科学数据规范引用关键问题探析*

王丹丹
河南科技大学管理学院　洛阳 471023

摘要：[目的/意义]明确实现规范引用需要解决的基本问题、重点和难点问题，并分析相关利益群体在这一过程中需要承担的责任和未来努力的方向，以引导其更好地厘清思路，推动实践。[方法/过程]借鉴国内外相关研究及实践经验，从科学数据规范引用价值认同入手，从如何引用、引用什么和何时引用三方面阐述科学数据规范引用的基本问题与难点。以此为基础，分析数据规范引用给相关利益群体带来的机遇和提出的要求。[结果/结论]理想的科学数据引用规范应该明确回答如何引用、引用什么和何时引用 3 个基本问题。如何引用，就是需要制定面向科学数据的元数据规范、建立能够承认作者贡献的评价体系并提供支持科学数据引用的参考文献管理工具；引用什么，就是需要明确引用数据的版本、粒度、数据验证方式以及引用对象等问题；何时引用，就是需要厘清使用数据的情景。相关利益群体应该看到科学数据规范引用给各自带来的机遇以及提出的要求，不断调整以推动数据引用的实践。

关键词：科学数据　数据引用　数据出版
分类号：G250
DOI：10.13266/j.issn.0252-3116.2015.08.006

数据出版是指在互联网上公开数据，并且支持除数据提供者之外的研究人员或者组织机构下载、分析、再利用以及引用数据。从广义上讲，任何将数据上传到互联网或者数据库并支持开放获取的行为都可以称之为“数据出版”。对科学数据而言，数据出版是指数据达到可引用和可追溯的状态，核心内容是为数据提供标准的数据引用格式和永久访问地址[1]。因此，数据引用是推动数据出版的先决条件。然而，数据引用是一个崭新的研究领域，在目前还处在一个从无到有的状态。对实现数据引用的几个关键问题进行阐释、探讨，有助于我们更好地理清思路，推动实践。

1　数据规范引用的意义

数据引用(data citation)是提供数据参考的实践，类似于研究者在论文中对文献的引用。因此，在很大程度上，标识数据的来源是传统出版物引用传统的延续，其目标是对数据创建者所作出的贡献给予承认，并表明数据的可利用性[2]。实现数据规范引用是基于以下 3 个方面的考虑：

1.1　使数据创建者的贡献得到承认

在传统意义上，数据不会被出版，已出版的文献中也没有正式承认数据的作用。然而，为了使数据真正可获取，数据创建者需要花费时间和精力提供充分的元数据并监管、保存数据。因此，创建者最关心的问题是这样做能够为其带来哪些益处。实现数据规范引用的初衷，就是想将传统的文献评价体系引入数据领域，使用类似引文的评价方式，对数据创建者以及相关利益群体所做出的贡献给予承认并量化贡献度，即提供一种使所有为数据产生付出努力的人的工作都得到认可的系统。因此，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，将引用排名纳入科学评价体系[1,3]。

1.2　实现数据使用的溯源

开展新的研究首先要理解先前的工作是如何产生的。引用为现有的工作提供了该领域发展进步的线索，是理解某一主题随着时间产生变化的重要途径。数据引用亦是如此。同文献引用类似，数据引用提供了一种数据定位或参考机制，有利于读者快速定位并

*本文系国家社会科学基金项目“数字图书馆用户数据资源化管理研究”(项目编号：2013CTQ012)和河南省教育科学规划课题“高等教育领域开放用户数据实践与应用研究”(项目编号：[2014]-JKGHC-0085)阶段性研究成果之一。
作者简介：王丹丹(ORCID：0000-0001-9214-4906)，副教授，E-mail：wangdan430806@163.com。
收稿日期：2015-02-26　修回日期：2015-03-12　本文起止页码：42-47，53　本文责任编辑：杜杏叶

获取数据,增加了数据重用性和共享能力,也提升了信息利用的效率与收益,有助于加快科学发现与创新的步伐[4]。比如通过数据与文献的互操作,读者在阅读文献的时候,既可以访问论文的原始数据甚至重复研究的过程,也能够从数据开始找到与之相关的所有文献,由此提高科学生产力[5-6]。

