

# 国外 Data Curation 研究综述

王 晴 (山西大学经济与管理学院 太原 030006)

**摘要** e- Science 环境下科学数据的有效管理和开发利用至关重要, Data Curation( DC) 是一项持续性的动态管理活动,具有重要的短期效益和长期价值,DC引起了国外学界的普遍关注。大学社区和科研机构就此展开了一系列卓有成效的实践探索,图书情报学院相继开设了具有针对性的教育培训课程以塑造未来的 DC 专业人才,DC 实施主体针对发展现状提出了改进方案或战略规划以支持 DC 的可持续发展。

**关键词** 科学数据 data curation e- Science 机构知识库 文献综述

## An Overview of Data Curation Research Abroad

Wang Qing ( School of Economics & Management, Shanxi University, Taiyuan, 030006)

**Abstract** Effective management, development and utilization of scientific data is a vital work in the e- Science environment, data curation is a continuous dynamic management activity, DC have important short- term benefits and long- term value, caused foreign scholars' s widespread concern. Foreign university communities and research institutions have carried out series of effective practice and exploration, library and information science schools have set up targeted DC education and training programs in order to shape the future DC professionals, the operation main constituent of DC proposed improvement programs or strategic plans aiming at the present development situation to support the sustainable development of DC.

**Keywords** scientific data, data curation, e- Science, institutional repository, literature review

“e- Science”一词通常指借助于网络技术,旨在推进数据、资源和通信的统筹和分布式共享而进行的大规模科研合作,观测数据和实验数据催生了大量的典型和通用数据类型,数据收集和开发的 Curation 工作对结果验证、后续试验和积累分析起到至关重要的作用。尽管如此重要,但是只有极少量的科研产出得到了有效管理和维护( curated),为了促进科研投入收益的最大化、研究得以验证和拓展,减少数据损失并提高科学解释的精确性,应当施行强有力的 Curation 实践以化解潜在的风险,确保数据的溯源性( provenance)、完整性( integrity)以及再利用的可靠性( reliable)。科研数据量的指数级增长以及不断涌现的新技术,给数据管理和开发活动带来挑战和危机,同时,科研数据的有效管理和充分开发能够促进知识创新,两者是不可分割的互动关系,国外学者在 21 世纪初就对数据的管理及相关问题作出了探索性研究并付诸实践,这对改善科研数据的应用环境起到了理论支撑和实践指导作用,积极借鉴和吸收国外的先进经

验对促进我国的科研创新有重要意义。

## 1 Data Curation 科学内涵及价值

### 1.1 Data Curation 的基础定义

英国 Digital Curation Centre( DCC)<sup>[1]</sup>对 Digital Curation( DC)下了明确的定义:DC 是指在数字化研究数据的生命周期内产生的维护、保护和增值数字化数据的动态管理活动,目的是减少长期研究价值的威胁,降低数字老化的风险,置于可信的数字化存储库中的“Curated Data”,可以增进英国研究界的数据共享并减少研究数据创建工作中的重复劳动,增强高质量科研的可用性以提高现有数据的长期价值,这里的“Digital”实质上是“Digital Research Data”的缩写,DCC 认为 DC 和“Data Preservation”是正在进行中的动态过程,需要深思熟虑以及充足时间和资源的投入,参与者必须做到心中有数并采取行动以促进整个数据生命周期 Curation 和 Preservation 的有效性。美国伊利诺伊大学( UI)图书馆与信息科学( LIS)研究生院认为,

