

美国高校图书馆数据管理计划 服务及启示*

王思明

(东北师范大学信息科学与技术学院, 长春 130117)

摘要: 对美国排名前10的高校图书馆数据管理计划服务进行调研。在全面了解服务现状基础上, 分析其服务开展的动因, 并对服务活动流程、服务对象、协作机构、活动内容、效果评估等内容进行梳理总结。最后提出我国高校图书馆开展数据管理计划服务的建议。

关键词: 数据共享; 研究数据管理; 数据管理计划; 高校图书馆

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2018.12.005

作为人类发现、探索、解释自然与社会的忠实记录, 科学数据逐渐获得与传统文献相同的学术地位^[1]。作为科研活动所必需的基础性生产资料, 它是学术记录的重要组成部分, 是支撑研究结论的重要事实或结果。科学数据日益被组织成为可发现的、可获取的、可互操作的、可重用的资源^[2]。2010年初, 美国国家自然科学基金会 (National Science Foundation, NSF) 发布了项目管理指南^[3], 规定2011年1月18日起所有提交至NSF的申请项目必须以附件形式提交一份不超过2页纸的数据管理计划 (data management plan, DMP)。美国国立卫生研究院 (National Institute of Health, NIH) 同样在其资助指南中指出, 任何在1年内直接资助费用超过50万美元的研究项目必须提供一份数据共享计划, 对项目的研究数据进行明确的描述与管理, 用以保证数据的共享与重用^[4]。2018年5月, SPARC Europe (欧洲学术出版与学术资源联盟) 与国际数字监护中心 (Digital Curation Center, DCC) 合作, 发布了《开放数据驱动因素研究报告》^[5], 报告中明确指出积极建立数据管理计划是推动开放数据驱动发展的第二大因素。

数据管理计划作为研究人员记录他们描述、存储、保护, 以及共享其科学数据的计划和措施, 不仅有助于

图书馆的科学数据管理团队了解研究人员的数据管理实践, 发现科学数据管理服务的需求, 也能促进图书馆的科学数据管理团队与本校内部的相关利益机构之间建立合作网络, 保证科学数据管理服务的有效推进^[6]。鉴于此, 本文对国外高校图书馆的数据管理计划服务实践进行梳理, 进而为国内高校图书馆开展数据管理计划服务提出建议。

1 美国高校图书馆开展数据管理计划服务的现状

根据美国U.S.News《美国新闻和世界报道》2018年发布的美国大学综合排名^[7], 选取排名前10的国立大学, 即对普林斯顿大学 (Princeton University)^[8]、哈佛大学 (Harvard University)^[9]、哥伦比亚大学 (Columbia University)、麻省理工学院 (Massachusetts Institute of Technology)^[10]、芝加哥大学 (University of Chicago)^[11]、耶鲁大学 (Yale University)^[12]、斯坦福大学 (Stanford University)^[13]、杜克大学 (Duke University)^[14]、宾夕法尼亚大学 (University of Pennsylvania)^[15]、约翰·霍普金斯大学 (Johns

*本研究得到国家社会科学基金项目“图书馆书目数据社会化应用研究” (编号: 15BTQ018) 和中央高校基本科研业务专项资金项目“科学数据引用行为及其影响因素研究” (编号: 130021822) 资助。

Hopkins University)^[16]进行调研。发现哥伦比亚大学图书馆尚未开展数据管理计划服务,故本文以其余的9所高校图书馆的DMP服务为研究对象。

1.1 概念界定

最初的数据管理计划来源于联邦资助机构、非营利的资助机构和一些企业资助者对于资助申请人提交一份“数据管理计划”文件的要求。为满足这些机构的要求,高校图书馆开展数据管理计划的服务工作。有关数据管理计划的定义,9所高校图书馆都引用典型研究资助机构的概念来说明,如NSF及NEH(National Endowment for the Humanities)有关数据管理计划的解释。其中,NSF将数据管理计划定义为数据共享的扩展,并进一步将数据管理计划描述为一个不超过2页的文档,用来描述如何管理和存储来自研究的所有数据;NEH将数据管理计划描述为数据共享政策方案,同时提供数据共享指南来帮助研究人员对于数据管理计划撰写。图书馆通过给出这些机构对于数据管理计划的明确概念,来帮助研究人员更好地理解研究数据及数据管理计划的重要性。同时,为进一步加深研究人员对于数据管理计划的理解,9所高校图书馆均选用DMPTool进行解释。