1.3 提供验证科研过程的途径

允许其他人获取基本数据有助于对数据和假设进行重新解读,也有助于发现任何可能存在的错误或不一致之处。通过数据引用获取研究过程的原始数据集,按照实验步骤进行操作,可以重现研究过程,验证科学研究的结果[7]。

2 数据规范引用的基本问题与难点

针对科学数据引用的需求,一些数据中心已经制定了相应的数据引用规范。如以帮助研究人员发布、挖掘、访问和利用科学数据为目的的澳大利亚国际数据服务中心(Australian National Data Service,ANDS)的作者指南为研究人员提供了以下指导:"发布我的数据(Publish My Data)",帮助研究员用元数据出版研究数据;"注册我的数据(Register My Data)",给研究人员的数据提供永久标识符[8];"识别我的数据(Identify My Data)",协助研究人员和研究机构宣传其研究数据;"引用我的数据(Cite My Data)",提供具体的数据引用规范[9-10]。英国数字化监管中心(Digital Curation Centre, DCC)发布了数据引用与链接指南[11]。这两个指南都是基于"如何引用、引用什么、何时引用"3 个基本问题来阐述数据引用的要求的[12]。

2.1 如何引用

如何引用的问题,具体而言就是需要制定面向科学数据的元数据规范、建立能够承认作者贡献的评价体系并提供支持科学数据引用的参考文献管理工具。

2.1.1 元数据 尽管数据引用是传统文献引用的延续,但是与传统文献不同,数据自身具有不同的结构和格式,因此数据作为引用对象可以有许多不同的引用方式。引用的关键内容是描述数据的信息——元数据,已经出现了一些面向科学数据的元数据推荐方案,如来自国际组织 DataCite[13]、OECD(The Organisation for Economic Co-operation and Development)[14] 和 DDI(Data Documentation Initiative)的方案。数据引用实质上是引用对象元数据的子集。对于数据集而言,数据标题和数据发布者(出版商或拥有数据的机构)对于发现和获取数据集本身帮助不大,除非明确知道数据的具体位置,例如知道其在某个数据库或具体学科知识库中。因此对于数据引用而言,提供每一条数据的位置就非常重要。网址是不可靠的,因为网站可以改变、移动甚至整个消失。但是唯一标识符是固定不变的,因此可以作为获取数据的永久依据,DataCite 提供这样的服务。在现有的引用规范中,最常用到的标识符是 DOI,其次是句柄系统 Handles(DOI 系统基于此构建)。大多数数据引用规范建议使用 DOI,这是因为 DOI 已被广泛采用,实施障碍较小。然而,标识符固然重要,文化观念的转变也同样重要,例如天文学领域为数据提供了唯一标识符,使其可以链接到论文,结果却以失败告终,其原因一方面是研究者缺乏意识,另一方面是使用这一系统的出版商和数据中心的数量不够多。

2.1.2 承认贡献 数据署名权是知识产权中的一项基本权利,数据引用是体现数据署名权的最佳方案[3]。

为了使数据创建者的工作得到承认,应将数据创建者作为另外一个重要的元数据元素。但是目前面临两个困难:一是排序问题。如果按照姓名首字母排序,那么无法体现做出最大贡献的创建者。同时,如何界定贡献的顺序也是一个很有争议的话题,因此数据中心应该就数据著作权的分配给出详细的说明;二是当融合多个数据集生成一个新数据集时,会产生所有权叠加的问题。在这种情况下,单独引用每一个独立的数据集就不再合理,而在引用记录中列出数据创建者的数目也不切实际[15]。与此同时,许多数据中心也要求通过数据引用得到承认、证明其存在的价值。作为整合生物多样性数据的全球基础设施机构,GBIF(Global Biodiversity Information Facility)深入考虑了这一问题,推荐的元数据不仅包括出版商,也包括贡献者[16]。