Digital Curation 是主动的、持续性的数据管理,这一管理活动伴随着数据的生命周期并对学术、科学和教育大有益处,DC 通过认证、归档、管理、保存和表示来促进数据发现和检索、保持数据质量、增加价值并提供重复利用<sup>[3]</sup>。Adrian Cunningham<sup>[3]</sup>从档案工作者的角度分析了 Digital Curation、Digital Archives 和 Digital Libraries 的联系和区别,他认为 Digital Curation 不仅是档案材料的数字化采集管理,数字化记录的 Curation 实质上是截然不同的 Curatorial 活动,是为了保证数字归档(digital archiving)这一不同术语的使用而进行的名词划分,从记录保存的视角看, Digital Archives 一词已经被误用和曲解,因此掩盖了与档案资源的采集和长期管理等根本性问题,数字归档是一个贯穿于记录保存过程的介入活动,正如开放档案信息系统(OAIS)参考模型的缺陷性一样,忽略了档案活动的预先摄取需求, Digital Curation 的价值在于跨越整个数字信息生命周期的各个流程并通过相关专业的努力从而团结成一个连贯的整体。

### 1.2 Data Curation 的内涵界定

Lord P. 和 A. Macdonald<sup>[4]</sup>对“Curation”、“Archiving”和“Preservation”三个相似的活动做了内涵界定。他们认为,“Curation”是管理和推广数据使用的一系列活动,以确保它符合当前的宗旨、发现和再利用,对于动态数据集,这意味着不断丰富和更新以保持灵活的适用性,高水平的“Curation”活动也将涉及维护注释和其他已发表材料的链接。“Archiving”是一项“Curation”活动,目的是确保数据的最佳选择、存储和获取,它的逻辑和物理完整性是指随着时间的推移,能够不断地保持科学数据的安全性和真实性。“Preservation”是包含在“Archiving”内的活动,保持特定的数据项,随着时间的推移能够通过技术变革来保证数据可以被访问和理解,因此,从广义上讲,“Preservation”是“Archiving”的一个方面,而“Archiving”是“Curation”根据需要而选择的一项活动,三者的管理关系随着时间的变化而不断调整和深化。Helena Karasti 和 Karen S. Baker 等<sup>[5]</sup>认为,e-science 环境下的“Data Curation”具有丰富的科学内涵,公共资助数据的开放存取是一项令人钦佩的活动,也是实践中难以彻底贯彻执行的政策,同时也为科研的实际行为、高质量数据的 Curation 活动以及长期管理工作带来了前所未有的挑战,“从事科研”实践中的数据挑战是普遍存在的,包括自动数据采集的 e-Science 将改变常规的基础性科学决策类型和所拥有的生态基础,从一开始就要考虑周全面对 Data Curation 不断变化的主体意识必不可少的应对措施,Curation 和管理工作(Stewardship)虽然都是针对数据展开的,但是对数据本质(Nature of Data)、生命周期以及科研环境的看法却不尽相同,e-Science 文献的 Data Curation 贯穿于数据摄入(Ingestion)、归档(Archive)和交

付(Delivery)等指引和程序的组织和监督过程中,而 Data Stewardship 只是提供了一个宏观的概念框架,包括回溯过去、了解现状以及未来规划等,涵盖了从数据规划到数据采样,从数据归档、利用到再利用等一系列过程,包括数据定义、需求分析、质量保证、反馈、再设计以及数据交换等数据关怀(Data Care)和信息基础工作。

### 1.3 Data Curation 的价值意义

Curation 激发了自信的和可靠的数据再利用,它对数据的创造者、研究人员、资助者和数据用户产生了即时和短期收益以及长期价值,大英图书馆的 Maureen Pennock<sup>[6]</sup>认为,DC 具有提高科研数据质量、提供可靠工作数据的获取、允许结论的外部验证、发布良好的记录保持标准、允许大数据的分析和开发、管理不同版本的动态数据集、确保有价值的知识和数据以及允许数据集的创新结合 8 项短期收益,有利于科研过程性证据的记录、收集/分析数据和初期的财务投资的潜力最大化、识别和确保数据的可用性和非重复性、为解决技术过时问题提供框架、增强数据出处的可验证性以及允许未来用户能够有效再利用或借鉴 6 大长期价值。Curation 不仅仅是一个简单的技术关注,还应该解决组织和文化问题:①增强不同网站之间的持续沟通性以避免不同系统和体制的目标所造成的冲突;②辨别和确定每个站点上员工的责任义务以促进 Curation 关注;③制定政策和程序以处理和定义好管理、传输、存储、归档、保存、访问和重用等实践活动;④加大培训和教育力度以确保符合定义的良好实践。当然,科学家和研究团队仍面临着实践和技术方面的问题,包括开源软件和开源标准、适宜的注释和元数据、关联数据层次、标示符和引文格式、监测和更新旧的存储设备、存储库的鉴别和扩展以及数据迁移和再利用等,e-Science 研究的分布式共享本质依赖于协同开发的可重复技术解决方案。

