

# 影响科学数据引用的外部因素分析\*

黄国彬<sup>1</sup>, 刘馨然<sup>1</sup>, 姜颖<sup>2</sup>

(1. 北京师范大学政府管理学院, 北京 100875; 2. 对外经济贸易大学图书馆, 北京 100029)

**摘要:** 为引导科学合理地引用科学数据, 提高科学数据的质量和权威程度, 有必要关注影响科学数据引用的关键因素。本文采用文献调研法和网站调研法, 站在科学数据引用者的立场, 结合国内外已有研究和实践经验, 梳理科学数据引用的主要外部影响因素, 包括科学数据引用规范与格式、科学数据权利保护与引用许可、科学数据发布平台与管理软件, 并指出对这三方面的有效处理和应对是科学数据引用者规范其引用行为的主要考虑点。

**关键词:** 科学数据引用; 影响因素; 引用规范; 引用许可

**中图分类号:** G255

**DOI:** 10.3772/j.issn.1673-2286.2017.06.001

科学数据是科技活动中产生的基本数据, 包括观测数据、考查数据、实验数据、统计数据等<sup>[1]</sup>。科学数据引用指科研工作者引用科学数据作为论文观点的支撑数据, 并以参考文献、脚注或文中注等方式, 对其所引用数据提供数据参考的做法<sup>[2]</sup>。对科学数据进行引用能有效促进科学数据传播, 也可以在后期更好地对科学数据进行评价和计量。在20世纪80年代已有学者指出, 其科研过程产生的数据文件应该被引用, 从而确定特定数据文件的可用性<sup>[3-4]</sup>。随着科研模式逐渐趋向于数据密集型, 在进行科学研究时对数据集进行使用并加以标引变得更加重要<sup>[5]</sup>。科学数据引用的影响因素指由于科学数据创建者或拥有者对科学数据的拥有方式、共享与被利用要求, 以及科学数据自身的领域、类型、保密性、著录要求等对科学数据引用者的引用动机、目的、规模、条件、方式与成本等产生作用的各种情形。为引导科学合理的科学数据引用, 提高科学数据的引用质量和权威程度, 有必要关注科学数据引用的关键影响因素。本文结合国内外已有研究和实践经验, 站在科学数据引用者的立场, 梳理科学数据引用的主要外部影响因素, 以为促进科学数据引用的发展提供借鉴和帮助。

## 1 国内外研究综述

目前, 国内外对于科学数据引用影响的研究不多,

基本上是从科学数据引用的发展现状与影响因素两方面进行调研。(1) 科学数据引用的发展现状。屈宝强等通过调研, 总结了目前科学数据引用在学科、引用位置、引用行为、引用格式和引用工具利用等方面的现状, 并且分析了基于引用的数据发现和评价研究状况<sup>[1]</sup>, 为今后科学数据利用中的权益分配、数据引用标准研制、数据引用实现方式、数据计量等方面的深入研究提供建议。张静蓓等从数据引用关键问题、数据引用规范、文献管理工具及数据引用利益相关者四个方面对国外数据引用研究现状进行梳理与总结, 其中数据引用关键问题包括数据共享与引用意识、引用对象的唯一标识符等问题, 文中指出引用意识与数据引用规范是推动科学数据引用发展的关键<sup>[6]</sup>。王丹丹从如何引用、引用什么以及何时引用三个角度出发, 探讨科学数据引用的关键点和难点, 并讨论了元数据、数据版本、数据粒度和文献管理工具等问题, 对研究者、出版商、数据中心和图书馆提出数据引用规范要求<sup>[7]</sup>。(2) 科技文献引用影响因素的探讨。鞠秀芳等讨论了影响科技文献引用的因素, 发现期刊的引用行为具有一定的复杂性, 其过程受引用动机、引用习惯、期刊编辑等内在动机和外在因素的影响<sup>[8]</sup>。马智峰从引用者动机、文献引用力度、信息检索工具の利用、信息资源的发现与获取、文献时效性、期刊编辑与管理等角度分析科技文献标引的影响因素, 发现文献标引规范、信息检索渠