2.1.3 支持工具 目前已有许多帮助用户高效管理和快速生成参考文献的文献管理软件,可以对各种资源进行引用。其中大多数软件都有增加新参考类型的选项,例如期刊论文、书或网站。引用所需要的元数据依据所选择的类型而变化,写作时表现为将不同的格式增加到文献引用中。随着这些工具的广泛应用,确保数据作为引用选项之一将有助于研究者更容易地引用数据。调查表明,目前很少有工具提供预先定义的数据选项,尽管许多工具允许设定个性化的字段选项,可以用来定义数据引用的格式,但很少有专门集成科学数据引文类型的软件[17]。

2.2 引用什么

不同数据对象有不同的结构和格式,即便在同一学科中也存在较大差异。这取决于研究本身,可以是

图像、数值、文本数据，也可以是物理对象、从单个工具、成千上万个传感器或个人观察中所产生的少量的数据点或成千上万条记录。在这种差异下，确定需要引用什么对于研究者而言就变得十分困难。有学者将这一问题总结为数据身份，指出为确保数据的可证实性和研究的可重复性，需进一步明确所引用数据的身份，以实现对数据对象的准确引用[18]。然而，数据集的界定在不同学科中存在差异，在不同方法中存在差异，在不同研究对象之间也存在差异。这就需要在制定引用规范时考虑以下几个基本问题：

2.2.1 数据版本 数据再生产能力是指，当引用数据的时候，数据对未来需要证实这一工作的研究者也同样应该是可用的。然而，确保数据保持不变不现实，因为太多原因都会造成数据的改变，比如新方法或技术优化可能造成原始数据的更新，或者数据是随着观察时间的变化以不断增加的方式产生的。以社会科学领域为例，纵向的研究会收集多年的数据。因此，社会科学数据知识库 GESIS 或 UK Data 存储每一次附加或波动的数据作为单个可引用的对象。然而，对于数据知识库而言，如何维护每一个动态数据集是一个未解的难题。比如 Argo Project 项目用大约 3 500 个类似于鱼雷的设备测量海水的温度和盐度[19]，每年产生 10 万条数据记录。就这样的规模而言，存储独立版本的数据所需要的存储量远不是现有的知识库所能负担的。

目前，引用大规模的动态数据集有两种方法：一是英国大气数据中心（the British Atmospheric Data Centre，BADC）所采用的、为完整且不可改变的数据集分配标识符的方法。在这种情况下，数据不存在版本问题。如果收集数据持续很长时间，那么在数据收集期间数据不具备可引用性；二是针对定期更新的数据，记录基础数据及随后的变化[20]。在这种情况下，所引用的数据是基础数据和随后变化的结合体[21]。只是当数据产生变化时，如何定义新版本成为一个新的难题。国家冰雪数据中心（the National Snow and Ice Data Centre，NSID）的建议是由研究者自己来决定什么构成了最主要和最基本的版本。英国国家资料库（UK Data Archives，UKDA）的指南对什么情况下需要定义新版本、什么情况下不需要进行了详细的说明[22]。总之，就引用数据而言，版本应该是一个可以克服的问题，但是需要研究者和数据中心协商制定有关版本和如何引用具有版本的数据的规范和方法。

2.2.2 数据粒度 数据引用规范的政策制定者应与相关科研人员就数据引用的粒度问题达成协议。展示数据引用的粒度非常重要，因为数据集的再生产性和可发现性可能因此受到消极影响。

P. Buneman 以国际药理学联盟（International Union of Basic and Clinical Pharmacology，IUPHAR）数据库为例，讨论了数据粒度问题，认为属于某一个数据知识库或中心的数据集，可以被组合成复合数据集，也可以被分解为各个子集。在各种不同粒度层面上的引用需要清楚地标明数据中心以及数据集的结构，因此数据中心有责任为每一个可利用数据层面设计引用格式。在引用时，可能是一个完整的数据知识库，可能是数据知识库中一个记录集，也可能是特定的数据记录[23]。这意味着，引用工具应支持对所需要的任何层次进行引用。地球环境科学领域的 Pangaea 设计了针对不同层次数据引用的引用格式[24]。

