

国内 Data Curation 研究综述

王 晴 (武汉大学信息管理学院 湖北 430072)

摘要 Data Curation(DC) 是图书情报学新兴的前沿领域,国内学者对 DC 概念的认知也略有差异,需对其中文译法重新界定。DC 活动表现出多样化的基本特征和价值意义,高校图书馆在 DC 服务中发挥着重要作用,并以多种形态开展 DC 服务,应以 DC 研究为契机,抓住发展机遇,在协同互动的理念指导下实现 DC 服务的本土化。

关键词 Data Curation 科学数据 图书馆服务 机构知识库 文献综述

Review of Data Curation Research in China

Wang Qing (School of Information Management, Wuhan University, Hubei, 430072)

Abstract Data Curation is an emerging frontier domain of library and information science, Chinese scholars have slightly different cognition of the concept of DC, it is necessary to redefine its Chinese translation. The basic features and values of DC are diverse, university library plays an important role in DC activities, and carries out the DC services in a variety of forms, and should seize the chance for development based on the chance of DC research, and the localization of DC services needs the guidance of advanced concepts of coordination and interaction.

Keywords Data Curation, scientific data, library services, institutional repository, literature review

科学研究的深入发展带来数据的不断更新,现有的数据管理模式和方法对于这样越发复杂的大数据,显得捉襟见肘。如何提升数据管理和利用的效率,扩大数据的开放共享,已成为国内学者普遍关注的课题。在对数据的管理、维护和再利用方面,国外研究已然走在理论和实践的前沿,我国学者也意识到开展相关研究的必要性,由此,“Data Curation”概念被引入本土并激发了许多专家、学者的研究旨趣,一些引荐性和探索性的文章见诸于图情类核心期刊,但同时也暴露出一些问题和矛盾,比如对 DC 的译法就存在争议。本文从 DC 的概念认知和译法界定入手,对目前国内学者发表的有关 DC 主题的文献进行全面梳理和总结,为进一步的 DC 研究提供借鉴和参考。

1 Data Curation 的概念及译法界定

1.1 Data Curation 的概念认知

从 CNKI 收录的有关“Data Curation”和“Digital Curation”文献分布情况来看,最早出现在 2011 年(1 篇),2012 年增至 8 篇,截止到 2014 年 3 月底,共检索到 32 篇文章(不计时滞),其中有 26 篇文章被 CSSCI 收录,

说明国内学者对 DC 关注的专业视角相对集中,所发文章的质量也较高,但从论文题名和关键词来看,不同学者对 DC 的翻译也不尽相同,初涉该领域的研究者若以中文题名检索,则很容易漏检其他文章,说明 DC 对于国内图书情报领域来讲是一个前沿性课题,亟待统一译法和用词,以免对后续研究造成混乱^[1]。

任平^[2]将 DCC(National Digital Curation Centre) 译为英国数字医疗中心,而 DCC 组建于 2004 年 3 月,根据其文章参考文献(网络资源)的访问日期可以看出,这篇文章的内容来源最早是 2004 年 6 月的外文网站,说明任平对于 Digital Curation 概念的认识要早于国内其他学者,并指出了 DCC 的核心功能和研究领域。他认为 DC 除了具有数据归档(Data Archiving) 和数字保存(Digital Preservation) 的全部含义外,还具有在学术生命周期中对科学数据进行主动管理和评价鉴定的功能,而这是确保数据共享利用的关键含义。2008 年 2 月,外交部将“首届非洲‘数据医疗’大会在南非召开”为题的新闻发布在外交部官方网站上,文中指出科学数据的长期数字化保存和共享已经受到各国的高度重视,DC 涉及多国间的软硬件系统建设和知识产权

政策等问题¹。张智雄¹在“2009(第四届)教育信息存储大会(北京站)”主题演讲中指出科学数据是 e-science 的灵魂,数据已经成为科研的重要组成部分,但由于数据格式和数据洪流两大危机的存在导致很多数据不能被很好地管理、保存和重新利用,为了激发数字资源的可持续利用功能,需要采取一系列数字化措施使其增值,DC 就是其中一项重要的管理活动,他将 Curation 译为“掌管”,并认为这是数据得以长期利用的重要数字保证。