\* 本研究得到国家社会科学基金项目“云计算环境下图书馆信息资源安全政策法律研究”(编号: 11CTQ004) 资助。

道的拓宽等可正确指导文献标引工作<sup>[9]</sup>, 提高科研水平和科技论文质量。

国外对科学数据引用的研究主要集中在三个方面。

(1) 数据引用现状。Piwowar等发现在同类数据集文献中, 开放数据集的被引率高于非开放数据集, 并且在公开数据源后, 数据重用的强度也在稳步上升<sup>[10]</sup>。(2) 基于数据引文索引(Data Citation Index, DCI)分析数据引用实践现状。Cho通过DCI分析数据引用现状, 发现遗传学和生物技术学是产生科学数据最多的学科领域, 而经济学和社会学的科学数据被引频次较高(如人口数据和就业数据); 通过数据分析发现UK Data Archive和ICPSR为被引频次最高的两个数据库; 并指出科学数据不仅可以提高研究过程的效率和透明度, 还可以通过数据整合和重新诠释创造新的科学数据<sup>[11]</sup>。

(3) 数据引用标准与数据引用最佳实践。Hapgood等提出气象数据难以与现代引用标准相结合, 并鼓励领域专家在数据政策与现实需求间寻找平衡点<sup>[12]</sup>。

目前, 国内外对科学数据引用的研究多集中在科学数据引用的实践与发展现状上, 对科学数据影响因素的研究较少, 而科学数据引用的影响因素决定科学数据的引用者是否可以引用、出于何种目的引用、以什么方式引用、引用时的著录方式等。因此, 本文从科学数据引用者角度, 指出影响科学数据引用的因素包括内因和外因, 而二者的区分是以影响引用者引用行为的主观和客观条件来划分的。其中, 主观条件包括引用者的引用动机、引用态度等, 本文将之归为影响引用者引用行为的内部因素; 而影响引用者引用行为的客观因素则包括所引用科学数据的可获得性、科学数据所涉及的领域、科学数据的类型、科学数据的引用格式、科学数据的引用版本等, 将其归为影响引用者引用行为的外部因素。现有研究较多探究科学数据引用规范、粒度、版本、唯一标识符及文献管理工具等的构成和实践发展现状, 较少阐述其对科学数据引用的影响。本文在已有研究基础上, 还将从科学数据引用政策、协议、版权、检索与发布平台等角度探讨影响科学数据引用的外部因素; 同时, 在相关研究及科学数据引用实践经验的基础上, 考虑资料的可获得性, 将影响科学数据引用的外部因素归纳为科学数据引用规范与格式、科学数据权利保护与引用许可、科学数据发布平台与管理软件三方面, 并对这三方面内容进行深入剖析, 以期为进一步研究和探讨提供参考。

## 2 科学数据引用规范与格式

科学数据引用规范与格式是对科学数据的规范性、完整性、可追踪性、可重复验证的限定。对科学数据引用者来说, 能否遵循引用规范和规定的引用格式, 与其引用的科学数据可信度具有直接关系。同时, 若科学数据引用者拟引用的科学数据存在引用版本、引用粒度、唯一标识符、元数据等著录内容不完整等现象, 则会降低其引用质量和权威程度。

### 2.1 科学数据引用规范

引用规范指对科学数据引用作出规定的文本, 是宽泛的、全面的约束性要求。本文对科学数据引用规范的讨论包括科学数据的引用规模及科学数据提交与引用规定。

#### 2.1.1 科学数据引用规模

科学数据引用规模指引用数据的类型、界限与数据量。Hayes等将科学数据引用对象分为数据库、数据集、数据产品和数据记录<sup>[13]</sup>。数据引用最常见的内容为数据集, 一个数据集下通常会包含若干子集。目前, 国际组织、学术出版社、数据中心等仍没有给出数据集的统一定义, 因此会对科学数据引用产生影响。首先, 引用者在对数据进行引用时通常存在模棱两可的现象, 而引用本身也是一个比较模糊的过程<sup>[13]</sup>; 其次, 数据集通常是动态的, 数字对象唯一标识符并不能解决此类问题<sup>[14]</sup>, 这些不确定性在一定程度上影响科学数据的引用; 再次, 数据产权问题, 数据集通常是多方合作产物, 同一个数据集有多个作者, 对这种数据的引用很容易造成知识产权纠纷, 阻碍科学数据引用实践的发展; 最后, 在科学数据引用过程中应避免个人特征的出现, 因此应注意对于敏感数据的引用。由此可见, 明确界定科学数据引用界限能够避免由于界限不清晰而带来的产权不清等问题, 能够让引用者更方便、合理地对科学数据进行引用。