1.2 服务动因

随着技术的进步,数字文件的共享变得更加容易,

资金提供者(公共和私人)和期刊越来越需要数据共享以实现研究的可再现性和透明度。数据管理已经成为研究人员需要遵循的重要实践,各个高校图书馆也逐渐将数据管理计划服务引入其整体服务体系。其中,麻省理工学院图书馆将数据管理计划服务引入数据管理服务体系,并明确指出数据管理计划在提升研究人员的研究影响力、节省研究团队的时间、允许自身和其他人在理解和使用研究数据等方面具有明显的作用;芝加哥大学图书馆将数据管理计划作为其研究管理服务的首要环节,并表明对于研究人员而言,一个有效的数据管理计划不仅能够防止错误进而提升研究分析的质量,还能够防止数据丢失。本文将数据管理计划服务的动因归结为:①满足资助机构、期刊等数据共享的要求;②保证项目研究的连续性;③降低数据的丢失、被盗、错误使用的风险,确保数据的完整;④易于数据的长期存档及保存;⑤节省研究团队的时间和资源,有助于预防错误,以提高分析的质量;⑥允许在学科内与跨学科之间共享数据,增加原始数据的价值;⑦促进研究数据的可发现、可获取、可复制及可验证;⑧保存和共享数据以帮助自己的研究获得认可,增加研究结果的可靠性;⑨鼓励数据引用以增加研究的影响。

1.3 协作机构及服务对象

通过调研图书馆开展服务的协作单位及服务对象,发现9所高校图书馆各具特色,见表1。

表1 国外数据管理计划服务概况

机构名称	本校协作单位	服务对象
普林斯顿大学图书馆	研究完整性和保障办公室、IT中心、数字仓储	研究人员、研究团队
哈佛大学图书馆	HUIT信息安全办公室、研究副教务办公室、社会科学定量研究所、FAS研究管理服务办公室	研究人员、研究团队
芝加哥大学图书馆	信息学研究中心、计算机研究中心、机构知识库	研究人员、研究团队
耶鲁大学图书馆	RDCG(Research Data Consultation Group)涵盖所有学科,其成员包括数据管理、元数据、信息技术和保存方面的专家,自然科学、社会科学和人文学科领域的专家,以及计算研究中心	学生、研究人员、图书馆员、数据专家
麻省理工学院图书馆	数据存储库、信息与技术学院	学生、教师、研究人员、研究团队
斯坦福大学图书馆	IT中心、研究办公室、研究计算中心、社会科学资源中心、数字存储库	学生、研究人员、研究团队
宾夕法尼亚大学图书馆	改为加州数字图书馆系统、数据保护中心、研究委员会、个别学院、研究支持事务厅、研究计划发展办公室、宾州医学院学术计算系统、技术信息系统办公室	研究人员、研究团队
杜克大学图书馆	数据存储库、研究支持办公室	研究人员
约翰·霍普金斯大学图书馆	研究管理办公室、商业与政府项目办公室	研究人员

1.3.1 协作机构

9所高校图书馆在开展数据管理计划服务时都会选择与本校其他部门合作,其中研究办公室、计算机中心、数据存储库等机构是普遍的合作单位。图书馆的主要职能是提供撰写数据管理计划各内容要素相关资源、解答疑问、相关技能培训等活动。多部门合作的数据管理计划服务成功弥补了图书馆能力不够的问题,也保障了服务的深入性、专业性、全面性。

1.3.2 服务对象

高校图书馆开展的数据管理计划主要服务的对象还是研究人员,其中不仅包含独立的研究人员也包含某个研究团队。麻省理工学院图书馆的服务对象较为广

泛,除本校的研究人员、研究团队外,还包括图书馆员及数据专家。但9所高校图书馆中明确将学生作为服务对象的情况较少,仅有耶鲁大学、麻省理工大学图书馆及斯坦福大学图书馆有所涉及。