杨鹤林¹将 DC 译为“数据监护”,他对目前国内少量博客将其翻译为“数据典藏”和“数据诊断”的译法持有值得商榷的态度,并指出 DC 是图书馆服务的一个新兴领域。他对这一译法并没有给出合理的解释和说明。沈婷婷和卢志国⁹采用“数据监管”来解释 DC 的内涵,文中指出“监管”=“监视”+“管理”,认为这样的描述符合活动本身的可持续性和可管理性特点。崔宇红¹将 DC 翻译为“数据管理”,显然,这一译法与“Data Management”存在矛盾,她认为互联网的发展使得“Curation”这一原本属于艺术文化领域的特定称谓泛化了,“Digital Curation”这一概念最早是由英国数字保存联盟(RIN)和国家空间中心(UKSA)在 2001 年的一次研讨会上提出的,而“Data Curation”一词最早则由微软公司的研究员 Jim Gray 在 2002 年提出,在 10 余年的发展历程中得到了学界和业界的热烈回应。同时需要指出的是,Data Curation 和 Digital Curation 概念的关系,Digital 是数字的、数据的意思,更强调数据的数字化形态,而且从相关外文文献可知,前者大多为美国学者所使用,后者则为英国机构所倡导,而从实质上讲,它们所指向的研究对象是一致的,即科学数据,因此,本文将两者视为相近词汇,不做严格区分,以便于将所涉及的研究内容有序地联系起来。

1.2 Data Curation 的译法界定

英国联合信息系统委员会(Joint Information Systems Committee, JISC)在 2004 年即对 DC 中“数据”一词的含义做了说明,DC 中的“Data”是“原始的研究数据”而非计算机编码的数字化数据,并就“Data Curation”、数据归档(Data Archiving)和数字保存(Digital Preservation)三个容易混淆的相似概念做了区分,归档是内容层面的活动,保存是技术层面的活动,而 DC 则是对动态数据集的持续性更新、补充和完善的管理活动。我们可以从两个方面来对 DC 进行译法界定:其一,DC 的对象是科学数据,这是国内学界已达成的共识,不存在任何争议,因此,DC 的内涵及外延是基于科学数据展开的,应以科学数据在科研活动中的地位、作用和影响为理解前提,那么,科学数据的价值在于其开放共享性和有效增值性,这也是 DC 活动的出发点和落脚点,说明 DC 的概念中应充分体现出数据的可增值性;其二,“Curation”的词源是中世纪拉丁文 cūr(a),

后演变为英文词汇 Care⁸,显而易见,Care 的中文意思是“照顾”,若直译为“数据照顾”,则用词显得过于生活化,为了避免这一缺陷,可将 Care 的中文词义进行延伸,同样作为名词的“护理”无疑是最佳选择,这样也将中文译法学术化,完全符合 DC 的本意。总的来说,DC 的内涵不仅在于其管理功能,还涵盖其再利用潜力,对 DC 的翻译不能局限于其活动过程的阶段性特征或单一功能,而应从 e-science 环境中去体会其内涵,因此,笔者倾向于将 DC 翻译为“数据护理”以显示其内在价值和外在效用。

2 Data Curation 的特征及价值意义

2.1 Data Curation 的基本特征

DC 服务的核心要素是科学数据,这些数据可以通过科学实验、观察测量或二次计算获得,包含了文本、图像、音频、视频、算法、软件以及模型等在科研过程中可以实现的多样化格式,这些数据本身有着显著特点:(1)数据的来源多样化,数据间的逻辑关系复杂。高校学科设置类别繁多,学科之间的知识结构和研究内容也各不相同,所产生的数据内容也多种多样,同时,随着交叉学科研究的不断发展和深入,学科之间的联系(尤其是相近学科之间)也趋于复杂化,这对于数据的甄别、筛选、分类和规范化处理带来了不小的难度,也增加了对于具有跨学科知识背景的馆员的需求量。(2)由于学科本身的差异,数据存储方式也有很大差别,比如,影视文学与工程测量,有的已经存入专用数据库或保存为特殊格式,“图书”与“图纸”的差别将影响 DC 服务内容的细化工作。(3)由于高校学科建设进展以及高校内不同学科的重视程度和规模大小迥异,资源配置方面就会有差别,每个学科所能生产的数据量也是不一样的,需要考虑数据保存的优先顺序和方法类别⁹。