#### 2.1.2 科学数据提交与引用规定

科学数据提交与引用规定指数据中心对数据创建者提交数据和数据引用者使用科学数据而制定的规

则。数据中心作为科学数据的存储及发布机构,承担数据监护的重要责任<sup>[15]</sup>。规范数据提交流程及数据内容,能够保证数据中心的数据质量,为数据引用奠定基础。Dryad明确表示其接收的数据集语言应为英文,提交形式多样,但数据大小不能超过20GB;此外,Dryad接收的数据大多经过同行评议,对于部分非同行评议但来源可靠的数据,也会考虑接收(如学位论文、专著等)<sup>[16]</sup>。

科学数据的引用可以使科学数据更容易被发现与重用,追踪数据来源并验证研究过程,目前许多数据中心都有数据引用规定,要求研究人员在使用数据时进行恰当引用。ICPSR (Inter-university Consortium for Political and Social Research)<sup>[17]</sup>和罗普中心公众舆论研究<sup>[18]</sup>提供了科学数据集引用指南,并给出具体引用案例。也有一些机构没有给出具体引用规范,仅要求对科学数据的引用进行说明,如美国国家海洋和大气管理局的国家数据中心仅请求在文本中得到承认或致谢;美国地质调查中心要求数字地图必须正式引用,而表格数据只需在致谢中说明即可<sup>[1]</sup>。由于不同数据中心对科学数据引用的要求不一致,因此产生多种科学数据引用格式,不利于引用的规范化。大量科学数据引用规范会影响研究者的使用,各数据中心应加强联系,研制统一的数据引用规范,降低引用难度,以推动科学数据引用实践的发展。

## 2.2 科学数据引用格式

引用格式指在引用科学数据时需使用的具体规格样式,如要求数据使用者通过致谢、脚注等形式提供科学数据来源信息。引用格式以引用规范为参照,而引用规范则以引用格式作为具体体现。本文对引用格式的讨论包括元数据与唯一标识符、数据版本和粒度以及与科学文献相关联。

### 2.2.1 元数据与唯一标识符

元数据是描述数据的数据,是科学数据引用的关键内容。目前,DataCite、OECD (Organization for Economic Cooperation and Development)等已出台一些面向科学数据元数据的方案,但并没有形成统一的科学数据引用标准。张静蓓等探讨目前致力于科学数据引用的国际组织所要求的核心元数据集构成(其中包括Author/Creator, Publication date, Title,

Publisher/Archive, URL, Persistent Identifier等),数据中心或其他机构可根据数据特点选择元数据项进行描述<sup>[15]</sup>。许多学科都有元数据标准,但对元数据标准的遵循还需各利益相关主体的支持与协作,只有科学数据创建者、科学数据管理机构及行业组织、科学数据引用者对元数据标准达成一致,才能最大限度地促进数据的引用和再使用。在此过程中,出版商的作用明显,《科学》和《自然》对基因数据引用采取的强制性政策是很好的例证<sup>[7]</sup>。元数据标准的制定应该在通用性与专用性间寻找平衡,既不能过少以致数据描述不完整难以被准确识别,也不能过多导致冗余。

唯一标识符指唯一标识数据集的名称或字符串,其独立于数据并永久标识数据<sup>[19]</sup>。《数据引用共同原则》提出数据要拥有唯一标识性,数据引用应该拥有一种机器可识别的、全球唯一的、能够被广泛使用的可持续方法<sup>[20]</sup>。使用唯一标识符能精确定位数据集,减少研究者工作量。Honor等指出数据共享和重用虽被广泛认可,但缺乏被广泛接受的数据引用标准,难于追踪科学数据使用情况并衡量其影响<sup>[21]</sup>。针对唯一标识符的分配问题,仍有许多争议,这需要标识符注册机构、数据中心、出版社等多方主体共同协商解决。唯一标识符能精确识别数据使用者所引用的数据集,保护数据创建者的著作权,并且对科学数据引用的规范性、完整性、可追踪性、可重复验证性具有重要意义。