M. Altman 则认为数据粒度问题本质上是数据深度引用的问题，建议在简单的水平上进行引用，即数据的子集的引用可以是引用数据集整体并且在引用具体内容中描述，类似于期刊论文的某些引用元素。例如，若引用论文中的数字，则把主要文本中给出数字的论文作为一个整体来引用[18]。DCC 综合上述两种观点，指出数据引用的粒度应以满足研究者需求为准，若不够确切，则在文章中引用数据的地方提供所引用数据集的细节[25]。为支持对更复杂数据的获取，GBIF 正在寻找一种新的解决方案，既考虑版本又考虑粒度问题[26]。

2.2.3 数据验证 考虑到数据引用的版本和粒度问题，对研究者而言能够证实在出版物中引用的数据是同一数据也非常重要。很多科研人员强调需要引用未产生变化的数据，但是采用什么方法保证这一点很少提及[7]。M. Altman 的建议是使用统一数字指纹 UNF（universal numeric fingerprints）。UNF 算法产生了一个短的、对数据而言独一无二的字符串，概括总结数据的内容，并具有独立的格式。如果数据自身的任何元素发生了变化（简单的在软件系统或操作系统间移动的情况除外），UNF 将不同。使用 UNF 不仅可以证实不同文献中所引用的数据是否是同一个，也可满足数据隐私或匿名化处理的要求。然而 UNF 目前并没有被广泛采纳[20]。

2.2.4 引用数据或数据论文 数据可以以原始数据集的方式出版，也可以和数据论文一起出版。数据论文是一种学术出版物，用可检索的元数据记录描述一种特殊的在线可获取的数据集，或一组数据集，根据标准化的学术实践的要求进行出版[27]。已经有许多期刊开始出版数据论文，但是并没有看到有关什么时候

应该引用数据论文而非数据集的建议。与引用数据集相比,引用数据论文可以给数据创建者带来更高的声望。此外,由于数据论文可以提供有关数据集质量的更多的证据,更丰富的元数据也有助于促进数据的再使用,与简单地通过数据知识库使数据可获取相比,数据论文也为同行评议数据提供了一个机会。因此,对数据论文的引用被称之为"金色引用",对数据的引用被称之为"银色引用"。由于数据论文自身引用数据集(使其再使用时可发现),因此,存储的数据集不再需要被引用。但是有一些数据库如 Dryad 不但要引用原始文献,也要引用 Dryad 数据包。

2.3 何时引用

理想的情况下,对所有相关的先前的研究都应该进行引用,以明了这些研究成果在新知识产生过程中的哪一个部分发挥了作用。为了确定合适的引用方案,需要考虑以下几种所需引用的数据的情形:

一是论文中包含解释的数据。鉴于这类数据很难被定位和引用,美国国家信息标准组织(The National Information Standards Organization,NISO)的建议是将文章整体作为引用对象,分配一个 DOI 来将该内容与文章进行链接[28]。

二是作为论文补充材料的数据。NISO 的建议是如果该数据是文章的必要内容,那么在引用数据的同时需要将所属文章也作为引用对象。考虑到长期保存的需要以及减少同行评议者评阅大规模补充材料的负担,一些出版商对可以作为补充材料的数据进行了严格的限制[29]。

三是存储在仓储中的数据以及在数据期刊中发表的数据。对这一类型的数据进行引用,可以有效提升数据的可发现性,并实现数据的关联,提供有关数据的情景信息。很多作者指南都鼓励作者在指定的数据仓储中存储数据,并为其数据添加访问号或 DOI。关于引用位置,不同指南有不同规定。有些建议将数据引用和文献引用一样放置在参考文献列表中,有些规定单独设置数据引用部分。对于更详细的描述数据集及其创建过程的数据期刊中的论文,应该同时引用数据论文及其原始数据集[30]。