DC 服务是基于科学数据这样的“材料”展开的,为了促进科研投入产出的收益最大化、验证和拓展研究结果,减少数据损失以提高科学解释的精确性,应当开展强有力的数据护理活动以化解潜在的风险,确保数据源的完整性以及再利用的可靠性。可见,DC 活动主要包括以下特征:(1)持续性的管理活动,这一管理活动始终伴随着数据的生命周期并能启发学习教育,DC 通过数据的采集、归档、保存和描述来促进数据价值的再发现以保证重复利用;(2)保持数据的增值利用,对数据进行护理活动的最终目的是提高数据的利用效率,挖掘出所有的潜在价值,由于科学数据具有时效性强的特点,及时有效地维护数据的内在价值并使其外化为更多的科研动力是 DC 的任务之一;(3)跨区域的协同活动,科学研究的跨区域协同合作是大数据时代对数据资源的终极关怀,是信息社会下公众获取数据、利用数据的具体表现,DC 服务是这种

价值理念的更新和应用,同时,跨界协作需要一系列的规则、协议或制度来协调主体行为,也需要大数据技术、语义网、关联科学的支持。

2.2 Data Curation 的价值意义

2.2.1 满足科研的现实需求

有效的数据保存机制应该是所有科研项目工作的组成部分,部分数据的特殊性和稀缺性要求科研从业者必须珍视,同时,项目的后续工作也会再利用这些数据,因此,对这样一笔重要的战略资产进行妥善的保管、维护和开发是科学研究的必然要求。科研人员在创作过程中会采集外部数据以支撑自身的研究需要,通过数字化媒介获取的数据内容和其他信息在数量上已经超越了本身的产出,面对这样的复杂“数据集”,如何检索利用成为一项重要的工作,如果缺乏有力的管理措施,这些数字化形式保存的内容将面临着来自数字退化和数据时效性等方面的威胁。数据的护理工作改变了“数据密集型”学术时代的某种悲观局面,促使科研数据“重获新生”^[19]。科研项目组将数据交由可信的 DC 小组进行护理,在一定程度上缓解了科研人员的工作压力,分担了科研“边缘”工作,同时也为数据的交流、共享提供了契机,因此,DC 活动有益于建立起轻松、开放、协同的工作氛围和环境,极大地满足了科研机构 and 人员对于数据管理和维护的现实需求。

2.2.2 扩大数据的开放共享

科学数据是科研活动的阶段性产出,是一种具有潜在经济价值的战略资源,也是政府部门、科研机构或其他组织制定计划、咨询决策的参考资料,在社会经济活动中发挥着重要的作用。开放共享是促进科学数据流动和转化的有效途径,对缩小信息鸿沟,促进信息公平也有积极的意义。将那些处于封存或半封存状态的科学数据开放,不仅有益于科研内部的知识共享和转化,也能提高科技创新能力,促进经济社会的可持续发展。

面对以 TB 级规模不断增长的海量科学数据资源,如何帮助用户便捷地获取和共享这些分布在数据库中或是 Web 上的大数据,越来越多的应用程序对科学数据本身的关注度在提高,一些具有前瞻性的研究项目也将数据的开放共享机制纳入研讨的范围,在实践中,图书馆也在积极准备,这已成为图书馆深化服务体制改革的重要风向标。学科服务已经成为高校图书馆工作的主要内容,大量增长的亟待开放和共享的科学数据为学科服务增添了新的内容,高校图书馆利用科学数据开展学科服务是工作创新和发展的的重要途径,并且已经在高校中兴起和实践。在 e-science 科研背景下,由于数据量的庞大和内容的复杂,独立从事科研活动的实验室、研究院所可能会面临着数据保存不善的风险,科研人员对数据共享也有较为浓厚的