## 2 Data Curation 的技术模型及应用

### 2.1 技术模型

科研数据 Curation 问题吸引了大量的关注,随着计算能力的提升、新技术的应用和数据的指数级增长,如何保存和开发数据已经成为知识转化的重要影响因子,能否获得科学数据的最大利益在一定程度上取决于学术图书馆的努力程度。康奈尔大学阿尔伯特·曼恩(Albert R. Mann)图书馆参与了 USAEP 项目的数据归档和分配工作,Gail S. Steinhart 和 Brian J. Lowe<sup>[7]</sup>认为,“Data Curation”的一个核心问题是某个特定学科的数据存储库(data repositories)是否得到最佳管理,但是促进和支持数字化数据(digital data)的基础设施仍受制于不均衡发展、学科、机构和研究范围的可用性等因素,研究图书馆的角色是为研究生命周期(RLC)中 DC 活动提供服务和设施,最终目的是促进转移完成

的数据集和元数据至领域特定机构知识库 (IR) 中。为此,他们通过辨别和使用特定领域的元数据标准来建立数据集等方法,建立了一个基于探索机制和 MIT s DSpace 的 DC 战略模型,该概念模型的设计理念是为了促进个人或研究团队科学数据实现更大范围的共享,该模型由工程参与者(个人实验室、研究小组和研究人員等)、临时区域(图书馆服务)、机构知识库(数据和元数据保存)以及搜索引擎和门户网站等构成。

科学数据一经发表就应该保持永久的可用性,以提供知识创新的源泉,使用者需要元数据以便能够准确理解数据的细节,包括工具的设计和构建、何时何地如何采集数据以及处理步骤的详细信息等, Jim Gray 和 Alexander S. Szalay 等<sup>[8]</sup>描述了一种记录、发布和维护斯隆数字巡天(Sloan Digital Sky Survey)数据的方法,认为不能够复制或重构的数据需要被保存下来,比如元数据。在数据发表(或出版)过程中涉及作者、出版者、创作者和消费者四大利益相关者,出版物由图书馆(Curators)保存和索引,图书馆员记录元数据被视为一种数据 Curation 行为。数据本身是无用的,除非有一份关于数据采集的时间、方式和派生的详细清楚描述,他们将 Curation 形象地描述为“捕捉蜉蝣”(Capturing the ephemera),强调数据利用的时效性和保存的及时性, Curators 应该尽可能地获得更多的元数据,如项目设计文档、讨论、程序、软件和操作日志等。

## 2.2 应用实例

Data Curation 对科学数据的数字化、共享、整合和利用等活动起到关键作用,例如在生物多样性研究中,DC 包括填写丢失的数据、修正错误和保持计量等工作,标本的数据需要被数字化处理以适应管理需要。DC 的工作流程通常包括使用多种工具或服务来支持工作,DC 的每一个步骤都被视为一个使用编程语言创作的单独程序以获取相应的服务和工具,开普勒(Kepler)科学工作流程技术提供了一种替代的解决方案,通过 workflows 的建设、调度和管理来实现 Curators 对文档数据的自动化操作。L. Dou 和 G. Cao 等<sup>[9]</sup>通过自制的 Kurator 工具展现这种方法的有效性,一个标本数据的数据 Curation 流程如图 1 所示。