四是未出版的私有数据,由于缺少指向链接或提供访问这些数据的入口,因此目前来看尚没有合适的解决方案。

3 数据规范引用对相关利益群体的要求

建立良好的数据引用机制是研究者、出版商、数据中心或图书馆的共同责任。因此,在推进数据引用实践的过程中,需要这些利益相关群体不断转变自身角色:

3.1 对研究者的要求

首先,应该认同数据引用规范。在依赖观察数据,特别是基于大规模的公共设施产生的数据(卫星、天气、海域数据),或产生成本很高的数据(大强子相撞数据)进行研究的学科,对于如何处理和参考可获取的数据已经形成了最佳的实践,例如在生命科学领域,数据共享和引用的实践走在了前列,推动并建立百慕大原则。基于该原则,基因序列数据在产生后 24 小时内实现公开发布。在其他学科这一现象很少见,研究者通常按照出版商的政策规定来处理数据。为保证出版政策与现有实践相匹配,需要研究者社群的广泛参与。而作为研究者以及出版商的代表的学协会(Learned Societies),在统一标准、减少差异方面可以发挥作用。学协会应该积极采取行动,基于现有实践制定涵盖其整个学科范围的引用规范。此外,学协会出版的研究论文大约占整个论文数量的 30% 以上,因此其出版政策在提高研究者数据引用意识方面也可以发挥一定的作用[31]。

其次,应积极创建数据集元数据,并和数据一起提交到数据知识库中。基本元数据(最小数据集)就是包含在引用中的元素。也正因为如此,数据集需要唯一标识符,以使其可以很容易从相关论文和数据标题中区分出来。许多学科都有自己的元数据标准,然而如何遵照元数据标准还需要得到相应的支持,这也正是图书馆和数据中心可以提供支持的领域。

最后,应有意识地使用唯一标识符,比如 DOI。对研究者而言,若唯一标识符使用起来无障碍,而且使用的益处也非常明确,那么应无条件地积极使用唯一标识符。同时,参考文献管理软件应该设置研究数据集选项,提供 DOI 或其他的唯一标识符,以降低研究者的工作量,从而推动数据引用最佳实践的产生。

3.2 对出版商的要求

首先,应依据现有的引用规范或指南,制定统一的数据引用标准。出版商应与某一学科领域的数据中心协作制定这些标准,以帮助作者和评阅者更清楚地了解其引用要求,并确保作者和评阅者遵守这一要求[32]。目前在生态学、进化生物学和环境科学领域只有不足 10% 的期刊给出数据引用规范[18]。同样,期刊只有通过同行评议和编辑过程强制实施引用要求,才可以使目前的状况得以改善。

其次，应在数据集元数据标准的采用方面发挥积极作用。高质量的元数据对于数据再使用和促使数据引用的重要性不言而喻，出版商和数据中心都意识到了这一点。为了使用元数据标准，首先需要达成一致，并使作者、数据生产者、数据存储者/图书馆，最重要的是数据出版商认同并接受这一标准。出版商在与这些利益相关者合作的过程中发挥着积极的作用，有助于推动现有规范的统一，并通过同行评议确保元数据的质量。过去《科学》和《自然》对基因数据所采取的强制性政策就是很好的例证。

第三，应积极推动数据唯一标识符的使用。唯一标识符是简短的名字或字符串，是独立于数据位置并永久且唯一的标识数据集。期刊、数据中心所提供的推荐和政策中应该详细描述数据唯一标识符的引用要求。

最后，鼓励出版商进行数据出版。多数人已经意识到通过引用链接到基本数据有助于提高传统文献的价值，甚至呼吁建立新型的数据出版物——数据期刊，这样可以增加一种重要的基本元素，通过知识产权的形式解释数据的特殊价值。

