文 献 与 数 据 学报
第 1 卷　第 3 期　　2019 年 9 月

国外科学数据管理经验及其对我国“双一流”高校图书馆的启示

郭佳璟　樊　欣
（华中科技大学图书馆，武汉　430070）

摘　要：[**目的 / 意义**] 从国家战略和图书馆发展两个层面阐述了科学数据管理的重要性和必要性。[**方法 / 过程**] 分析了英国、美国、新加坡高校近二十年来创建的典型 DC 项目及其实践活动，归纳其科学数据管理的特点，总结了 DC 发展的一般规律。[**结果 / 结论**] 在此基础上，结合当今数据变革的时代背景，提出了我国“双一流”高校图书馆开展科学数据管理的途径，为我国“双一流”高校图书馆科学数据管理提供了一种新思路。

关键词：“双一流”高校图书馆　科学数据　数据管理　数据监管　机构知识库　知识产权
分类号：G250
DOI：10.31193/SSAP.J.ISSN.2096-6695.2019.03.03

0　引　言

科学数据管理（Data Curation，简称“DC”）主要对可以用于科学研究的数据或者科学研究中产生的数据进行收集、整理、保存、共享和跟踪。在英国，科学数据管理被称为“Digital Curation”，在美国被称为“Data Curation”。然而，此领域并不如英文直译过来的“数据管理”一样简单。原因有三，一是数据拥有很高的学科专业背景；二是对数据的精准性要求较高；三是管理的数据要能够被合理的发现和多次利用，并且能让数据服务于未来十年二十年甚至上百年的科研工作。因此，DC 在实现知识资产保护的同时，还要保存科学数据的长久价值，以提高科研人员创新研究的效率。

科学数据是科学研究中产生的重要资产和资源，是科学研究持续或协作研究的重要基础。但有些科学数据收集难度大、成本高，导致利用率较低，例如航空航天类数据；有些科学数据生长

[作者简介] 郭佳璟（ORCID：0000-0002-2588-4581），女，华中科技大学图书馆，助理馆员，硕士，研究方向为图书馆及信息科学、数字化资源，Email：guojiajing@hust.edu.cn；樊欣（ORCID：0000-0001-8476-6705），男，华中科技大学图书馆，馆员，硕士，研究方向为通信与信息系统、图书馆学，Email：fanxin@hust.edu.cn。

周期不稳定或者较长，难以得到有效的跟踪和更新，影响更深入的科学研究，例如社会心理学的数据；有些科学数据生长较快，但传播途径较广，导致容易被污染，例如宏观经济学数据，等等。这些数据上的问题，增加了科学研究的成本和难度。因此，建立有效的科学数据管理平台，不仅可以保护知识产权，减少学术造假问题，帮助科学研究工作者专注到研究当中去，而且可以大大提高科学数据的使用率，提高科学研究的成效。高等学校图书馆，作为办学的基本条件、科学研究的重要支撑，其科学数据的保存将是大数据时代图书馆馆藏的重要内容和为用户服务的重要基础，图书馆应该高度重视并积极建立完善的科学数据管理业务流程和管理平台，更好地为高等学校教学科研事业的发展提供保障。

1　科学数据管理的重要性

1.1　科学数据管理是国家科技发展的战略需要

1.1.1　美国联邦政府颁布的有关科学数据管理法律法规

美国联邦政府长期坚持对科学研究数据的管理与共享，这是美国促进科技发展、提升创新能力和创新效率的重要手段。早在 1966 年，先后经过 4 次修订的《信息自由法》（Freedom of Information Act）颁布实施，为美国政府信息和数字资源公开化奠定了法律基础[1]。随后，美国相继颁布实施了一系列国家法律，为日后的科学数据管理开放共享体系打下了坚实的法律基础，例如 1966 年的《信息技术管理改革法》、1974 年的《隐私法》等。1987 年白宫新闻办公室发布了里根总统旨在推动科学技术不断发展的 11 项具体措施，明确要求农业部、商务部、能源部、卫生和人类服务部以及国家航空航天局分别制订“技术分享”计划，以便与美国的工商企业和高等院校长期共同从事基础研究和应用研究[2]。2002 年美国又颁布了《电子政府法》和《联邦信息安全管理法》。2013 年 2 月 22 日，美国奥巴马政府再一次承诺，纳税人有权访问并获得科学研究数据。并且在政府的备忘录中提及，联邦政府将向公众公开超过 1 亿美元研发支出产生的研究成果[3]。