### 2.2.2 数据版本和粒度

数据版本指由于数据变动使数据集表现出不同状态,如随着时间的推移数据量不断增加,或应用新的研究方法产生新数据等。在数据版本不断变化的情况下,如何保证数据的可验证性成为影响科学数据引用的关键。数据中心对动态数据集的维护也是影响科学数据引用的一个难题。如Argo Project<sup>[22]</sup>,用3 000多个设备探测海水温度和盐度,每年产生10万条数据,该数据规模需要相当大存储量的数据中心才能承担。王丹丹总结了目前引用大规模动态数据集的两种方法,一是完整的数据集分配标识符,这种情况下的数据集只有一个最终版,存在数据收集期间不能被引用的问题;二是制定基础版,对新增数据进行补充,针对这种情况,国家冰雪数据中心建议由研究者决定基础数据版本,英国国家资料库则对定义新版本的情况进行说明<sup>[7]</sup>。而OECD则制定动态数据集元数据来解决动态数

数据集的维护问题,其元数据项包含数据集更新时间、数据集常规更新频率等<sup>[19]</sup>。就引用数据而言,需要研究者和数据中心协商制定数据版本和引用不同版本数据的规范和方法<sup>[7]</sup>。科学数据的及时更新,能够提高引用质量与权威程度,但由此带来的引用版本信息的揭示同样需要引起重视。

数据粒度指数据引用的层次<sup>[7]</sup>。一个数据集往往可划分为多个子集,而一个数据集可能又是另一个大数据集的子集,数据间的层次关系较多变与复杂,这种特性也让引用者对数据引用产生困扰。由于不同学科引用粒度存在差异,因此数据中心需明确适合其数据的最佳引用单元。如在生物学科中,共有15个数据聚合器,却有11种不同的引用格式,这会使研究者感到困惑<sup>[23]</sup>。为解决数据引用的粒度问题,需要科学数据引用规则或政策的制定者给出合理、明确的标准规范,来帮助数据引用者更好地引用数据,同时也需要数据引用管理工具协助生成对各层次数据的引用。如PANGAEA针对不同层次数据引用设计相应的引用格式;DCC认为数据粒度应满足研究者需求,可在文章引用数据处提供所引数据集细节;GBIF希望找到一种既满足版本问题又满足粒度问题的解决方案,并强化工具的整合,以支持复杂数据的引用。

数据的版本与粒度反映数据更新、修订及引用层次问题,在引用时通过对版本信息、引用粒度的注明,方便为审稿人和使用者(包括科技文献读者、对同一科学数据进行引用的其他学者)提供比较完整的科学数据信息,保证科学数据引用的规范性、完整性、可追踪性及可重复验证性。

### 2.2.3 与科学文献相关联

将科学数据与科学文献相关联,可以使读者了解数据的使用情况,为科学研究进展提供借鉴;同时,也可以为数据发布平台设计更好的服务模式提供基础。根据欧盟PARSE.Insight研究显示,超过84%的科学家认为将科学数据与文献进行关联是有用的<sup>[1]</sup>。目前已有部分科学数据与科学文献实现关联,如通过PubMed检索号实现GenBank到PubMed的链接,以保证在GenBank中使用检索基因序列能够查找到相关文献<sup>[24]</sup>。通过数据与文献的双向链接,读者在阅读文献时,既可以访问原始数据甚至重复其研究过程,也能从数据出发找到与其相关的所有文献,由此提高科研效率。

## 3 科学数据权利保护与引用许可

科学数据权利保护与引用许可能够明确数据创建者,如果数据创建者或拥有者声明他人不可以擅自利用其科学数据,那么科学数据使用者进行引用时就需要经过创建者的允许。本文对科学数据权利保护与引用许可的讨论包括科学数据创建者的态度与立场、科学数据版权保护的法律、科学数据的引用政策以及科学数据的引用协议。

