托其管理。目前该倡议提出的模板得到了其他领域学术期刊的认同。Dryad 作为一个公共数据知识库，是所有接受 JDAP 模板的学术期刊所推荐的数据仓储之一。目前有 107 家期刊实现了与 Dryad 的集成出版[16]。因此，笔者选取 Dryad 集成出版服务为该模式的典型案例并重点分析。Dryad 将数据集信息与来源文献信息集成在一个文件包中，可在该包中访问来源文献和下载科学数据。此外，Dryad 要求，在来源文献中必须插入存储在 Dryad 中的支撑数据的 DOI 号。

在 Dryad 集成出版中涉及三个主体，分别是作者、期刊、数据仓储。论文由作者提交至期刊，由期刊进行出版；而数据集由作者提交至 Dryad，由 Dryad 进行出版，但二者需要集成，因此出版流程较为复杂[17]，如下：(1)作者向期刊提交论文手稿；(2)当论文审核通过后，期刊利用自动通知系统告知 Dryad，之后，Dryad 为预期数据创建临时记录，并将临时记录的链接发送给期刊(3)同时期刊通知作者向 Dryad 提交数据并向作者提供临时记录的链接；(4)作者将数据及其描述文件上传至临时记录中；(5)Dryad 为记录分配一个 DOI 号，并将其同时发送给作者和期刊；(6)Dryad 工作人员审查数据，对每个文件执行基本检查，并发布数据包；(7)期刊保证其所有版本的论文都在文中和文后参考引文中附有数据集的 DOI 号。此外，当论文在审核中时，Dryad 会将数据集的临时记录链接发送至期刊，以供同行评审人员对数据集进行审查。

3.2　科学数据独立出版

并不是所有的科学数据都有来源文献，例如水文数据、气象数据等，是通过某个长期观测项目持续产生的，从来源层面并没有直接相关的文献。对于此类科学数据，通常是以独立学术成果的形式出版。该出版模式下，数据归档是出版的必要环节。根据出版流程，可将独立出版模式划分两种。一是直接通过数据归档机构出版；二是将数据归档后，由数据期刊出版。

3.2.1　数据知识库出版

在该出版模式中，数据知识库是科学数据出版的核心机构。数据生产者根据其科学数据的特点和资助机构的要求，选择合适的数据知识库存储其科学数据，数据知识库按其政策收取或免去费用，对科学数据实施长期监护，并向每个数据集提供 DOI 号，以保证科学数据的可访问性、可理解性和可共享性。目前，数据知识库的学科范围较广，从自然科学到社会科学；涉及的数据多为特定时空范围内通过

大型仪器设备或大型调查而产生的，其类型主要包括观测数据、统计数据和调查数据等。其中较为典型的数据知识库出版案例为英国的信息环境数据中心（The Environmental Information Data Centre，EIDC）。EIDC 由世界级的研究中心生态与水文中心（Centre for Ecology & Hydrology，CEH）主办，出版由 CEH 长期监测全英水文环境而获得的观测数据[18]。与集成出版下的两种出版模式不同，EIDC 出版信息较为单一，仅有数据集本身的相关信息，主要包含基本出版信息、数据集内容信息和数据集获取信息三个方面。

由于该模式涉及的数据出版主体仅为数据知识库，因此出版流程较为简单。EIDC 的出版流程为：(1)数据生产者提交数据，同时须提交数据集的描述文档；(2)数据知识库审查数据质量和元数据质量(3)数据知识库存储数据，将数据存储在安全的位置并定期备份和审查；(4)数据知识库为数据分配 DOI 号；(5)公开发布数据集，数据用户可通过 CEH 数据目录检索数据集，出版完成。

3.2.2 数据期刊出版

数据期刊出版是指将科学数据作为独立的学术成果，以数据论文的形式通过数据期刊出版。该出版模式下，出版的内容包含数据论文与科学数据两个实体。数据论文是正式发表的一种论文类型，但与传统学术论文不同，其内容并不包含基于科学假设和科学问题的研究结果，而是对数据采集、获取、处理等过程和方法的描述，其目的是让科研群体更好地发现、获取、理解与复用数据，从而促进科研创新。数据论文发表在数据期刊上，而科学数据存储在数据知识库中，但数据论文须与其描述的公开发布的科学数据链接，这个链接通常可以是 URL 或数据知识库分配给数据集的 DOI [19]。