1.1.2　中国政府出台的有关科学数据管理办法

科学数据是国家科技创新发展和经济社会发展的基础性战略资源。我国在科学数据开发利用、开放共享和安全保护等方面还有很大进步空间。近年来，中国政府越来越重视科学数据的开放与共享。2017 年 12 月 8 日，习近平总书记在主持中共中央政治局第二次集体学习时强调，实施国家大数据战略，加快建设数字中国。可见，发展迅猛的中国开始将构建数字中国放在了非常重要的位置[4]。2018 年 3 月 17 日，国务院办公厅印发了关于科学数据管理办法的通知，首次从国家层面出台了《科学数据管理办法》（以下简称《办法》）[5]。《办法》第 1 条明确定义了科学数据，即“在自然科学、工程技术科学等领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据”。同时《办法》第 8 条明确指出，“有关科研院所、高等院校”在科学数据管理方面的职责为：“建立加强规范科学数据管理，做好数据的授权和保密工作，推动科学数据开放共享，建立有效的激励机制。”《办法》的出台旨在大力推进科学数据资源的开放与共享，特别是国家科技

计划项目产生的数据，要求进行强制性汇交，否则项目不予验收，为科学数据的收集和管理奠定了制度基础。

1.2　科学数据管理是图书馆服务转型的必然选择

1.2.1　科学数据管理是传统图书馆的基本职能

耶鲁大学著名英国文学教授廷克（Chauncey Brewster Tinker）1924 年在给毕业生的演讲中指出，大学有三大标志性要素：学生、教师和藏书[6]。高等学校图书馆是学校办学的三大支柱之一，是学校的文献信息中心，是为教学和科学研究服务的学术性机构[7]。传统图书馆员的核心工作是资源的收集、整理、加工、保存，用户完全依赖图书馆获取信息资源，图书馆的职能之一是保存人类科学文化遗产，可以说，科学数据管理是传统图书馆的基本保存职能。

1.2.2　科学数据管理是现代图书馆的转型要求

随着计算机技术和网络技术的飞速发展，信息资源的载体和传播方式发生了技术性变革。用户获取信息的行为也相应发生了变化，其越来越熟悉和习惯利用网络获取信息资源，对图书馆的依存度大大降低，因此图书馆寻求第二个突破口，新型服务能力成为衡量图书馆价值的重要指标。图书馆积极开展信息资源的整合、线上信息平台的建设，多样化、个性化、智能化的服务措施纷纷出台，其目的是帮助读者在海量的信息资源中，快速准确地找到合适的资源，提升用户对图书馆的依存度，满足用户对所需文献的信息需求。然而，努力寻求变化去适应用户行为的图书馆似乎仍然面临两大不可逾越的问题：一是纸电资源大量重复，各类信息资源不完整、不规范，同时各类信息资源零散地存储在繁杂又相对独立的图书馆资源系统中；二是图书馆多样化的服务难以实现质的飞跃。当资源分散在图书馆各个独立的系统中，且数据脏乱的时候，图书馆员需要花费大量的时间在各个系统中寻找，对资源进行细化、清洁、整合，再推送给用户。因此，零散、重复存储在图书馆各独立系统中脏乱的数据，成为图书馆提升服务质量、提高服务效率、降低服务成本的绊脚石。有效整合图书馆的纸电资源，将数据从繁杂的独立系统中提取、清洗、整合，成为当今图书馆必须解决的问题。

因此，高校图书馆构建适合自身馆藏特征和学科服务特色的数据管理平台，实现对馆藏数据的拆分、归类、清洗、整合与共享，不仅是国家科技发展的战略需要，也是图书馆信息服务转型升级的必然选择。

2　国外开展科学数据管理的实践

2.1　国外开展科学数据管理的起步

早在二十年前，英国和美国就已经开始 DC 的实践。英国数据监管中心（Digital Curation Centre，简称 DCC）和美国 DataCite 是最早成立的非盈利的专门研究对科学数据的收集、存储、保存和出版的专业机构。英国 DCC 于 2004 年成立，是开展 DC 工作的领头羊。其成立之初，致力于辅助英国高等教育和继续教育中的数字化保存和管理（digital preservation and curation）工作。后来得益于 SCARP（Disciplinary Approaches to Sharing, Curation, Reuse and Preservation）项目，DCC 的工作重心转向对数据存储、共享、再利用、管理和保存等的深入研究。在英国爱丁堡大

学、格拉斯哥大学和巴斯大学这三所高校以及继续教育研究中心的长期支持下，DCC 研究并提出了一套全面完整的 DC 业务流程模板，即数据管理的全生命周期模型（The DCC Curation Lifecycle Model，见图 1）[8]。这一流程模板是 DCC 最重要的研究成果，是一套理想的具有指导性的对科学数据进行管理的方法论。其被后来全球开展 DC 的实践者所借鉴，各机构可完全根据自身特点和需求，选择开展的细化程度，创造适合自身的 DC 业务规范，从而达到对科学数据相应阶段的有效管理。