图 1 展示的是 Human Curators 如何将自动创建步骤交织在一起并实现有效运转的工作流程,根据各阶段的不同特点可将全部流程分为两个方面:①数据清洗服务,包括名称索引、精确化分类和建立支持可视化的数据集;②协作和用户交互功能,利用谷歌云服务来完成数据共享、显示和编辑,并将集成化的数据

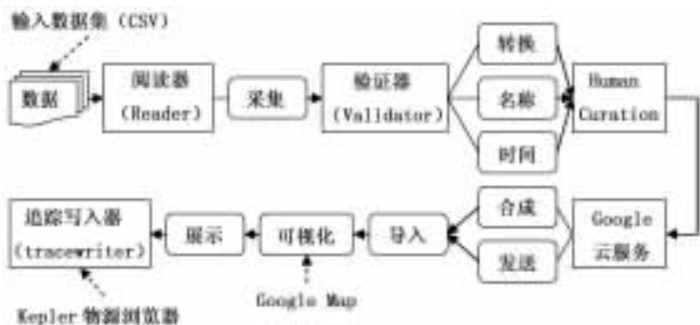


图 1 标本数据的数据 Curation 工作流程示例

输入谷歌电子表格,通过 e-mail 或 SMS 与 Curators 团队实现在线或远程共享, Curators 通过在线电子表格实现对分配的数据记录的评估和修改,并提供正确的数据或提出修改意见, Kepler 工作流程对这些内容进行调查后并根据实际需求修改原始数据集,形成报告后再次将电子表格导入谷歌以供 Curators 最终审查,在整个过程中 Kurator 工具包为专家们提供了一个解决数据 Curation 工作的创新方式, Kepler 操作者根据数据集不同的功能或质量问题汇集了不同的 Curation 目的,以编程方式精心策划可执行的任务和步骤,促进实现数据 Curation 的自动化。

DataStaR 是由康奈尔大学 Albert R. Mann 图书馆主办的数据暂存库<sup>[10]</sup>,被设想为一个平台和一组服务以方便在合作者之间的数据共享,鼓励教职员向特定领域存储库或机构库提交数字化数据和高品质元数据。为了说明 DataStaR 平台对于研究人员的实用性,可以考虑如图 2 所示的工作流程。

一个完整的工作流程包括以下几个步骤:①研究

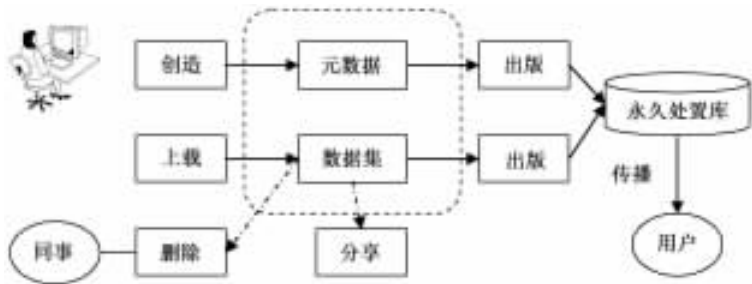


图 2 DataStaR 中的研究人员工作流程示例

人员将电子表格上传至 DataStaR 平台,并由系统自动生成元数据可供创作者选择;②研究人员将元数据权限分配给同事以支持他们及时下载数据;③数据馆员(Data Librarian)根据创作者的要求提出长期存储建议,至此,创作者的数据集已经完成,在创作者与馆员达成一致协议的基础上,确定最终版本并将数据集的副本存入康奈尔机构知识库(Cornell s IR);④创造者根据出版要求和相关政策适时调整、修改和更新元数

据记录;⑤DataStaR 提炼出用户需要的元数据。DataStaR 平台相比机构知识库( IR) 和学科库( DR) 的优势主要有:一是数据标准一般性;二是创作者、管理者和用户的互动合作;三是暂存性的存储任务而非永久性的保存。这种富有弹性的 DC 政策能够促进创作者的知识创新热情,同时也扩大了原有的学术交流圈。