该模式较为典型的案例是《地理数据期刊》(Geoscience Data Journal，GDJ)。GDJ 是 Wiley 旗下的在线开放获取数据期刊[20]，通过数据论文的形式出版地理领域的科学数据。GDJ 的出版流程包括以下几步：(1)将科学数据存储至推荐的数据机构库，并获得 DOI 号；(2)按照期刊的格式要求，撰写数据论文并在线提交；(3)数据论文和数据集接受同行评议；(4)发表数据论文，并在论文中附上数据集的 DOI 号。用户可通过数据论文获取数据集的描述信息及数据集本身。

4 科学数据的不同出版模式比较分析

总体来说，当前科学数据出版处于多种模式并存的阶段。虽然每种出版模式各不相同，但其出版过程都会涉及以下三个环节，即对数据集进行描述、质量审核、收取出版费用。下面以这三个环节为切入点，对当前科学数据的不同出版模式进行比较分析。

4.1 数据集描述

对科学数据进行描述是数据出版的第一步。这是由于科学数据与学术论文、科技报告、科技图书等传统的科学文献不同，从形式来看可能是一组观测数值、实验数据记录、问卷数据或者一段计算机代码。如果不对其变量含义、产生背景、获取方法等进行描述，则无法掌握科学数据的具体含义。所以不论是通过何种模式出版科学数据，都需要对科学数据进行描述。但不同的出版模式对科学数据进行描述的方式和描述的具体内容有所不同。文章根据每种出版模式的典型案例，对不同出版模式下的描述方式和描述内容进行了总结(表 1)。

表 1 四种出版模式对数据集的描述方式和具体对象

出版模式 \ 比较内容		描述方式	描述的具体内容
集成出版	提交至学术期刊，由学术期刊自行出版	非结构化	数据集基本信息、来源文献信息
	提交至公共的科学数据知识库，并委托其管理	结构化	数据集基本信息、来源文献信息
独立出版	数据知识库出版	结构化	数据集基本信息、数据集使用信息、数据集内容信息
	数据期刊出版	半结构化	数据集基本信息、数据集使用信息、数据集内容信息

就描述方式而言，主要包括结构化描述、半结构化描述和非结构化描述。结构化描述是指以元数据为描述工具对科学数据进行具体说明，通常遵循某个元数据标准，其描述结果可被计算机自动识别。例如，Dryad 和 EIDC 都以元数据为描述工具对科学数据进行结构化描述。前者遵循 Dryad 元数据应用方案(Dryad Metadata Application Profile)[21]，该方案是在都柏林核心的基础上结合本仓储集成出版的特点进行扩展的。后者采用 UK GEMINI 标准和 INSPIRE 标准[22]，这两个元数据标准分别是英国和欧盟在地理学领域的认可元数据标准。非结构化描述方式是指以文本的形式对数据集进行描述，并没有标准或固定的格式可供遵循，且描述结果仅人工

可读。例如 Thieme 对出版的科学数据以非结构化的方式进行描述，在数据集文件包中有一个以“Readme”命名的 PDF 文件，其内容是对数据集的简要描述。半结构化的描述方式介于上述两种描述之间，是指按照固定的结构对科学数据进行描述。数据期刊通常采用半结构化方式描述科学数据，这是由于数据期刊发表的数据论文即为对科学数据的描述，而数据论文一般具有固定的结构。例如《地理数据期刊》规定了详细的数据论文结构[23]，包括标题页、数据集信息页、正文、致谢、参考文献等。其中正文包含数据摘要、数据产生方法、数据格式与存储位置、数据集使用与再利用信息。

就描述的内容而言，包括数据集基本信息、数据集来源文献信息、数据集使用信息、数据集内容信息。数据集基本信息包括数据集的题名、作者、出版时间和摘要；来源文献信息是指数据集所支撑的文献的基本信息；数据集使用信息是指数据集使用所遵循的协议或规定；数据集内容信息是指数据集产生方法、数据集质量、数据集的时空覆盖范围等。属于集成出版模式的 Thieme 和 Dryad 描述的具体内容为数据集基本信息和来源文献信息；而属于独立出版模式的 EIDC 和 GDJ 描述的具体内容包括数据集基本信息、数据集使用信息和数据集内容信息。