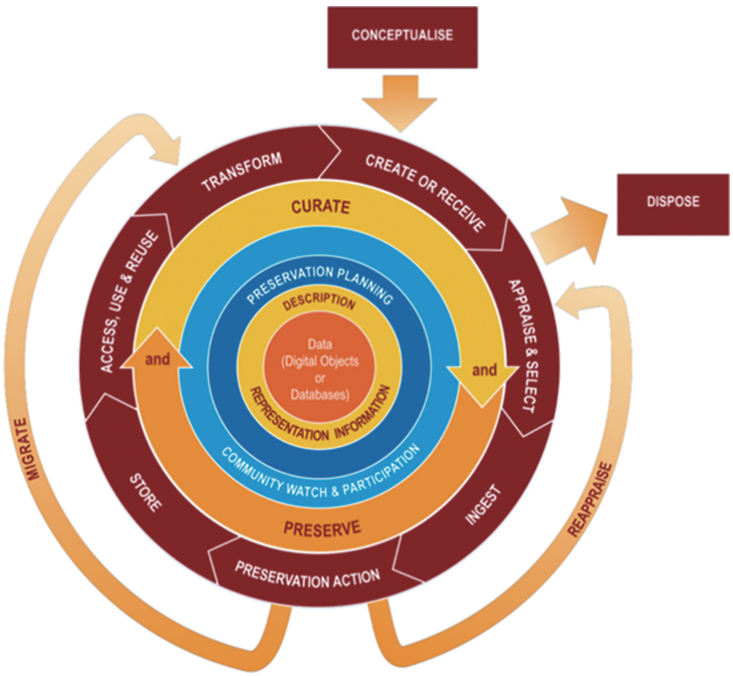

图 1　DCC 数据管理的全生命周期模型

如图 1 所示，数据管理的全生命周期模型包含以下四个方面：

（1）描述：用适当的模板和标准，对数据集进行全面充分的描述。描述包括数据的行政方面、技术方面、结构方面和元数据储存等。

（2）保存计划：根据数据生命周期的每个阶段，制定详尽的方案与措施。

（3）监管：创建对数据有效的监管环境，开发或者利用模板、工具、软件等措施，对数据进行长期、有效的监管。

（4）存储和发现：让数据集在每个生命周期阶段，都能得到安全的存储和妥善的利用。

从 DCC 数据管理的全生命周期模型可以看出，开展科学数据管理可以从两个方面入手。一是加强对成本极高的科学数据的管理与保护。对这种科学数据的管理，需要强大的学科背景做支撑，无论是评估还是对元数据进行描述和做数据保护计划。二是全面收集各类学科可以用到的数据，为开展研究降低难度和成本。这方面成功的关键是构建良好、便捷的数据存储、发现、共享平台。

美国高校在美国国家科学基金会（NSF）的支持下，约翰霍普金斯大学、康奈尔大学、田纳西大学、加州大学、新墨西哥大学等五所高校的图书馆于 2007 年开始，五年间先后成功开展数据监管研究项目，并成功创立了各自的盈利模式，有效地得到持续性的发展[9]。

近几年，新加坡国立大学、新加坡管理大学和南洋理工大学三所高校借鉴欧美的实践经验与研究成果，也着手开展科学数据管理实践，构建出符合自身学校特点的科学数据管理平台，同时开展了一系列服务。

2.2　国外开展科学数据管理的方式

纵观全球众多高校图书馆开展科学数据管理的成功案例和 DCC 的数据管理全生命周期模型研究成果发现，因为高校的特性不同，图书馆的定位、馆藏、馆员、需求、规模和经费不同，所以其选择开展科学数据管理的方式也有所不同。目前大致可以分成两种方式：服务主导型和数据主导型[10]。以下分别介绍国外科学数据管理实践的两种方式。

2.2.1　服务主导型

服务主导型的数据管理方式，更多的是针对某一特定科研项目，对科研活动全过程（包括科研项目前期计划、中期进展和后期验收维护）中产生的科学数据进行收集、整理、编目、跟踪等。这类科学数据的共性是：数据不易获取，收集成本很高，解读需要较高的专业知识，科研耗时长且具有可持续发展性。服务的对象往往是机构或项目组成员。

2.2.1.1　牛津大学的 EIDCSR 项目

牛津大学作为全球顶尖的研究型大学，拥有大量宝贵的科学数据。EIDCSR（Embedding Institutional Data Curation Services in Research）项目，旨在通过制度、业务流程和技术，构建完善的架构，对机构内已经产生的科学数据进行有效的、长期性的收集、存储、管理、保护和分享，并促进更多的研究项目加入进来，从而最大化地发挥科学数据的价值，营造更加安全、便捷、自由的科学研究氛围。