3.3 对数据中心/图书馆的要求

首先，应参与建立统一的数据引用规范，明确指出希望数据被如何引用，考虑各种粒度上的引用应该如何完成，并确保出版指南可将描述详细的引用规范提供给用户。此外，还应与出版商充分沟通，使出版商了解其要求。数据引用应尽可能地简单，这对于鼓励研究者引用数据非常关键，而数据中心/图书馆可在这方面发挥重要作用。在数据集登录页面（提供数据集相关信息的网页）提供推荐的数据集引用格式，或上载到参考文献管理软件的引用格式。登录页面也应该提供与相关出版物的链接方式，以提供更多的有关再使用某一具体数据集的价值的信息。GBIF 自 2008 年就制定了自己的数据引用规范，并在其白皮书中指出，数据引用需要进一步发展以达到与传统文献引用同等的地位，具体工作包括再思考数据的著作权、加强利益相关群体之间的沟通、明确数据出版商出版与引用数据的要求、设计更为灵活的引用格式以适应不同数据类型的差异，确保引用规范的持续性，并强化工具的整合，减少引用工作量[33]。

其次，应在支持并促进唯一标识符使用方面发挥积极作用。和出版商一样，数据中心所提供的引用格式也应该涉及唯一标识符。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合适的引用单元。在生物多样性学科中，有 15 个数据聚合器，有 11 种不同的引用格式，这使研究者感到十分混乱和困惑[33]。因此，同一个学科若能使用同一标准，引用的难度则会大大降低。这就需要数据中心做更多的工作以助力于规范的统一。

4 结语

规范数据引用有助于推动研究者开放共享数据并促进进一步创建、再使用和出版数据。研究者通过使其数据可获得引用荣誉，能够证明其工作的价值，反过来帮助他们获得进一步的资助，使其可以进一步创建、出版和再使用数据。对引用者而言，提供数据链接或引用的论文会获得更多的引用。虽然在研究数据引用方面目前已有很多探索和实践，但是仍然存在诸多问题和挑战，这主要是由技术方面的原因（如基础设施和标准）或文化及社会方面的问题（如作者没有意识到数据引用的要求或者不确定引用哪些数据、如何引用数据以及什么时候和在哪里引用数据等）造成的。为推进数据引用实践，应从以下几方面努力：

首先，需要花费时间让研究者接受这一事物，研究者如果没有看到数据引用的益处，会认为数据引用所需要花费的精力（提供充分的元数据使其他人可以引用自己的数据，确保以合适的方式引用了他们所使用的数据）不可容忍，因为这些工作没有任何回报。其次，需要制定有关数据引用的计量指标，并嵌入当前的认知和奖励系统中。一方面为数据中心和研究资助者进行决策或制定策略提供支持，另一方面强化数据创建者遵从数据引用规范的意识[3]。第三，使用永久标识符独一无二地标识和定位数据集。数据出版商（数据中心、数据库、图书馆、出版商）应分配标示符，使数据集易于被发现和使用。通过编辑出版的过程，一方面使科研人员引用数据时使用永久标识符的需求得以满足；一方面在参考文献中应列出使用永久标识符的引用，以实现数据引用计量指标的计算。第四，出版商和数据中心需要制定指南，以进一步鼓励研究者引用数据，也让同行评议者了解其所实施的引用标准和格式，以确保所引用数据是合适的并被包含在论文中正确的部分。出版商也应该与数据中心和资助者一起协商制定数据引用计量指标，用于监测和评价数据的引用情况，提供数据出版之后在多长时间内被引用和再使用的证据。第五，书目工具提供商应提供渠道，使研究者可以很容易增加其格式是数据集的参考文献，并上传数据中心所提供的书目元数据作为数据集。第六，对如何界定数据集作者及其贡献需进一步明确。

第七,应明确引用数据论文与引用数据集分别在什么情况下实施。第八,数据中心、出版商和研究者之间应加强交流,以确保所制定的任何新的标准和指南适用于现有的实践。

总之,不同利益群体如何协同工作,采用何种方式让研究者看到数据引用的价值和益处,约束其遵守出版商和数据中心制定的引用指南,仍然是一个需要不断探索的话题。

参考文献:

[1] 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(3): 104 -110.

[2] 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,2014,58(8):91 -96.

[3] 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(3):383 -390.