兴趣和期盼,图书馆正好可以借此时机开展数据的管理、维护和开发利用工作,将 DC 工作与学科服务相融合,提升自身的价值功能^[1]。

2.2.3 促进岗位的优化转型

随着 e-science 的快速发展,科学数据的再利用日益受到重视,国外相关机构已经设置专门的数据研究专家、数据服务馆员等岗位,将图书馆的服务领域回溯至科研学术交流的上游阶段。以往高校图书馆建设和维护的机构知识库主要是面向本校师生而展开,管理的内容也多以公开发表的论文、著作和报告为主,对研究结果中产生的原始数据缺乏有效管理,未来的高校图书馆将向科研支撑材料的管理和开发工作转型,这种趋势是显见的。叶兰^[1]在对国外的 60 个 DC 岗位设置与需求分析中发现,有的 DC 岗位已经兼具机构知识库的功能,并且积极地与其他相关部门展开密切合作。由于 DC 活动与科研的融合,也使得 DC 逐渐成为学术交流体系中的重要环节,未来图书馆的科研服务功能将由 DC 岗位担当,同时,应当看到的另一面是,馆员在参与 DC 服务的过程中,也扩大了自身的知识面,提升了交流能力,形成了新的不断更新的知识结构,这对于馆员的自身发展是有益的,因此,这样的岗位设置也会吸引更多的追求进步的馆员参与其中。总的来讲,新形势下的 DC 岗位需求给图书馆及其馆员、教学院系及其职员注入了新的改革元素。

3 Data Curation 的主体及服务形态

3.1 Data Curation 的主体角色

图书馆及图书馆员在 DC 领域的角色转变已经成为公众关注的焦点,尤其表现在科学数据的保存和管理提供支撑服务方面,图书馆在 DC 领域发挥如此重要作用的优势来自三点:(1) 图书馆具有文化传承、继续教育和传递信息等功能,e-science 环境下科研数据凸显的信息资源价值使得图书馆(尤其是研究型图书馆)在科学数据的增值服务方面具有不可推卸的义务和责任;(2) 图书馆具有较为完备的服务职能体系、不断更新的技术平台、充沛的办公空间和资金配置,以及拥有交叉学科背景和知识结构的资深馆员,都赋予了图书馆参与 DC 活动的强大动力;(3) 图书馆在信息资源建设、分类编目、语义描述及规范、数据挖掘及加工等方面具有专业的学科知识和丰富的业务经验作为支撑^[1]。

e-science 环境下的研究图书馆需要重新定位自己的功能角色,以便于对数字内容进行维护和管理,而现在的科研和学习活动更加趋向于跨学科性、虚拟性和合作性,跨机构的协作能够为科学数据的有效利用和增值提供活力,这就要求图书馆必须建立起新的服务角色:(1) 致力于创造可持续性的科研环境,而不是建立孤立的流程和系统,或者以单一视角去考察现

状;(2)提供可靠的DC服务来提高它们存在的实用性,以支持更加丰富的知识创造活动;(3)制作OA期刊或者电子出版物来扩大资源的利用范围。同时,图书馆员的角色也更加多样化,主要表现在四个方面:(1)承担了权益管理专家的部分职能;(2)从信息管理的旁观者转变为师生信息消费的“研究者”;(3)成为系统建设者和组织设计师以及数字内容的传播者;(4)在科研机构内外部之间搭建起合作的平台并以多种方式维持这种健康的协作关系^[14]。构建机构知识库是众多高校和科研机构追逐的服务项目,但其中也不乏“面子工程”或至多是个“网上文库”,科学数据的生产者对IR不甚感兴趣的重要原因在于其服务的末端化,即用户不能够参与到这样一个流程中来,而是作为数据的提交者角色参与少量的业务,在一定程度上削弱了科研者共享数据的积极性,而一个新的服务理念是让IR“回到科研流程中”,开展DC活动正是面向科研流程的新服务,保存数据只是服务的切入点而不是终点,充分挖掘其科研驱动力和再利用价值才是最终目的,这要求图书馆要从静态的馆藏服务向动态的“科研代理人”转变,发挥科学数据交流、共享的桥梁和纽带作用^[15]。