牛津大学 EIDCSR 项目组包含三类专业人员：图书馆员、科学家和计算机中心工作人员。项目组成员共同努力，基于英国 DCC 提出的科学数据管理全生命周期模型，依托 3D 心脏项目，创建了如图 2 所示的科学数据管理框架[11]。这个框架中包含：（1）数据管理政策；（2）数据审计框架方法论；（3）数据存储、获取框架。

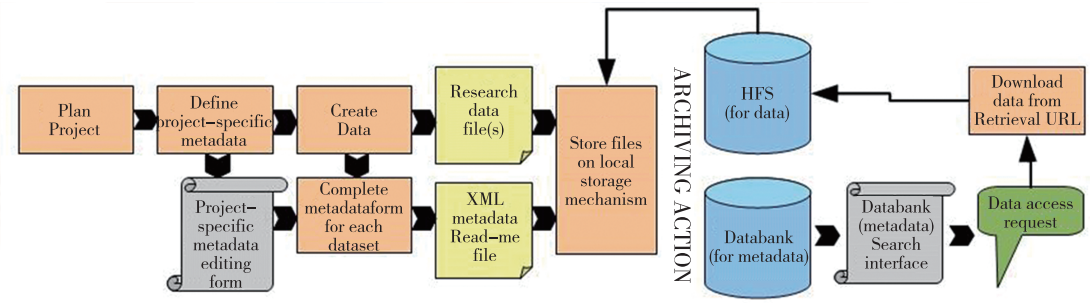

图 2　EIDCSR 科学数据管理框架

2.2.1.2　爱丁堡大学科学数据管理

爱丁堡大学科学数据管理成功的关键在于其拥有权威的科学数据管理顶层设计和清晰完善的DC 业务流程与管理制度。在组织框架方面，爱丁堡大学成立了专门的科学数据管理委员会，其成员基本覆盖了科学研究的各个领域，避免了在数据管理过程中服务人员出现知识盲区[12]。爱丁堡大学将科学数据管理分成了五大组成部分：数据管理计划、数据同步、数据存储、数据评估和数据共享（见图 3）[13]。每一部分均定义了目标集和每个目标的优先级，同时又针对每一个目标定义了需要做的工作、预计得到的结果和需要的人员与时间。

爱丁堡大学在管理政策方面还创建了一个良好、完善的研究社区，以帮助各类型的研究者，从定义科学研究的方向开始，予以全方位的支持。进而在这个过程中让科研产生的科学数据得到有效的管理、跟踪、保存与分享。