### 3.1 科学数据创建者的态度与立场

科学数据创建者的态度与立场指创建者是否同意他人使用其数据并对其进行引用。数据创建者在科研活动中收集或创造科学数据,对其进行描述、组织,并上传至数据中心,使其得到传播与重用。科学数据的产生需要耗费创建者大量的时间与精力,因此需要制定一定的科学数据奖励机制,使数据创建者的工作得到认可,以此鼓励创建者分享数据。张静蓓等总结出影响科学数据共享的制度因素(数据引用政策与规范等)、技术因素(统一的元数据标准、合适的数据库等)和个人因素(个人预期利益及共享可能带来的风险等)<sup>[25]</sup>。科研工作者对科学数据的共享是科学数据引用及重用的前提。科学数据创建者对数据共享的态度在很大程度上决定科学数据引用的进程和发展。Candela等指出,在学术交流过程中,科研人员应减少科学数据的交流阻碍,支持其流动和发展<sup>[26]</sup>。也有学者对科研人员共享科学数据的态度进行调研,结果显示科研人员对数据共享持积极态度<sup>[27]</sup>。科学数据创建者对科学数据引用的许可是科学数据引用实践的基石,对推动科学数据引用实践发展有重要作用。

### 3.2 科学数据版权保护的法律

科学数据版权指数据的权利归属,而科学数据版权保护的法律则将科学数据的版权属性纳入法律保护范围。目前,从国内外发展实践来看,科学数据的保护需要相应的法律法规作支撑;从世界各国现行的立法实践来看,科学数据并不是各国现行版权法律所完全认可的保护标的物。一方面,科学数据的创建凝聚着创建者的智慧,及设备、资金方面的大量投入;另一方面,从版权立法的初衷来看,平衡版权作品创建者

和版权作品使用者的利益关系,需要给予科学数据创建者必要的、在限定时间内的、适当形式的专有权,并加以排他性的保护。这种保护是对其创建科学数据所投入的财力与智力成本的必要弥补。而从科学数据的广泛传播和最大利用效益来看,科学数据一旦创建,需要被更多的科学数据使用者知晓、检索和再利用,以促进人类科学数据研究的进步。目前,虽然国内外高度重视科学数据的再利用,但在现有各国版权法律未能给创建者提供足够保护前提下,借助科学数据创建者对科学数据利用与共享的立场与态度,以及由此体现的数据共享协议的细化要求,可以在保护科学数据创建者合法权益的同时,促进科学数据的引用。因此,科学数据引用者在进行科学数据引用时,需充分尊重科学数据创建者或拥有者的权益诉求,参照现行版权法律规定,避免侵权行为的发生。若科学数据创建者或拥有者被禁止对科学数据在某种情形或某种规模、内容上的使用,则科学数据施引者应放弃其使用意图。

### 3.3 科学数据的引用政策

科学数据的引用政策指致力于科学数据管理的组织/机构对科学数据引用制定的原则和要求。科学数据引用的用途、再共享等方面的自由与方便程度,与科学数据的权利保护与引用许可密切相关。由于科学数据引用具有一定的政策约束,因此科学数据使用者对数据进行引用时必须遵循政策规定。目前,许多致力于科学数据引用的相关组织/机构(如DataCite<sup>[28]</sup>、DCC<sup>[29]</sup>、ANDS<sup>[30]</sup>、ESIP<sup>[31]</sup>、CODATA<sup>[32]</sup>和RDA<sup>[33]</sup>)都发布了科学数据引用政策,有助于对科学数据引用行为进行约束和规范,形成良好的科学数据引用氛围。与此同时,科学数据引用也需要建立一定的奖励机制,对数据创建者给予奖励可以激励科学数据的共享,对科学数据引用者的奖励(如优先发表)也可以激励其收集或创造更多的数据。王丹丹认为理想的数据引用应与传统文献引用方式一样,在参考文献部分对数据进行引用标注,并将引用排名纳入科研评价体系<sup>[7]</sup>。张静蓓等针对科学数据缺乏共享意识和引用意识的问题,提出应完善学术报偿体系,将数据引用纳入科研评价体系<sup>[6]</sup>。权威机构通过科学数据引用政策对科学数据引用行为进行引导与约束,形成良好的科学数据引用氛围,能够促进科学数据引用实践的进一步发展。