3.2 Data Curation 的服务形态

3.2.1 Data Curation 的层次模型

DC是一个面向知识资源增值的循环系统,DCC提出的DC流程主要由三个层次构成:(1)低层次是“研究过程”+“出版过程”的传统数据增值过程,通过出版发行的形式将研究者的数据资源实现在同行、图书馆、科研机构以及公众之间的信息共享;(2)中层次是基于数据研究和归档等环节的数字资源长期保存过程,以应对激增的科学数据并保持数据的真实性和安全性以此提高利用效率;(3)高层次是基于语义工具的知识发现过程,并由数据的护理者、机构知识库以及相应的协作机制共同组成,实现真正意义上的知识增值^[1]。每个项目所设定和分配的结构模块不尽相同,相对应的辅助性工具也存在差别,例如DCP项目组联合开发的基于DCC生命周期模型的DC工具包,细化层次模型中每一个程序、流程和步骤,旨在为图书馆员和科研人员之间建立起沟通的桥梁,这样一个工具包主要由用户指南、采访手册、采访工作表和DC文件模板组成,通过将不同学科的数据服务需求整合进DCP文件,根据不同专业科研人员的需求,制定因人而异或因学科而异的一般性解决方案^[16]。

3.2.2 Data Curation 的业务模块

不同的DC实施机构根据自身的服务对象和业务需求而采取不同的业务模块,但基本原理是相通的,也是具有可借鉴意义的。英国数据档案馆(UKDA)具有40余年的DC活动历史,它提出的DC业务模块由数据处理流程(9个步骤)、数据质量控制(4个层级)、

数字资源保存策略(2年修订一次)和数字资源仓储平台(三方协议)四个部分构成,UKDA的DC业务特色是对于数据质量的控制和数据标准的规范,并将信任机制作为保证DC有效运行的整体性框架^[17]。它对我国DC工作的借鉴意义在于:(1)对提交到IR或类似收藏机构的科学数据进行严格的筛选,并根据实际情况对数据分配不同的处理标准;(2)制定层级分明的数据质量检查和验证标准,并贯穿于整个DC过程以确保元数据的完整性和可用性;(3)建立面向用户需求的动态管理机制以优化工作流程,并及时更新软件和硬件以应对不断变化的用户环境;(4)对存储平台实施认证和授权制度,提高风险管理意识,并将这种综合性协议框架在成员内部达成一致性,与外部联盟实现映射,以提高DC政策的执行效率。

3.2.3 Data Curation 的领域界限

对科学数据的护理是一个涉及面广泛的协作性过程,从数据生命周期角度看,DC活动共涉及三个领域和两个迁移过程^[18]。私有研究领域是第一领域(D1),科研人员通过管理信息系统采集、保存和管理科研活动产生的数据资源,这些数据特征多表现为孤立、分散的并缺乏有效组织的特征,仅供个人访问和再利用,很少与外界存在数据交流。D1迁移机理是通过团队主管的授权,由IT人员设置严格的访问控制程序,实现一个简单的Curation领域。共享研究领域是第二领域(D2),科研项目或小组成员可以通过协作信息系统进入数据存储库共享数据,相对于第一领域而言,第二领域的更新速度、开放范围和利用效率都有所提高。D2迁移机理是得到团队主管的授权并由DC管理者具体操作实施有效共享服务。这些数据成果在发表后即进入第三领域(D3),通过关联数据或开放存取的形式实现开放共享,相比较前两个领域,D3保留较为规范的元数据标准,并由专门的DC管理机构实施最强的维护力度,同时应注意,在将数据汇入公共领域后,还需要解决数据格式兼容、系统访问识别码以及互操作技术可转换性等问题。