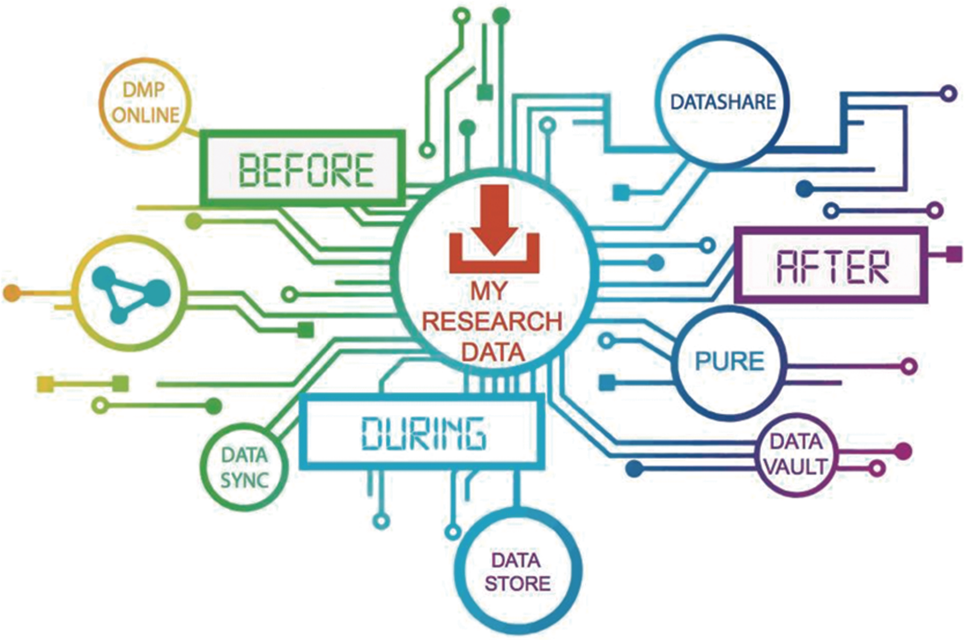

图 3　爱丁堡大学科学数据管理框架

2.2.1.3　新加坡管理大学图书馆

新加坡管理大学图书馆则是将科学数据管理融入研究服务模块当中，以学科研究为主线，先通过划分学科，将目前新加坡管理大学图书馆支持的研究服务分为 11 个学科，然后根据每个学科的特性，将用户需求分成更细小的分支。例如，将会计学科的用户需求分为寻找财务年报，如何使用 Bloomberg，如何使用 Eikon，如何使用 S&P Capatial IQ，MPA 白皮书，会计风险准则，如何开始做会计学术研究等 8 个分支。将学科专业中的用户需求归类后再进入详细页面，可以让用户在图书馆页面上寻求帮助的时候不模糊意图和目标。科学数据管理服务则归在了学科学术研究模块当中，提供了科研指南、endnote、科学数据管理、科研数据资源、出版指南、机构知识库、政策等服务[14]。新加坡管理大学图书馆不仅整理了高校内产生的宝贵的知识

资产，同时提供了高校外专业机构和企业的数据资源，给予该学科科研工作者较全面的科学数据资源。

2.2.2　数据主导型

数据主导型的管理方式，主要是对已经生成的相对较干净的数据进行存储、管理与共享。对于这类科学数据，市场获取途径相对较多，数据产生的周期相对明确。数据服务对象可以拓展，不局限于本机构内，而是可以延伸到社会上的其他专业机构。因此，数据主导型管理方式更加关注数据管理平台的功能和实用性，符合数据的特征和生命周期，能满足各种用户的需求。以数据主导型方式开展科学数据管理的案例如下。

2.2.2.1　康奈尔大学 DataStaR 平台

康奈尔大学 DataStaR 是根据用户、研究者需求，自行开发的语义平台。其亮点在于：利用语义 Web 技术作为 DataStaR 的元数据基础架构，将现有的元数据框架转化为 OWL 本体，融入 DataStaR 系统，可促进以前创建的表述在新的元数据中的应用[9]。换言之，随着科研数据的增多，基于 DataStaR 平台，数据可以不断融入新的集合中，得到更好的跟踪与更新。同时 DataStaR 平台从访问层区别、控制用户访问和获取数据的方式，不仅保证了数据的安全性，同时还给数据使用团队提升了一个开放自由的管理数据的环境。

2.2.2.2　约翰霍普金斯大学 DRCC 数据研发和监管中心 DDC 项目

DDC（Digital Data Curation）是约翰霍普金斯大学图书馆 DRCC（Digital Research and Curation Center)、SDSS（Sloan Digital Sky Survey）和 NVO（National Virtual Observatory）三所机构共同对数字化天文数据集开展的数据监管项目。因为庞大的天文数据主要为图像数据，不适合当时约翰霍普金斯大学图书馆主要用文档的方式存储资源，所以 DDC 开发构建了数据监管系统的架构，将已有的数据仓储和电子出版系统对接，利用 OAI-ORE ReM 数据模型，将数据和论文建立关联，同时保存了完整的元数据信息[9]。

2.2.2.3　南洋理工大学 DR-NTU（Data）平台

南洋理工大学数据管理平台是基于哈佛大学开发的 Dataverse 开源软件构建的数据存储库。Dataverse 是基于科学数据生命周期特点开发的数据监管平台，可满足从做数据管理计划开始，对数据进行详尽的记录、跟踪、保存、访问、发现、共享、引用和分析等功能要求。同时，Dataverse 还增加了数据可视化、模板定制、用户评论、格式自动切换等人性化的功能。DR-NTU (Data) 平台是在专业科学数据管理平台 Dataverse 上进行的二次开发，设计的功能不仅尽可能地满足科研工作者使用数据时可能需要的功能，同时还将科学数据管理从科研工作者的角度，分成了五大模块：创建和添加数据集、修改数据集、元数据、搜寻数据集和引用数据集。为每个模块撰写了白皮书，让用户尽可能全面地了解每个模块的目的与使用方法。