### 3 Data Curation 的教育培训及研讨

数字化媒体、复杂的科学数据以及不断增长的数据需求都在深刻影响着数据保存和开发利用, Data Curation 至少需要三项技能,即图书、档案和博物馆领域选择、保存和阐释数据的专门知识,计算机科学与工程领域数据处理、探索和存储方法,以及材料科学领域认识数据内涵、来源和重点的专业知识。为了加快数据共享进程和知识转化效果以及迎合这种科研现实需求,一个新的职业——Curator 诞生了,数字世界里的 Curator 将掌握超越技术、科学和图书馆事业的多元化技能:①技术( Technology),纯粹的 Data Curation 不太可能需要高水平的编程知识, Curator 只需使用脚本语言去复制数据或验证副本的正确性,但这些需要建立在 Curator 具有专业化的知识背景和良好的认知能力;②主题事项( Subject Matter), Curator 无需像天文学家一样具备技术优势,而需要经过专业化的训练以培养一种对相关科学领域的研究程序和使用结果的理解能力;③图书馆事业和归档( Librarianship and Archiving), Curator 需要与用户互动的技能以实现有效的信息组织和科学的价值评估。Michael Lesk<sup>[11]</sup>认为未来的 Data Curator 职业将会有分工协作、职业融合和变革等特点,并对职业定位、职业路径和职业道德等现实问题作了探讨,他认为 Curator 在大数据时代将有很大的市场需求并受到追捧,图书情报教育机构应做好教育和培训准备。

Rebecca L. H- P 和 Yan Quan Liu<sup>[12]</sup>对北美 LIS 学院网站进行调查以确定足以解决海量数据( data deluge) 问题的 Data Curation 课程数量,回顾了课程的描述、目标和教学大纲,对比了课程的目标、要求、主题、任务以及确定课程的项目,结果发现:北美共有 52 所大学的 LIS 学院涉及 DC 教育,16 个研究所提供 DC 课程,不断增长的 DC 教育机构反映出 LIS 学院正在积极响应社会对 DC 专业人才的迫切需求,因此,更多的 LIS 学院应该将 DC 教育及时添加到他们的现有课程当中以跟进大数据时代的发展趋势,同时应加强协作以确定最佳的课程目标和学习成效。他们在调查中也发现了一些问题,如虽然有少数的大学制定了 DC 教育文件,但是仍然缺乏一个整体性的制度框架。

Data Cure Education Program( DCEP) 是美国伊利诺伊大学( UIUC) 香槟分校图书馆与信息科学( LIS) 研究生院<sup>[13]</sup>提供的 ALA 认可的 DC 教育理学硕士项目,由

博物馆与图书馆服务研究所( IMLS) 在 2006 年资助成立,专注于数据采集和管理、知识表示、数字化保存和归档以及数据标准和政策四大主题,提供学术和产业研究所需的基础性理论和技术。DCEP 已与国内基础科学、社会科学和人文科学数据中心建立了教育合作关系,旨在培养能够胜任 DC 管理工作的新一代 LIS 专业人才。DCEP 设立了人文和科学两大咨询委员会以保障教育质量并加强业务指导。Digital Curation Curriculum( DigCCurr) 是美国北卡罗来纳大学教堂山分校( UNC- Chapel Hill)<sup>[14]</sup>举办的 DC 教育培训项目,该项目由 IMLS 资助并于 2006 年 7 月初正式启动,2006~ 2009 年成功开办了第 1 期,目前进行的是第 2 期( DigCCurr II)。DigCCurr 项目旨在开发国际化博士水平的课程和教育网络,文化、科学、商业、卫生、教育和政府部门要获取和再利用数字资源,必须拥有博士级别的 DC 员工,因此,该项目正是为这种未来需求而准备的,涵盖卡罗莱纳州 DC 博士生奖学金计划( CDCDF)、DC 博士生课程设置、专业机构、DC 交易所( DCE) 以及 DC 研讨会 5 项内容。2012 年初北卡大学举办了“Enabling the Curation of Digital Collections”( ECDC) 学术会议<sup>[15]</sup>, ECDC 是一个专注于 DC 工具和方法的高度互动的教育研讨活动,来自不同领域的 19 位专家从“DC 需求和行为”、“特定存储库管理环境”、“元数据和文档”以及“数据转换、处理和访问”4 个方面进行演讲和演示,2013 年初延续上次主题又召开了“Curate Gear 2013”研讨会,主要讨论 DC 的专业化应用以及最新进展。