使用何种描述方式描述科学数据取决于出版过程中涉及的出版主体。出版主体包含数据知识库的出版模式，通常采用结构化的描述方式如集成出版模式下委托数据知识库管理模式和独立出版模式下数据知识库直接出版模式；出版主体涉及学术期刊的出版模式，采用非结构化的描述方式；出版主体为数据期刊的出版模式采用半结构化的描述方式。此外，四种模式在描述的具体内容方面既有共性也有差异。数据集基本信息为描述的最小信息集合，四种出版模式在描述数据集时均包含了数据集的基本信息。来源文献信息为集成出版模式的必备描述项目，这是由集成出版的数据与文献集成特性决定的。而独立出版的两种模式的描述内容除了数据集基本信息外，还包含了数据集使用信息和数据集内容信息。因此，与集成出版模式相比，独立出版模式对数据集的描述更为详细和全面。

4.2 质量审核

质量审核是学术出版的核心环节。而科学数据作为学术成果的一种类型，在出版过程中必然要经过质量审核这一关键环节。传统学术文献已有成熟的质量审核机制，如学术论文通过同行评议的方式进行质量审核。而科学数据作为一种新型的学术资源，其质量审核机制尚未成熟和统一，具体表现为不同出版模式的质量审核主体和对象不尽相同（表2）。在科学数据出版中，质量审核主体主要包括期刊编辑、数据知识库工作人员、同行评议人员和数据集生产者；审核对象包括数据集本身、元数据、数据论文等。其中对数据集本身质量而言，包括技术质量与科学质量。科学数据的技术质量是指数据集本身的完整性和描述的充分性，而科学质量是指数据集收集方法的评价、科学数据的合理性和再使用的价值[24]。元数据是关于数据的数据，是对数据集的结构化描述信息，其质量主要体现在是否完整、准确地以可理解和可获取的方式描述数据集。数据论文是解释数据的学术性文章，一般有其固定的结构，因此在数据论文的质量审核中需要考虑文章的科学性、结构的完整性等。以所举的四个典型案例对此进行具体说明。与学术论文一同提交至 Thieme 的原始数据，是通过期刊编辑和同行评议人员对原始数据的质量进行审核。期刊编辑主要审核的是数据集是否完整，是否论文内容一致。同行评议人员审核的对象主要是学术论文，但在审核过程中为了查验论文的科学性，可向编辑提出申请以审查与论文相关的原始数据，客观上实现了原始数据的科学质量的审核[25]。Dryad 通过与学术期刊合作实现科学数据与其来源文献集成出版。在出版过程中，Dryad 工作人员与同行评议人员共同完成对科学数据的质量审核。其中 Dryad 工作人员在数据提交至仓储时对数据的技术质量进行审核，例如检查上传的数据集是否能够打开、数据在传输过程中是否有损坏等，而同行评议人员主要是来源文献所在期刊的评审专家，在评审论文的过程中对数据集进行访问以审查数据集的科学质量。EIDC 对其出版的科学数据，由其工作人员对数据集进行质量审核，主要审核的对象是数据集的技术质量和元数据。该数据中心提出了元数据质量审核的标准：准确性、可用性、可读性和可检索性[26]。而 GDJ 的质量审核主体为同行评议人员，对数据集的技术质量和科学质量、元数据、数据论文三方面进行了全面的审核[27]。

表 2　四种出版模式的质量审核主体与对象

出版模式	比较内容	审核主体	审核对象
集成出版	提交至学术期刊，由学术期刊自行出版	期刊编辑 同行评议人员	数据集技术质量和科学质量
	提交至公共的科学数据知识库，并委托其管理	数据知识库工作人员、同行评议人员	数据集的技术质量和科学质量、元数据
独立出版	数据知识库出版	数据知识库工作人员	数据集的技术质量、元数据
	数据期刊出版	同行评议人员	数据集的技术质量和科学质量、元数据、数据论文

总体来说，除了独立出版模式下的数据知识库出版模式外，其他三种模式的科学数据质量审核主体中都有同行评议人员。对于集成出版模式下的两种出版模式而言，同行评议人员审核的重点是学术论文，而对科学数据的科学质量审核处于辅助地位。而数据期刊出版模式下，同行评议人员的审核重点包含数据集及围绕数据展开的数据论文和元数据。此外，数据知识库模式的审核主体为数据知识库工作人员，其为数据管理专家并非某一领域的学科专家，因此对数据集的质量审核侧重于数据集的技术质量和元数据全面性、准确性等，并不能深入审核科学数据的科学质量。因此，相比而言，独立出版模式的质量审核更为全面和深入。