3　我国开展科学数据管理的现状、问题与途径

3.1　我国组织机构开展科学数据管理的现状

近年来，国内关于 DC 的讨论越来越多，而且也有相关组织机构开展了科学数据管理的实

践。例如：“十二五”期间，中国科学院面向科技创新和科研信息化需求，启动“科技数据资源整合与共享工程”项目，建设了负责全院科学数据资源整合、存储、备份和共享服务的中科院数据云（Data Cloud Of CAS）[15]；从 2009 年开始，中国人民大学中国调查与数据中心，在中国国家自然科学基金重点项目资助下，建设经济与社会数据共享平台（CNSDA），实现对经济、综合（家庭）、健康、社会类科学数据进行清洁、编码、整合、管理、存储、共享等功能[16]，等等。国内高校图书馆开展科学数据管理比英美发达国家起步晚，分析我国高校图书馆在开展科学数据管理方面存在的问题，发现与英美发达国家的差距，并找到解决问题的途径和方法，有助于加速推动我国高校图书馆科学数据管理的进程以及提升开发利用、开放共享的水平。

3.2 “双一流”高校图书馆开展科学数据管理的问题

近年来国内高校图书馆越来越重视 DC 工作，目前基本都是采取数据主导型的方式开展科学数据管理，即建设机构知识库平台，来实现对本校产生的科学数据（如期刊论文、会议论文、学位论文、专利等）的收集、存储和共享。科学数据是科学研究的基础，它不仅应该在机构知识库中被当成知识资产，受到合理的保护，同时应该被图书馆合理地管理，在机构知识库平台上被发现与再利用，持续地创造知识。同时，数据类型也不仅仅是公开发表的论文和学位论文、专利等，应该更强调科研创新中产生的各种有价值的数据以及数据的动态更新和维护管理。通过分析国内高校建设的科学数据管理平台以及平台使用情况，发现其存在以下几个主要问题：

（1）目标不清晰。开展科学数据实践之初，未能明确定义符合自身特点的科学数据管理的初衷和目的。

（2）未能设计业务流程。缺乏对科学数据管理服务的深入研究，未能设计出一系列符合中国高校特征的科学数据管理业务流程。

（3）没有规划平台功能。构建平台之初，未能根据自身馆藏特点，规划定义平台的功能和使用目的。

（4）平台的发现性低、定位单一、利用率小是普遍存在的问题。

（5）机构知识库对科学数据管理的功能有待进一步提升。

日本国立情报学研究所 2017 年 11 月 7 日发布信息，宣布该所与欧洲核子研究中心、日本国立物质材料研究所合作，联合开发下一代机构知识库系统。该系统不仅保持原有机构知识库的功能，而且着力研究数据管理，将二者融合为一体[4]。显见现阶段的机构知识库对科学数据的管理存在不足，需要进一步研究科学数据管理的全生命周期，提升机构知识库对科学数据的管理的功能和水平。

3.3 “双一流”高校图书馆开展科学数据管理的途径

国外科学数据管理实践的成功案例，对我国高校图书馆开展 DC 具有重要的启示和借鉴意义。要想在高校图书馆成功开展和推广 DC，应当从以下三个途径来考虑。

3.3.1　明确“双一流”高校开展科学数据管理的目标

图书馆开展 DC 的目的是更好地应用科学数据，实现知识服务，为学校师生开展创新研究提供支撑和保障。因此，开展 DC 首先需要从机构自身的特性出发，从学校特点、学科特征、专业组成、图书馆馆藏与布局、图书馆经费等方面去定义和规划开展科学数据管理的初衷与目标。

如果完全照搬欧美科学数据管理先行者的所有模式，并不一定能满足师生学习和研究的需求，不一定能满足学校教学和科研的需求，那么这样的科学数据管理就不是成功的。正如新加坡的三所大学（NUS, NTU, SMU）照搬 DCC 的数据管理业务流程，尽管都构建了良好的数据管理平台，可实现用户检索、上传、下载、访问、使用等数据管理的所有功能要求，但 NTU 调研发现，87.5% 的研究人员从未使用过本校的数据管理平台[17]，新加坡三所大学开展的科学数据管理案例也并未在国际上成为典范。北京大学、武汉大学、复旦大学等高校均已建设机构知识库平台，并且近几年都非常重视平台的推广与应用，但科学数据管理并未在本校形成良好的反响与盈利模式。可见，照搬欧美先行者成功案例，并不一定能在本土收到同样的成效。

高校图书馆全面调研与分析学校的特点、专业特色、科研方向、服务需求等，是制定科学数据管理目标的前提。图书馆不仅需要全面地聆听、理解校内科研工作者的需求，同时要与学校相关职能部门进行充分沟通，明确开展科学数据管理的出发点与服务目标，合理制定建设和开展项目需要实现的具体内容，根据目标内容决定科学数据管理的类型。

不同的目标内容决定了科学数据管理的类型。如果采用服务主导型，则需更加重视学校优势学科、特色学科或重大的科学研究项目，制定符合科研全过程、全方位的科学数据管理与服务，制定符合科研项目特定需求的科学数据管理业务流程，对科学数据进行全生命周期管理。如果采用数据主导型，则需要对数据来源和经费两个方面进行考量，着重构建符合用户行为特征与需求的科学数据管理平台。通过加强平台的互通性和功能特点，建立有效的服务机制，促进更多的用户使用科学数据管理平台，从而促进科技创新。