### 4 Data Curation 的发展策略及展望

约翰斯·霍普金斯大学( Johns Hopkins Univ.) 的机构知识库是作为数字图书馆系统的一个组成部分来运行的,强调长期存储的重要性和必要性,它是一个基于师生确定需求的服务集合。Sayeed G. Choudhury<sup>[16]</sup>认为,JHU 所属的 IR 本质上是一个构成数据档案的网关,支持 Data Curation 成为不断发展的网络基础设施的一部分,由于本校 IR 的学术认可度较低且出版流程不完整等原因,学者忽视了本校 IR 的需求而向其他 IR 提交数据,借此获得学术圈的认可。JHU 正在发展一种所谓的“数据科学家”或“数据人道主义者”称谓以改善图书馆和学术社区之间的关系,这些努力折射出一个现实问题:IR 若要获得成功,必须将自己的角色定位于服务多样化的分布式学术社区,并不断拓展业务流程以满足创作者的出版需求。

在生物信息学领域,Curated Data 是数据集成的先决条件已经为人们所熟知,但是忽略了一个问题:整合与分析数据过程中的编目和 Curation,以编程方式获取数据服务是生物信息学家进行试验常用的手段,工作流程中的数据整合依赖于有能力获知何种服务

是存在的,以及何时何地去发现,随意命名或缺失文档将导致数据服务的终止。工作流程本身就是一个复合的过程,当数据可被发现和理解时,才能够被池化(pooled)和重用(reused)。Goble C 和 Stevens R 等<sup>[1]</sup>认为,恰当的 Curation 应包含语义标记以促使过程能够被发现、保持和便于再利用,且语义注释是数据整合全过程的必要措施,他们提出了一个重要的公式,即“Data Curation”+“Process Curation”=“Data Integration”+“Science”。从公式可知,DC 不仅是对数据的采集、保存和利用,也是生命科学研究中数据整合的基础性阶段。

Data Conservancy<sup>[18]</sup>是一个围绕 DC 研究技术发展和社区建设的数据管理蓝图,最初由 NSF 的 DataNet 项目资助,运作主体由大学图书馆、国家数据中心、国家研究实验室以及信息科学研究和教育项目构成。数字化研究数据只有通过持续的制度承诺才能得到有效的管理和保存,研究数据 Curation 是一个多方面的问题,需要将技术、组织结构、知识和技能通过互补的方式融合在一起才能完成 DC 协作。DC 包含了一个共同的愿景:科学数据 Curation 是一种收集、整理、验证和保存数据的方式以激发更广泛的知识创新,填补了机构面对跨学科研究挑战的解决方案。DC 实例的整体性结构框架分为两大部分:一是由软件架构、设施组件、部署配置、Web 服务、API 交互等软件基础设施,以及由服务器和索引器构成的硬件基础设施及系统可扩展性功能;二是由人员编制和技能构成的组织架构。DC 制定并实施了可持续发展战略:①技术可持续性=模块化的框架+开放源码许可+发展承诺;②财务可持续性=成本分类(硬件、人员配置、管理成本)+金融模式和策略(政府拨款、机构资助、社区成员、服务费用、补助资金)+签署正式协定;③人力可持续发展是确保 Data Curation 连续性和一致性的关键;④面向团队协作的社区建设也是 DC 的重要任务。