4 Data Curation 的机遇及发展策略

4.1 完善政策引导机制,拓展学术研究领域

DC是一个涉及多方面参与力量的复杂性系统工程,既要有上级部门来组织、调配相关资源参与其中,还要有国家层面的政策性保障制度来支撑DC的发展^[19]。一方面要加大对学术造假行为的审查力度,杜绝数据源头上的偏差;另一方面,要规范数据共享和利用制度,倡导纯净的学术精神,对于不涉及国家秘密和安全的科研数据要及时有效地公开,不适宜公开的应书面说明理由并作为研究成果的附件一并提交给相关机构。高校图书馆应认识到DC服务对现有资源提出的更高要求,为此需要在制度、管理和业务等层面建

立起新的培训机制、评价体系和培养方案,主动寻找存在的差距,向国内同行、国外机构借鉴学习^[20]。在有利的 e-science 环境下,相关科研人员(包括直接参与 DC 活动的馆员)需要积极转变观念,投入到 DC 的理论和实践探索当中去^[21],至少有以下几点可以值得考虑:(1) 密切关注国际性的 DC 研讨会,如有可能,申请到参与机会,把国外专家请进来,现场授课,临场指导,效果会比直接获取网络信息资源要好一些;(2) 开展 DC 实验项目的应用研究,在实验中发现、分析问题并找到最佳的解决方案是践行 DC 活动的有效途径之一;(3) 提高面向 DC 服务的信息技术操作可行性,开发针对性的 DC 工具包是 DC 环节必不可少的流程,而且工具包质量直接影响应用效果,这一点不容忽视;(4) DC 服务涉及科学数据的知识产权和版权许可问题,在 DC 过程中要制定切实可行、行之有效的 DC 管理制度和政策法规,这既是控制数据使用的合法性措施,也是提高数据共享水平的政策性保障。

4.2 创新跨界合作模式,开展横向协同服务

科学数据工作已经从独立分散的基础性地位发展成为驱动科学研究的动力机制,21 世纪以来,欧美国家的基金会、学会、协会以及其他学术组织相继设立和资助了一系列有关开放数据的项目或计划,以促进科学数据的增值利用和开放共享。同时,应该认识到的是科学数据的增长趋势是爆发性的,数据结构的多样性和复杂性都给数据的开发利用带来了挑战。图书馆、科研机构以及资助组织都意识到协作效应的重要性,通过定期举办研讨会、出版学术专刊、成立工作组来共同解决 DC 应用中的现实问题。虽然 DC 在国内的发展仍处于初级阶段,但是这并不能抑制创新活动^[22]。首先,作为 DC 的实施主体,图书馆可以在人员分工、资金分配、资源分享等方面寻求业界的最佳组合,成立专门的协同工作组,定期开展业务研究和学术交流活动,扩大数据共享的范围;其次,设置数据馆员岗位,将具有计算机、信息管理以及图书馆学等学科背景的馆员纳入培训课程中来,补充新的业务知识和技能,进一步激发馆员对于 DC 活动的热情,以便于向高校师生宣传 DC 的服务理念和应用前景;还有,可以组织技术人员合作开发 DC 工具包或技术平台,为 DC 活动的有效开展提供技术支撑,另外,应促成图书馆员与科研人员的业务融合,鼓励馆员一开始就进入到科学研究的数据创建工作当中去,而不只是出现在数据生命周期的末端,这样有利于建立一个互惠互利的协同工作网络,提高数据标准、架构的可操作性。

4.3 设置数据馆员职位,开展 DC 培训项目

图书馆的岗位结构对其业务发展和服务延伸具有一定的调控和导向作用,为了应对不断增长的科学数据开发利用需求,国外的许多高校图书馆纷纷设置“数据馆员”(Data Librarian),同时招聘专业人才从事