[4] 张丽丽,黎建辉. 数据引用的相关利益者分析[J]. 情报理论与实践,2014(7):44 -48.

[5] 马建玲,曹月珍,王思丽. 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.

[6] 侯经川,方静怡. 数据引证研究:进展与展望[J]. 中国图书馆学报,2013(1):112 -117.

[7] 李丹丹,吴振新. 研究数据引用研究[J]. 图书馆杂志,2013(5):65 -69.

[8] 杜伟,张静. 科学研究数据的出版与获取[J]. 出版科学,2013(6):86 -90.

[9] Australian National Data Service. Cite my data[EB/OL]. [2015 -01 -23]. http://ands. org. au/guides/cite-my-data. html.

[10] Data citation[EB/OL]. [2015 -01 -23]. http://ands. org. au/cite-data/index. html.

[11] Ball A, Duke M. Data citation and linking[EB/OL]. [2015 -01 -23]. http://www. dcc. ac. uk/resources/briefing-papers/.

[12] Parsons M A, Duerr R, Minster J B. Data citation and peer review[J]. Eos, Transactions American Geophysical Union, 2010, 91(34):297 -298.

[13] DataCite International Data Citation Metadata Working Group. DataCite metadata schema for the publication and citation of research data version 3. 0[EB/OL]. [2015 -01 -23]. http://schema. datacite. org/meta/kernel-3/doc/DataCite-MetadataKernel_v3. 0. pdf.

[14] Green T. We need publishing standards for datasets and data tables[J]. Learned Publishing, 2009,22(4):325 -327.

[15] Rohlfing T, Poline J. Why shared data should not be acknowledged on the author by line[J]. NeuroImage, 2012,59(4):4189 -4195.

[16] Chavan V. Recommended practices for citation of data published through the GBIF network Copenhagen[EB/OL]. [2015 -01 -23]. http://links. gbif. org/gbif_best_practice_data_citation_en_v1.

[17] 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,2014,58(8):91 -97.

[18] Wynholds L. Linking to scientific data: Identity problems of unruly and poorly bounded digital objects[J]. The International Journal of Digital Curation,2012, 6(1):214 -225.

[19] Argo project[EB/OL]. [2015 -01 -23]. http://argoproject. org/.

[20] Altman M, King G. A proposed standard for the scholarly citation of quantitative data[EB/OL]. [2015 -03 -12]. http://www. dlib. org/dlib/march07/altman/03altman. html.

[21] 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.

[22] Corti L, Bolton S. Presentation: Data identifiers: How to ensure your data is cited properly[EB/OL]. [2015 -01 -23]. http://www. jisc. ac. uk/media/documents/events/2012/JISC_WEBINAR_DOIs_FINAL. pdf.

[23] Buneman P. How to cite curated databases and how to make them citable[EB/OL]. [2015 -03 -12]. http://homepages. inf. ed. ac. uk/opb/homepagefiles/harmarnew. pdf.

[24] PANGAEA data publisher for earth & environmental science[EB/OL]. [2015 -01 -23]. http://www. pangaea. de/about/.

[25] Ball A, Duke M. How to cite datasets and link to publications[EB/OL]. [2015 -01 -23]. http://www. dcc. ac. uk/resources/how-guides/cite-datasets/.

[26] Robertson T, Doring M, Guralnick R, et al. The GBIF integrated publishing toolkit: Facilitating the efficient publishing of biodiversity data on the internet[J]. PLOS ONE,2014,9(8):1 -7.

[27] Chavan V, Penev L. The data paper: A mechanism to incentivize data publishing in biodiversity science[EB/OL]. [2015 -03 -12]. http://www. biomedcentral. com/1471-2105/12/S15/S2.

[28] NISO Business Working Group 2012. Recommended practices for online supplemental journal[EB/OL]. [2015 -01 -23]. http://www. niso. org/apps/group_public/download. php/10055/RP -15 -2013_Supplemental_Materials. pdf.

[29] Maunsell J. Announcement regarding supplemental material[J]. The Journal of Neuroscience, 2010,30(32):10599 -10600.