Data Curation 过程、数据管理策略和获取工具的设计与发展是存储库面对科学和教育的重要目标之一,数字化 DC 是研究和发展中一个活跃领域,Curation 意味着对信息的精心策划的主动管理,涵盖数据的生产、保护(conservation)、保存(preservation)和获取(access),Curation 应支持数据/信息的再利用并从原有数据中催生出新的信息和知识。NEESgrid 系统集成团队 2004 年 3 月在芝加哥专门召开会议制定了一项关于 Data Curation 及相关问题的议程,与会者建议开展进一步的调查工作和其他相关的合作,包括开展 DSPACE 项目、建立 DC 研究中心、美国国会图书馆(LoC)数字保存项目、天空服务器(Skyserver)、综合注册信息系统(IRIS)、美国校际社会科学数据共享联盟(ICPSR)以及美国联邦地理数据委员会(FGDC)联邦政府地图项目等。数据和元数据管理与 Curation 是

NEESgrid 的重要组成部分,但自定义接口设计费用过高和研究人员之间的信息交换成为经费紧张时期数据共享的障碍,长期承诺和“商业”模式需要被用来维持和促进存储库数据的最佳使用,值得推荐的应对措施如下:①根据数据的有效性、质量、来源和状态来对元数据进行标记,数据管理方案应定义并实现访问控制和政策;②易于访问、组织有序和可靠记录的数据服务是最佳方案;③政策和指导方针应伴随着数据管理和 Curation 的所有阶段。从广义 DC 角度看,有必要制作和定义一个路线图以帮助获取和使用数据,从技术战略角度看,NEESgrid 社区需要确定 DC 是否为多样化的活动或严格控制过程<sup>[19]</sup>。

## 5 结语

Data Curation 的对象是科学数据,由于科学数据的多样化结构和标准以及庞大体量,执行得力的 DC 仍面临着诸多挑战。国外的研究已经将重点从概念引入转向标准设定和技术架构等方面,并在不同的研究机构中积极应用,迈出了 DC 理论与实践的重要步伐。相比国外的研究历程,国内学界对 DC 的内涵理解尚处于初级阶段,甚至连 DC 的中文译名都未能达成共识。实际上,DC 的内涵不仅在于其管理职能,还包括开发潜力,因此,对 DC 的翻译不能仅局限于其单一功能或阶段性特征,从国外的理论研究和实践应用来看,DC 的核心理念是对数据的保存、维护、管理和开发利用,笔者建议将 DC 译为“数据护理”以显示其内在的科研价值和社会效用。总结上文可知,国外 DC 研究的重点在四大方面:①来自跨领域的科学内涵界定;②基于不同技术模型的最佳实践探索;③面向专业人才塑造的教育准备;④立足可持续发展的战略规划。DC 是一个相对较新的研究主题,也是大数据时代科研创新的一项重要发展对策,应从基础研究抓起,以实践探索为依托,通过多领域跨学科的团队协作和业务分工促进 DC 在我国实现本土化。

## 参考文献

- [1] What is Digital Curation[EB/OL]. [2013-10-01]. <http://www.dcc.ac.uk/digital-curation/what-digital-curation>.
- [2] GSLIS at the University of Illinois. Specialization in Data Curation[EB/OL]. [2013-10-04]. [http://www.lis.illinois.edu/academics/programs/specializations/data\\_curation](http://www.lis.illinois.edu/academics/programs/specializations/data_curation).
- [3] Adrian Cunningham. Digital curation/digital archiving: A view from the national archives of Australia[J]. The American Archivist, 2008, 71: 530-543.
- [4] Philip Lord, Alison Macdonald. E-Science Curation Report, Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision[EB/OL]. [2013-10-01]. [http://www.jisc.ac.uk/uploaded\\_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf).
- [5] Helena Karasti, Karen S Baker, Eija Halkola. Enriching the notion of data curation in e-science: Data managing and infor-