DC 服务工作^[23]。数据馆员的内涵及外延与数据护理员(Data Curator)基本一致,并且在业务功能上相互配合。一个称职的数据馆员至少需要满足以下三项条件:一是拥有数据及相关专业硕士学位,或是经过 DC 课程培训的 LIS 硕士;二是具备至少两种以上计算机语言应用与信息分析能力,尤其是机构知识库的资源建设经验;三是具有高度责任感和服务意识以及团队协作能力。为了激发数据馆员充分发挥自身优势,调动工作积极性,作为管理机构的图书馆应及时补充制定一系列保障性规章制度,优化组织设计和资源配置。同时,作为馆员队伍的后备力量,高校图书情报课程设置也应努力向现实需求对接,招收面向 DC 服务的定向培训生,为了保证课程教育质量,需要注意以下几个方面:(1) 凸显 DC 专业教育的特色,注重与传统办学优势的衔接;(2) 开展 DC 应用工具实践项目,训练学生的数据采集、挖掘、分析及预测等动手能力;(3) 开展多样化灵活性的授课形式,让学生与数据馆员零距离交流,将实际需求纳入课程目录;(4) 动态追踪机制,对已经参加 DC 服务的毕业生实行继续教育,以便于他们掌握更多的实用技能和分析方法^[24]。另外,值得推荐的是,理论研究也是推动 DC 教育实践发展的动力源,理论与实践教育是双向互动的过程,两者的有机结合才能发挥 DC 服务的更大效能。

5 结语

Data Curation 研究在国外的发展已有 13 年,产生了一大批卓有成效的实践项目,也带动了相关的理论研究。国外的实践项目通过建立可信的数字保存机制来实现可靠的经济性支撑模式,并从微观的系统应用转向宏观的战略规划^[25]。DC 概念进入国内也只有 3 年多的时间,时间虽短,但兴趣不减,图书情报领域的学者陆续开展了系列研究,但相关的规模和质量显得略有单薄。从作者发文情况来看,只有 1/3 的作者发表了 2 篇及以上的文章且缺乏 DC 基金项目的支持,这也给我们带来了一个启示:国内学者还没有形成一个较为集中的研究氛围或团队,研究成果也多表现为引介国外 DC 项目进展、教育培训及案例分析等,将国外的 DC 服务理念内化到国内的实践项目中仍需时日。从目前国内研究成果分析,未来的研究重点将包括:(1) 面向跨界数据增值目标的 DC 服务协作机制;(2) 基于数据标准的 DC 业务流程规范化操作;(3) 围绕人才培养的数据护理专业教育实施方案;(4) 基于数据权利的法规政策研究;(5) 面向可持续发展的 DC 技术应用体系。我们应当看到,DC 研究不再是单一学科所能驾驭的,正如数据的开放共享理念一样,研究也需要协同互动。