4.3　出版费用

在数据出版过程中，通过对数据集描述和质量审核，为数据用户增加了数据集的价值。因此数据出版与传统学术出版相同，需要一定的出版费用。但不同出版模式的收费标准和费用的具体承担者有所不同(表 3)。具体而言，收费标准包括数据集大小、是否为资助项目、文章数量等。作为集成出版的一种，Dryad 按数据集大小进行收费。当数据集小于 20GB 时，每个数据包的费用为 120 美元；当数据集大于 20GB 时，每超过 10GB 多收取 50 美元[28]。EIDC 受英国自然环境理事会资助，因此该数据中心按数据的产生过程是否受到该理事会的资助为标准进行收费。一般而言，对于受资助项目产生的科学数据，该数据中心不收取出版费用。而非资助项目产生的科学数据出版时缴纳的费用取决于出版数据所花费的时间、精力和财力[29]。以数据论文的形式通过数据期刊出版科学数据，数据出版费用即数据论文的出版费用。GDJ 规定，每篇数据论文的出版费用为 1500 美元[30]。Thieme 出版的化学期刊 *Synthesis* 在作者指南中规定，不收取作者任何费用，包括论文及其原始数据的出版费用。

对于数据出版费用的承担者而言，主要包括学术期刊、数据生产者与资助机构。集成出版模式下，数据出版的费用由学术期刊独自承担或者学术期刊与数据生产者(在集成模式下同时为论文作者)共同承担。例如，Thieme 不收取数据生产者任何费用，这意味着数据出版的费用完全由期刊承担。而 Dryad 建议数据出版费用由期刊通过赞助的形式帮助作者负担数据出版费用，因此出版费用由学术期刊与数据生产者共同承担。独立出版模式下，数据出版的费用通常由数据生产者承担。但由资助项目所产生的科学数据通过该资助机构建立的数据仓储出版时，数据生产者无需承担出版费用。例如英国信息环境数据中心规定不收取英国自然环境理事会所资助的研究项目产生的科学数据的出版费用。

表 3　四种出版模式的收费标准和费用承担者

出版模式	比较内容	收费标准	费用承担者
集成出版	提交至学术期刊，由学术期刊自行出版	无	学术期刊
	提交至公共的科学数据知识库，并委托其管理	数据集大小	数据生产者、合作期刊
独立出版	数据知识库出版	是否受资助	自然环境理事会、数据生产者
	数据期刊出版	按篇收费	数据论文作者（数据生产者）

总体来说，四种科学数据出版模式在收费标准方面存在较大差异，集成出版模式下的将数据提交至公共的科学数据仓储并委托其管理模式的收费标准是数据集大小，而期刊自行出版模式不收取数据生产者任何费用。独立出版模式下的数据仓储直接出版模式采取的收费标准是科学数据是否由资助项目所产出，这是因为该模式下的数据仓储一般是国家数据中心或大型的公共数据仓储，由政府或科研资助机构的基金资助而建立的，所以，由资助项目产出的科学数据通过这些数据仓储出版时，无须缴纳出版费用。而数据论文出版模式采取的收费标准和学术期刊类似，按篇收费的标准简单明了，易于实施。在科学数据出版费用的具体承担者方面，四种出版模式存在一些共性，即有三种出版模式的费用承担者都涉及数据生产者，集成出版模式的费用承担者都包括学术期刊。

5 结语

大数据时代,科学数据在学术研究中的价值日益受到研究人员的重视,期刊出版商、数据知识库以及科研资助机构等主体越来越多的参与到科学数据出版实践之中,不同程度上推动了科学数据出版的发展。然而,与传统的学术文献出版相比,科学数据作为一种新型的学术成果,其出版体系还在发展之中。

文章从科学数据产生情形出发,结合现有的出版实践和前人理论研究成果,归纳出四种科学数据出版模式,并从数据集描述方式、质量审核方式、引用要求与出版费用等方面对不同出版模式进行了比较分析。总结出以下结论:(1)不同出版模式中涉及的出版主体不同,因此描述数据集的方式不同;科学数据独立出版模式比集成出版模式对科学数据的描述更为详尽。(2)与集成出版模式相比,独立出版模式对科学数据的质量审核机制更为完善。(3)由于科学数据多为开放获取出版物,因此数据生产者需要承担出版费用;但两种模式不同是,集成出版模式中,传统期刊对作者提供资助。

参考文献

1 Thieme. Primary data for chemistry available for the first time [EB/OL]. [2017－04－06]. https://www.thieme.de/statics/dokumente/thieme/final/de/dokumente/zw_synthesis/CFZ-Primary-Data-Press-Release.pdf.