[30] Penev L, Mietchen D, Chavan V, et al. Pensoft data publishing policies and guidelines for biodiversity data[EB/OL]. [2015 -01 -23]. http://www. pensoft. net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines. pdf.

[31] Macrina F L. Teaching authorship and publication practices in the biomedical and life sciences[J]. Science and Engineering Ethics, 2011,17(2):341 -354.

[32] Mooney H. Citing data sources in the social sciences: Do authors do it? [J]. Learned Publishing, 2011,24(2):98 -108.

[33] Lane M A. Data citation in the electronic environment[EB/OL]. [2015 -01 -23]. http://www. gbif. org/orc/? doc_id =4884.

(下转第 53 页)

2012 新生入馆教育个案研究[J]. 中国科技博览,2013(29):7.
[16] 田芳. 国内信息素养教育创新与实践[J]. 高校图书馆工作,2012(5):77 -79.
[17] 陈芳. 高校图书馆志愿者长效机制的建立——以三峡大学图书馆为例[J]. 科技情报开发与经济,2014(8):39 -41.
[18] 王珂. 高校图书馆新生入馆教育模式探究——以华中师范大学图书馆为例[J]. 教育教学论坛,2014(14):63 -64.
[19] 王丽娟,王浩,朱大伟,等. 基于90后大学新生心理特征的入学教育探索[J]. 学理论,2009(31):255 -256.
[20] 张羽,杨斌,张春生,等. 中国高校班集体制度对学生成绩影响的实证研究[J]. 清华大学学报(哲学社会科学版),2011(3):133 -142,158.
[21] 马伊凡,徐宁,杨萌,等. 高校新生入馆教育探索与创新[J]. 科技情报开发与经济,2013(3):39 -41.
[22] 袁红军. 高校图书馆读者教育培训能力研究——以中南六省区"211"高校图书馆为例[J]. 图书馆,2013(3):128 -130.
[23] 陈善礼. 经济与政法学院新生自主体验图书馆活动圆满结束[EB/OL]. [2014 -09 -01]. http://www.hbut.edu.cn/contents/59/13698.html.
[24] 陈善礼. 馆院联合举办宣传周 倡议"爱读书 爱图书"[EB/OL]. [2014 -09 -01]. http://www.hbut.edu.cn/contents/59/14086.htm.

On the Model of Library Admission Education for Freshmen Based on Peer Education

Chen Shanli

Library of Hubei University of Technology, Wuhan 430068

Abstract: [**Purpose/significance**] This paper tries to explore the connotation and operational mechanism of the model of university library admission education for freshman based on peer education. [**Method/process**] After reviewing the previous studies and analyzing the feature and need of post 90s, the paper presents a new model of university library admission education for freshman based on peer education, in which the freshmen act as peer educator. [**Result/conclusion**] The paper describes the overall architecture and the trial case, raises more issues which should be discussed in the future.

Keywords: library admission education for freshmen peer education university library post 1990s

(上接第47页)

Discussion on Key Problems of Scientific Data Citation

Wang Dandan

College of Management, Henan University of Science and Technology, Luoyang 471023

Abstract: [**Purpose/significance**] Identify the basic problems, keys and difficulties of data citation, then analyze the responsibility and the future directions of stakeholders, in order to help them to clarify which idea is better, and promote the practice of data publishing. [**Method/process**] Referring to the domestic and foreign related research and experience, starting from the value elaborates the basic problems and difficulties in scientific data citation. Based on this, analyzes the opportunities and requirements for stakeholders. [**Result/conclusion**] Results show that the reasonable standard of scientific data citation should answer three basic questions clearly. They are how to cite, what to cite and when to cite? The metadata, acknowledgement and tools for citation are keys to answer the question of how to cite; versioning, granularity, verification and citation object are cores of what to cite; when to cite is to clarify the scenarios of using data. Stakeholders should see the opportunities and requirements brought by scientific data standard citation, and adjust themselves continuously in order to promote the practice of data citation.

Keywords: scientific data data citation data publishing