3.3.2　建设一个良好的数据管理平台

开展科学数据管理的目标不同，其建设的科学数据管理平台也有所不同。图书馆可以根据数据的来源（如图书馆已有的元数据集、第三方的数据接入、在期刊上发表的论文数据集等）和数据服务的目标，设计数据管理平台的方案，一般数据管理平台应具备以下功能：

（1）能满足目标要求的各学科各类型的数据集管理。实现对高校成果的全面性、永久性的收集、保存、更新、共享等管理；平台科学数据集要长期、连续、反复地被使用。

（2）平台应具有良好的拓展性。尽量与现有的机构库、科研人员交流软件或者平台从技术上建立关联。文档与数据之间很难用统一的格式或技术来存储，但可以通过关联数据等将机构库中的文档与监护中的数据关联起来。这不仅能让科研人员简单快捷地获取所需的科研数据，提高机构库的使用率，而且能激发科研人员共享数据的意愿。如华中科技大学的“一张表”平台与图书馆机构知识库的平台就具备这种关联性，不仅便于数据互通互用，同时可以通过多方校核，提高数据的准确性，也使得图书馆构建的机构知识库平台不

变成孤岛。未来平台还可以与计算、分析平台对接，让科学数据管理平台能够贯穿整个科研项目。

（3）平台的数据集要具备可拓展性。如果数据集只有 CSV 或者 XML 等单一的存储结构，就可能导致数据集无法满足数据处理分析软件的格式要求，无法适用于各种数据使用方式，导致科学数据管理平台的数据不好用。数据集的可拓展性有两个方面的要求：一是要有不同的数据存储结构，以满足不同数据分析软件的格式要求；二是将图书馆的数据编织成知识网或知识图谱，更好地开展知识服务和智慧服务。

（4）平台的模块设计和界面设计要能让使用者迅速全面了解数据集的信息，判定是否满足自己的科学研究需要，从而减少科研工作中不必要的时间成本。

（5）安全且便于数据提交者缴存和更新科学数据，让用户能够在平台上完全放心地提交、管理、使用数据。

3.3.3 图书馆要组建一支结构合理的团队

高校图书馆开展科学数据管理是大数据环境下图书馆知识服务的重要内容，要做好这项工作，首先要组建一支学科专业背景结构合理、年龄结构合理、岗位职责明确的强有力的团队。至少由图书馆分馆馆长牵头负责，团队成员主要来源于具有较高技能、高素质或者学科背景的图书馆员，但又不乏其他职能部门、科研团队中的协作成员。如果采取服务主导型，则团队中需要更多具有较强学科专业背景的学科馆员，可以通过单一项目着手，创建一个能将 DC 业务流程全面贯穿科研活动，且能融入科研人员社交网络的科学数据管理业务服务流程。如果采用数据主导型，则需要考虑融入更多的信息数据和 IT 系统能力强的馆员，他们不仅仅需要搭建更加安全且能满足用户需求的数据管理平台，同时还需要：

（1）选择引进或自建的数据，或者授权用户提交的数据，扩充自身数据资源。

（2）对同类多源数据进行整合、清洁，从而让用户能以最低的成本获取较高质量的数据集。

（3）对每个数据集进行元数据创建和制定，从而实现对每个数据集长期、有效地跟踪记录与更新。

（4）创建 Linked Data，将图书馆资源数据编织成知识图谱，提升图书馆知识服务质量。

除此之外，这个团队一方面需要弄清开展科学数据管理的目标和特点，其中包括明确适用管理的数据背景、定义、范围及要求等。因此，需要与学校的相关职能部门、学院，科研院所或重点实验室等协调和交流，对人员的专业背景、交流能力等有很高的要求。另外，为了保证科学数据管理工作的可持续发展，需要制定合理的科学数据管理制度和业务流程。通过制度和业务流程的建立，明确科学数据管理各个环节参与人员的角色定位和职责，如图书馆、课题负责人、学院、科研管理办公室、IT 部门等，明确数据管理的知识产权要求，从而达到科学数据全生命周期的安全和有效管理。

4 结束语

2018 年国务院办公厅印发了关于科学数据管理办法的通知，彰显了国家在大数据时代对科

学数据管理的高度重视。高校作为科技创新重地，开展科学数据管理也就成了当务之急。科学数据管理的重要性在图书馆界基本达成共识，但工作量大、复杂，实施难度大，需要专业人员参与，建设周期长。如果图书馆在建设初期就能够明确目标，做好规划，组建科学合理的建设队伍和管理机制，DC 建设工作将会事半功倍。