参考文献

- [1] 张文彦,武瑞原,于洁.大数据时代的图书馆初探[J].图

- 书与情报,2012(6):15-21.
- [2] 任平.数字资源长期保存研究在英国[J].现代图书情报技术,2003(3):66-69.
- [3] 首届非洲“数据医疗”大会在南非召开[EB/OL]. [2013-03-27]. http://www.fmprec.gov.cn/mfa_chn/wjdt_611265/zwbdt_611281/t434384.shtml.
- [4] 张智雄.数字保存是数字时代信息中心的一项重要战略[EB/OL]. [2013-03-27]. http://www.edu.cn/li_lun_yj_1652/20091124/t20091124_424016.shtml.
- [5] 杨鹤林.数据监护:美国高校图书馆的新探索[J].大学图书馆学报,2011(2):18-21,41.
- [6] 沈婷婷,卢志国.数据监管在我国高校图书馆的应用展望[J].图书情报工作,2012,56(7):54-57,87.
- [7] 崔宇红.E-Science环境中研究图书馆的新角色:科学数据管理[J].图书馆杂志,2012,31(10):20-23.
- [8] Curate[EB/OL]. [2013-03-25]. <http://dictionary.reference.com/browse/curate>.
- [9] 沈婷婷,卢志国.科研项目不同阶段的科学数据监管方法[J].图书馆建设,2013(3):49-51.
- [10] 刘雄洲,王菲.国外数据存管实施现状及其对国内高校图书馆的启示[J].图书馆,2012(5):81-83.
- [11] 程莲娟.美国高校图书馆数据监护的实践及其启示[J].图书馆杂志,2012,31(1):76-78.
- [12] 叶兰.国外图书馆数据监护岗位的设置与需求分析[J].大学图书馆学报,2013(5):5-12.
- [13] 时婉璐,任树怀.数据策管:图书馆服务的新创举[J].图书馆杂志,2012,31(10):24-27,34.
- [14] 谭榕,元靖涛.新时代新角色:数字保存的监管[J].图书与情报,2011(3):45-48.
- [15] 杨鹤林.数据监护看美国高校图书馆的机构库建设新思路——来自DataStaR的启示[J].大学图书馆学报,2012(2):23-28,73.
- [16] 丁培.数据策展与图书馆[J].图书馆学研究,2013(6):94-98.
- [17] 殷沈琴,张计龙,窦方.欧洲科学数据监护的标准与实践——UKDA案例研究[J].图书馆杂志,2013,32(6):76-80.
- [18] 吴敏琦.Digital Curation:图书情报学的一个新兴研究领域[J].图书馆杂志,2012,31(3):8-12.
- [19] 张秋彦.高校科学数据监护研究[J].情报科学,2013,31(5):42-45.
- [20] 夏姚璜.国外数据监护教育和培训实践及启示[J].新世纪图书馆,2013(6):35-37.
- [21] 高红文,陈清文.国外数据监管研究综述及启示[J].图书馆学研究,2013(10):2-4,27.
- [22] 夏姚璜.欧美Data curation的实践及启示[J].信息资源管理学报,2013(1):53-58.
- [23] 孟祥保,钱鹏.国外高校图书馆数据馆员岗位设置与管理机制[J].图书与情报,2013(4):12-17.
- [24] 孟祥保,钱鹏.国外数据管理专业教育实践与研究现状[J].中国图书馆学报,2013(6):63-74.
- [25] 李文文,陈雅.国内外Data Curation研究综述[J].情报资料工作,2013(5):35-38.

[作者简介] 王晴,男,1986年生,武汉大学信息管理学院2014级博士研究生。

收稿日期:2014-03-31

欢迎订阅 2015 年

图书、情报、信息、资料工作者 自己的刊物

《情报资料工作》

CSSCI 来源期刊,全国中文核心期刊,中国人文社会科学核心期刊
双月刊,大16开,112页,全年定价288元,国内统一刊号CN 11-1448/G3

——图书馆、情报室、资料室、信息中心的理想文献、必藏刊物
——图书、情报、信息、资料、档案工作者的业务参谋、良师益友

《情报资料工作》杂志面向图书情报界、信息产业界和文献资料档案界,读者遍及高校、党校、科学院、军队院校、政府信息机构及公共图书馆系统。刊物一贯注重追求理论精品,面向工作实际,形成了求实创新的学术风格,是广大图书情报工作者进行学术研讨及业务交流的重要园地。

订阅方式:

- (1) 在全国各地邮局订阅,邮发代号:82-22
(2) 直接向中国人民大学书报资料中心市场部订阅
① 邮政汇款:
地址:北京9666信箱(邮编100086)
收款人:市场部
电话:010-82503412/38/40 62512171

② 银行汇款:

户名:中国人民大学书报资料中心
开户行:中国银行北京人大支行
账号:344156031742

欢迎致电《情报资料工作》编辑部,电话:010-62512296/62514977;或登录博客:<http://blog.sina.com.cn/qbzlgz>;登录网站:<http://qbzl.zlzx.org/CN/volumn/home.shtml>;或发邮件至 qingbaoziliao@263.net。