### 3.4 科学数据的引用协议

科学数据引用协议是能够将科学数据创建者与科学数据使用者对科学数据引用的权利和义务达成一致意见的一种约定。由于现有版权法等相关法律法规及引用、共享政策不能完全保护科学数据创建者的权利。因此,科学数据共享中心按照数据类型特点制定了相应的数据许可使用协议或创作共享协议,以约定使用者承认并尊重创建者的权利。数据许可协议提供双方共同认可的约定,保证科学数据引用双方的基本利益。在科学数据引用协议中,大部分协议明确提出数据引用的要求,并且提出要遵循特定数据引用格式规范。例如,UK Data Service的数据许可使用协议包含科学数据的引用格式、使用申请机制、使用和用途规定及特殊协议保护措施等内容<sup>[34]</sup>。协议要求数据使用者在引用数据时对数据进行规范性标识,并将数据按照创建者意愿分为“open data”“safeguarded data”“controlled data”。其中,“open data”没有使用限制,“safeguarded data”与“controlled data”都有单独的许可使用条款,对数据的用途进行了规定;“safeguarded data”只能用于科研目的,“controlled data”需要引用者提交使用申请,待数据创建者审核通过并将其加入特定用户组后,才可获得使用。在许可协议条件下,科学数据的引用者需尊重数据创建者的权利,严格按照协议内容对数据进行使用;此外,引用协议对衍生数据也有控制权。衍生数据是在原有数据基础上进行补充、加工等行为后产生的数据,因此科学数据引用者使用衍生数据时也需遵循相关的引用协议。

## 4 科学数据发布平台与管理软件

本文对科学数据发布平台与管理软件的讨论包括科学数据检索与发布平台、科学数据管理软件与发现工具。

### 4.1 科学数据检索与发布平台

科学数据检索与发布平台是能够对科学数据进行存储、描述、检索、获取、共享的数据库。科学数据的引用需要由科学数据的发布平台作为科学数据创建者与数据引用者的沟通桥梁,科学数据引用者通过发布平

台可以集中、方便地检索,发现并下载所需引用的科学数据。从国内外发展情况来看,科学数据的发布平台主要有单一科学数据发布平台和科技文献与科学数据的混合发布平台。

单一科学数据发布平台包括国家数据中心和机构数据库。国家数据中心,如国家地球科学数据中心、环境信息数据中心、国家冰雪数据中心等;机构数据库,如罗普中心公众舆论研究、哈佛大学Harvard Dataverse等。科技文献与科学数据混合发布平台,如Dryad、ICPSR、Figshare和GenBank等。

## 4.2 科学数据管理软件与发现工具

科学数据引用软件可以使数据使用者更方便地生成引用规范,科学数据发现工具可以使数据使用者更容易发现与检索所需数据集。

Kotarski等对常用文献管理软件进行调研,指出在18种常用文献管理软件中,只有3种软件(EndNote、Papers和Sente)具备科学数据引用管理功能,分别通过“数据库”“表格”和“数据文件”选等选项进行引用管理;其余15种软件中,有5种(Bibus、Bookends、JabRef、Reference Manager和Scholar's Aid)允许用户自定义引用类型,如“数据集”选项,但并没有专门针对科学数据引用的元数据;另外10种软件(Biblioscope、Citavi、Docear、Mendeley、Pybiographer、Qiqqa、Rebase、Ref Works、Wiz Folio和Zotero)则根本没有涉及科学数据选项<sup>[35]</sup>。目前,常用的文献管理软件缺乏对数据引用的支持,对科学数据等非文本资源类型引用的关注度还不够。随着参考文献管理工具的广泛应用,确保科学数据的引用管理作为其功能之一将有助于使用者更规范、便捷地引用数据。