科学数据管理是一项系统工程，涉及方方面面，在科学数据管理平台建设、数据清洗和数据模板建设过程中，需要相关业务公司合作开发，但图书馆员参与全过程的重要性不容小觑。开展科学数据管理的图书馆员不仅需要具备图情知识、学科专业素养和一定的专业知识，还需要具有对数据的敏感度和数据管理知识，保障 DC 系统正常运行和维护。科学数据是在科研活动中产生的，因此对科学数据的管理也需要学校教学与科研人员的积极参与、学校和学院相关部门的大力支持以及各级领导的高度重视，为科学数据管理提供有力的组织保证，这将更加有利于提升科学数据管理水平和共享使用成效。在科学数据管理平台建设中除了要满足数据管理的全生命周期各项要求外，还要特别重视数据作为科技创新的成果或中间结果，要对科学数据使用者和生产者的行为进行规范，体现对科研人员智力劳动成果的尊重，加强对其知识产权的保护。

【参考文献】

[1]王炼 . 美国联邦政府科学数据管理政策及实践 [J]. 全球科技经济瞭望 , 2018(7):47–51.

[2]李明德 . 美国发展科学技术的具体措施 [J]. 国际科技交流 ,1987(8):1–2.

[3] MICHAEL S. [2013–02–22].https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding–public–access–results–federally–funded–research.

[4]吴建中 . 吴建中：大数据成为重要战略资源，它背后是怎样一场全球运动？[EB/OL]. [2018–04–03]. https://www.jfdaily.com/news/detail?id=81295.

[5]国务院办公厅 . 国务院办公厅关于印发科学数据管理办法的通知 [Z].2018.

[6]吴建中 . 走向第三代图书馆 [J]. 图书馆杂志 ,2016,(6):4–9.

[7]中华人民共和国教育部 . 教育部关于印发《普通高等学校图书馆规程》的通知 [Z].2015.

[8] Curation Lifecycle Model [EB/OL]. [2019–02–01].http://www.dcc.ac.uk/resources/curation–lifecycle–model.

[9]蔚海燕 , 卫军朝 . 国外高校数据监管项目的调研与分析 [J]. 图书情报工作 , 2014(22):38–47.

[10]卫军朝 , 张春芳 . 国内外科学数据管理平台比较研究 [J]. 图书情报知识 ,2017(5):97–107.

[11] EIDCSR – Final Report – v.1.1 – 14/02/2011 [R].2011.

[12]王海彪 , 卫军朝 . 科学数据管理关键因素研究——基于爱丁堡大学科学数据管理实践及启示 [J] 图书馆杂志 ,2017(1):20–26.

[13] Research Data Management Roadmap [EB/OL]. [2017–09–30].http://www.ed.ac.uk/is/rdm–roadmap.

[14]汪满容 , 刘桂锋 , 刘琼 . 新加坡高校图书馆科研数据管理服务调研与启示 [J] 图书馆学研究 ,2018(9):64–71+22.

[15]黎建辉 , 周园春 , 胡良霖 , 等 . 中国科学院科学数据云建设与服务 [J] 大数据 ,2016(6):3–13.

[16]刘兹恒 , 曾丽莹 . 我国高校科研数据管理与共享平台调研与比较分析 [J] 情报资料工作 , 2017,(6):90–95.

[17]付少雄 , 陈晓宇 , 赵海平 , 等 . 新加坡高校的科学数据管理实践体系 [J] 图书馆论坛 2019,(2):141–148.

郭佳璟，樊欣 . 国外科学数据管理经验及其对我国 “双一流”高校图书馆的启示［J］. 文献与数据学报，2019，1（3）：026–037.

Data Curation Experience of International Universities and Insights for Domestic “Double First-Class” University Libraries

GUO Jiajing　FAN Xin

(Huazhong University of Science and Technology Library, Wuhan 430070, China)

Abstract: ［**Purpose/significance**］This paper illustrates the significance and necessity of Data Curation (DC) from the perspectives of national strategies and library development. ［**Method/process**］Typical DC projects and activities are analyzed according to the experience of universities in the United Kingdom,United States and Singapore for the past two decades, based on which we summarize the features and development patterns of DC. ［**Result/conclusion**］Regarding the upcoming era of data revolution, we propose the roadmap of facilitating DC development for “Double First-Class” university libraries, which delivers a novel ideal.

Keywords: “Double First-Class”; University libraries; Science data; Data management; Data curation; Institutional repositories; Intellectual property

（本文责编：周　霞）