- mation infrastructuring in the Long Term Ecological Research (LTER) network[J]. Computer Supported Cooperative Work, 2006, 15: 321-358.
- [6] Maureen Pennock. Curating e- Science Data[EB/OL]. [2013- 10- 01]. <http://www.dec.ac.uk/sites/default/files/documents/resource/briefing-papers/curating-e-science-data.pdf>.
- [7] Data Curation and Distribution in Support of Cornell University's Upper Susquehanna Agricultural Ecology Program[EB/OL]. [2013- 09- 30]. [http://ecommons.library.cornell.edu/bitstream/1813/7517/1/Steinhart\\_and\\_Lowe2007.pdf](http://ecommons.library.cornell.edu/bitstream/1813/7517/1/Steinhart_and_Lowe2007.pdf).
- [8] Jim Gray, Alexander S Szalay, Ani R Thakar, et al. Online Scientific Data Curation, Publication, and Archiving[EB/OL]. [2013- 10- 01]. <http://research.microsoft.com/pubs/64568/tr-2002-74.pdf>.
- [9] Dou L, Cao G, Morris P J, et al. Kurator: A Kepler package for data curation workflow[J]. Procedia Computer Science, 2012 (9): 1614-1619.
- [10] Gail Steinhartby. DataStaR: An Institutional Approach to Research Data Curation[EB/OL]. [2013- 10- 03]. <http://ecommons.library.cornell.edu/bitstream/1813/12668/1/iqvo1313steinhart.pdf>.
- [11] Michael Lesk. Curators of the future[J]. New Technology of Library and Information Service, 2013 (3): 1-7.
- [12] Rebecca L. Harris- Pierce, Yan Quan Liu, Is data curation education at library and information science schools in North America adequate?[J]. New Library World, 2012, 113(11): 598-613.
- [13] Center for Informatics Research in Science and Scholarship. Data Curation Education Program( DCEP) [EB/OL]. [2013- 10- 03]. <http://cirss.lis.illinois.edu/CollMeta/dcep.html>.
- [14] About Our Projects- DigCCurr II[EB/OL]. [2013- 10- 03]. <http://ils.unc.edu/digccurr/index.html>.
- [15] Alex H Poole, Christopher A Lee, Angela P Murillo. Curate Gear: Enabling the Curation of Digital Collections[EB/OL]. [2013- 10- 03]. <http://www.dlib.org/dlib/november12/poole/11poole.html>.
- [16] Sayeed G Choudhury. Case study in data curation at Johns Hopkins University[J]. Library Trends, 2008, 57(2): 211-220.
- [17] Carole Goble, Robert Stevens, Duncan Hull, et al. Data curation + process curation= data integration + science[J]. Briefings in bioinformatics, 2008, 9(6): 506-517.
- [18] The Data Conservancy- Blueprint for Data Management[EB/OL]. [2013- 10- 03]. [https://dataconservancy.org/wp-content/uploads/2012/04/data\\_conservancy\\_blueprint\\_march\\_201211.pdf](https://dataconservancy.org/wp-content/uploads/2012/04/data_conservancy_blueprint_march_201211.pdf).
- [19] Kincho H Law. Summary Report on NEESgrid's Data Curation Summit[EB/OL]. [2013- 10- 04]. [http://www.neesgrid.org/documents/TR\\_200443.pdf](http://www.neesgrid.org/documents/TR_200443.pdf).

[作者简介] 王 晴,男,1986年生,山西大学经济与管理学院2011级情报学硕士研究生。  
 收稿日期:2013- 10- 09

欢迎订阅 2014 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:华夏银行北京知春支行

账 号:81941031

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://qblz.zlzx.org/CN/volumn/home.shtml>;或发邮件至 [qingbaoziliao@263.net](mailto:qingbaoziliao@263.net)。