汤森路透于2012年推出DCI<sup>[36]</sup>,对推进科学数据引用标准统一和增强数据价值有重要意义。DCI的数据来源于数据库,它为数据集建立索引,使其能够与文献相链接。Robinson-garcía等对DCI数据引用记录进行分析,发现虽然科学和工程领域的科技数据集引用最多,但在社会科学和艺术与人文研究中,科学数据的作用更大;同时,有88.1%的数据没有得到引用<sup>[37]</sup>。通过对科学数据发现工具的研究,能够促进全球科学数据的发现与引用,寻找科学数据引用规律,使其更好地为科学研究服务。

## 5 结语

随着科学技术的发展以及科学数据在科研活动中的广泛使用,对科学数据引用行为的研究越来越迫切。科学数据引用不仅体现对科学数据创建者基本权利的尊重,也为科学数据共享提供依据与保障。为提高科学数据引用质量与权威程度,本文基于已有研究与实践经验,梳理出影响科学数据引用的外部因素主要包括科学数据引用规范与格式、科学数据权利保护与引用许可、科学数据发布平台与管理软件,以期为科学数据引用研究的发展提供借鉴。

## 参考文献

- [1] 屈宝强,王凯.科学数据引用现状和研究进展[J].情报理论与实践,2016(5):134-138,114.
- [2] WU Z X.Data citation resources[EB/OL].[2017-03-31].[http://www.ands.org.au/cite-data/resources.html#What\\_is\\_Data\\_Citaiton](http://www.ands.org.au/cite-data/resources.html#What_is_Data_Citaiton).
- [3] DODD S A.Bibliographic references for numeric social science data files:suggestedguidelines[J].Journal of the Association for Information Science and Technology,1979,30(2):77-82.
- [4] WHITE H D.Citation analysis of data files use[J].Library Trends,1982,30(3):885-888.
- [5] BALL A, DUKE M.Data citation and linking[EB/OL].(2012-06-21)[2017-03-21].<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking>.
- [6] 张静蓓,吕俊生,田野.国外科学数据引用研究进展[J].图书情报工作,2014(8):91-95,138.
- [7] 王丹丹.科学数据规范引用关键问题探析[J].图书情报工作,2015(8):42-47,53.
- [8] 鞠秀芳,郑彦宁,潘云涛.期刊引用操纵行为研究综述[J].西南民族大学学报(人文社会科学版),2013(4):224-228.
- [9] 马智峰.参考文献的引用及影响引用的因素分析[J].编辑学报,2009(1):23-25.
- [10] PIWOWAR H A,VISION T J.Data reuse and the open data citation advantage[J].Peerj,2013,1(3):e175.
- [11] CHO J.Study about research data citation based on DCI(Data Citation Index)[J].Journal of the Korean Library and Information Science Society,2016,50(1):189-207.
- [12] HAPGOOD M,KNIPP D.Data citation and availability:striking a balance between the-ideal and the practical[J].Space Weather-the International Journal of Research & Applications,2016,14(11):919-920.