2 马建玲,曹月珍,王思丽,王楠,李慧佳.学术论文与科学数据集成出版研究[J].情报资料工作,2014(2):82－86.

3 Reilly S,Schallier W,Schrimpf S, et al. Report on integration of data and publication[EB/OL]. [2017－04－06]. http://libereurope.eu/wp-content/uploads/ODE-Report On Integration Of Data And Publication.pdf.

4 Brase J,Schindler U. The publication of scientific data by World Data Centers and the National Library of Science and Technology in Germany[J]. Data Science Journal,2006(5):205－208.

5 李红星,吴立宗,南卓铜,王亮绪,王建,赵彦博,张耀南,马瀚青.科学数据联合出版模式与内容研究[J].遥感技术与应用,2016(4):801－808.

6 高雅,翁彦琴,董文杰.ENCODE 项目科学数据出版模式创新[J].中国科技期刊究,2015(8):808－812.

7 Kansa EC, Kansa SW. We all know that a 14 is a sheep: data publication and professionalism in archaeological communication [J]. Journal of Eastern Mediterranean Archaeology and Heritage Studies. 2013,1(1):88－97.

8 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.

9 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.

10 周波.科学数据元数据标准的设计与应用[D].东南大学,2013.

11 邱春艳.科学数据与期刊文献的关联实现研究[J].图书馆杂志,2015(8):29－33.

12 同 1.

13 Thieme.科学数据出版[EB/OL].[2017－04－06]. https://www.thieme-connect.de/DOI/DOI? 10.1055/s-2008-1067226.

14 同 3:52.

15 Dryad. Joint archiving policy[EB/OL].[2017－4－6]. http://www.datadryad.org/pages/jdap.

16 Dryad. Submission integration: current status [EB/OL].[2017－04－06]. http://wiki.datadryad.org/Submission_Integration:_Current_Status.

17 Dryad. Submission integration: implementation[EB/OL].[2017－04－06]. http://wiki.datadryad.org/Submission_Integration:_Implementation.

18 EIDC. About [EB/OL].[2017－04－06]. http://eidc.ceh.ac.uk/about.

19 Chavan V,Penev L. The data paper: a mechanism to incentivize data publishing in biodiversity science[J]. BMC Bioinformatics,2011,12(15):1－12.

20 GDJ. Overview[EB/OL].[2017－04－06]. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049－6060.

21 Dryad metadata application profile [EB/OL].[2017－04－06]. http://wiki.datadryad.org/Metadata_Profile.

22 EIDC. Metadata guidance [EB/OL].[2017－04－06]. http://eidc.ceh.ac.uk/support/metadata-guidance.

23 GDJ. An example of how to structure a Data Paper [EB/OL].[2017－04－06]. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049－6060/homepage/ForAuthors.html # dataset_submission.

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 Thieme. Primary data for chemistry available for the first time [EB/OL].[2017－04－06]. https://www.thieme.de/statics/dokumente/thieme/final/de/dokumente/zw_synthesis/CFZ-Primary-Data-Press-Release.pdf.

26 EIDC. Guidance for metadata authors [EB/OL].[2017－04－06]. http://eidc.ceh.ac.uk/help/catalogueHelp/createedit/metadataauthorguide.

27 GDJ. Guidelines for reviewers[EB/OL].[2017－04－06]. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049－6060/homepage/guidelines_for_reviewers.htm.

28 Dryad. Data publishing charges[EB/OL].[2017－04－06]. https://www.datadryad.org/pages/payment.

29 EIDC acquisition policy [EB/OL].[2017－04－06]. http://eidc.ceh.ac.uk/about/eidc-acquisition-policy.

30 GDJ. Article publication charges [EB/OL].[2017－04－06]. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049－6060/homepage/article_publication_charges.htm.

作者单位:北京师范大学政府管理学院,北京,100875
收稿日期:2017 年 4 月 26 日

(转第 33 页)