- [13] HAYES P J, HALPIN H, SHETH A. In defense of ambiguity[J]. International Journal on Semantic Web & Information Systems, 2008, 4(4): 1-18.
- [14] DUERR R E, DOWNS R R, TILMES C, et al. On the utility of identification schemes for digital earth science data: an assessment and recommendations[J]. Earth Science Informatics, 2011, 4(3): 139.
- [15] 张静蓓, 田野, 吕俊生. 科学数据引用规范研究进展[J]. 图书与情报, 2014(5): 100-104.
- [16] DRYAD. What kinds of data does Dryad accept?[EB/OL]. [2017-03-31]. <http://datadryad.org/pages/faq#depositing-acceptable-data>.
- [17] ICPSR. Citing Data[EB/OL]. [2017-03-31]. <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/citations.html>.
- [18] Roper Center. How to cite Roper Center data[EB/OL]. [2017-03-21]. [http://www.ropercenter.uconn.edu/data\\_access/data/how\\_to\\_cite.html](http://www.ropercenter.uconn.edu/data_access/data/how_to_cite.html).
- [19] 李丹丹, 吴振新. 研究数据引用研究[J]. 图书馆杂志, 2013(5): 65-71.
- [20] 宋宇, 真溱, 汤珊红. 数据引用的共同原则[J]. 情报理论与实践, 2015(8): 145.
- [21] HONOR L B, HASELGROVE C, FRAZIER J A, et al. Corrigendum: data citation in neuroimaging: proposed best practices for data identification and attribution[J]. Frontiers in Neuroinformatics, 2016, 10: 43.
- [22] Project Argo[EB/OL]. [2017-03-31]. <http://argoproject.org/>.
- [23] LANEM A. Data citation in the electronic environment[EB/OL]. [2017-03-31]. <http://www.gbif.org/resource/80798>.
- [24] 韩涛. 科学数据与科学文献相关性研究——以生物信息学为例[J]. 图书情报知识, 2008(3): 42-46.
- [25] 张静蓓, 吕俊生, 田野. 国外数据共享行为影响因素研究综述[J]. 图书情报工作, 2014(4): 136-142.
- [26] CANDELA L, CASTELLI D, MANGHI P, et al. Datajournals: a survey[J]. Journal of the Association for Information Science and Technology, 2015, 66(9): 1747-1762.
- [27] 何琳, 常颖聪. 科研人员数据共享意愿研究[J]. 图书与情报, 2014(5): 125-131.
- [28] DataCite. Citation formatter[EB/OL]. [2017-03-31]. <https://www.datacite.org/citation.html>.
- [29] DCC. Cite Datasets[EB/OL]. [2017-03-31]. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>.
- [30] ANDS. Citation and identifiers[EB/OL]. [2017-03-31]. <http://www.ands.org.au/working-with-data/citation-and-identifiers>.
- [31] ESIP[EB/OL]. [2017-03-31]. <http://www.esipfed.org/>.
- [32] CODATA-ICSTI data citation standards and practices[EB/OL]. [2017-03-31]. <http://www.codata.org/task-groups/data-citation-standards-and-practices>.
- [33] RDA. RDA events[EB/OL]. [2017-03-31]. <https://www.rd-alliance.org/>.
- [34] UK data service[EB/OL]. [2017-03-31]. <https://www.ukdataservice.ac.uk/>.
- [35] KOTARSKI R, REILLY S, SCHRIMPF S, et al. Report on best practices for citability of data and on evolving roles in scholarly communication[EB/OL]. [2017-03-31]. <http://rep4-vm.awi.de/31396/1/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf>.
- [36] 汤森路透发布Data Citation Index促进对全球研究数据的发现[EB/OL]. [2017-03-31]. <http://www.thomsonscientific.com.cn/press/press20121102/>.
- [37] ROBINSON-GARCÍA N, JIMÉNEZ-CONTRERAS E, TORRES-SALINAS D. Analyzing data citation practices using the data citation index[J]. Journal of the Association for Information Science & Technology, 2015, 67(12): 2964-2975.

## 作者简介

黄国彬, 男, 1979年生, 博士, 副教授, 硕士生导师, 研究方向: 信息法学、信息分析, E-mail: huanggb@bnu.edu.cn。  
刘馨然, 女, 1993年生, 硕士研究生, 研究方向: 信息分析, E-mail: bnuliuxinran@163.com。  
姜颖, 女, 1984年生, 硕士, 馆员, 研究方向: 信息素养教育, E-mail: uibeijing@126.com。

## Analysis of External Influencing Factors of Scientific Data Citation

HUANG GuoBin<sup>1</sup>, LIU XinRan<sup>1</sup>, JIANG Ying<sup>2</sup>

(1. School of Government, Beijing Normal University, Beijing 100875, China;

2. Library of University of International Business and Economics, Beijing 100029, China)

**Abstract:** In order to guide scientific and reasonable scientific data citation, to improve the quality and authority of scientific data, it is necessary to pay attention to the key influencing factors of scientific data citation. By using the literature research method and website research method, combining with the existing research and practical experience at home and abroad, this paper sorts out the main external influencing factors of scientific data citation in scientific data referrers' view. The external influencing factors of scientific data citation include scientific data citation standards and formats, scientific data rights protection and reference permission, scientific data platform and management software. Effective processing and response to these three aspects are the main consideration for scientific data referrers to standardize their behaviors.

**Keywords:** Scientific Data Citation; Influencing Factors; Citation Standards; Citation License

(收稿日期: 2017-04-18)