1.4 服务内容

1.4.1 服务定位

调研发现,9所高校图书馆的数据管理计划服务都处于图书馆研究数据管理下,并被定位在整体研究数据管理首位,同数据捕获及格式转化、数据的描述与记录、数据共享与存储、数据引用研究等数据管理服务模块组成基础服务层。基础服务层、相关资源层及培训指导层上下配合,形成体系化的流程(见图1)。

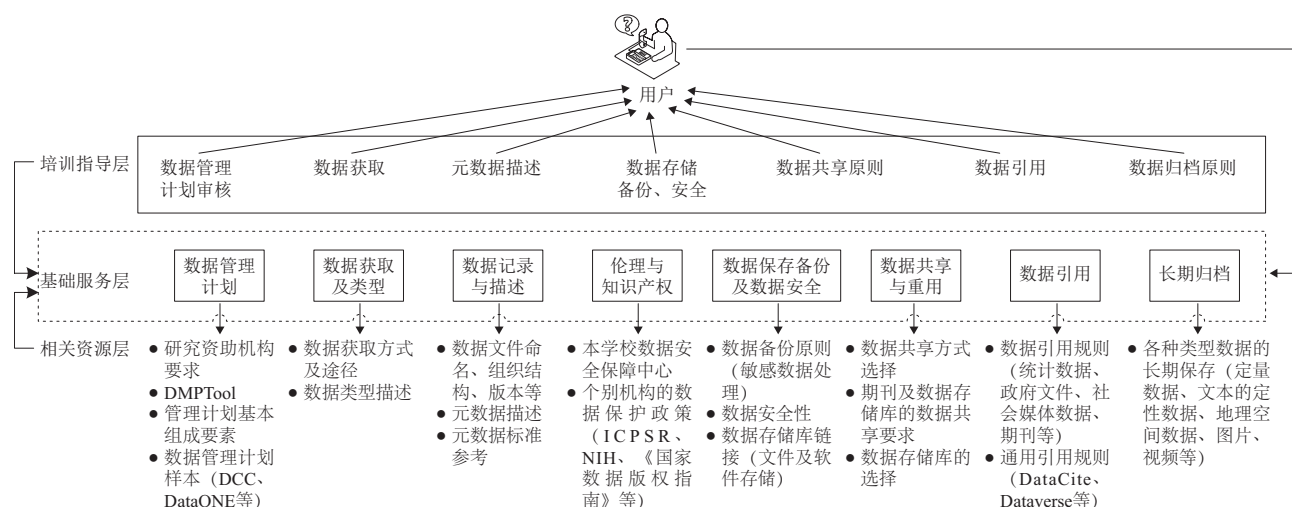


图1 9所高校图书馆研究数据管理服务体系

1.4.2 组成要素及相关资源

数据管理计划服务,关键点就是要先明确数据管理计划中的要素类型,以帮助研究人员明确主要的方向。在要素类型的选定上,9所高校图书馆通过提供通用文档,给出研究人员编写DMP过程时可能需要考虑的要素。调研发现,高校图书馆对于要素的选定较多集中在定义研究团队成员的角色和职责、数据类型和格式、捕获方法和文件命名(元数据)、道德与知识产权(保护及隐私)、数据访问、共享和重用、短期存储和数据管理、数据归档及长期保存等要素类型上,同时图书馆会通过相应资源链接的方式帮助研究人员理解并撰写数据管理计划中的每一项内容。

数据管理计划组成要素的相关支撑资源主要包括3种类型。

(1) 研究授予机构或研究资助机构的数据管理/数据共享要素类型要求。如,NIH和、NSF、National Oceanic and Atmospheric Administration等,这种资源链接主要是帮助研究人员保证自身撰写的要素内容符合研究资助机构的总体要求。

(2) 数据管理计划参考样本。如DMPTool、New England Collaborative、DataONE等中所给出的数据管理计划的样例,给出参考模板以指导研究人员解决在实际撰写中的困扰及问题。

(3) 单一学科的在线数据存储库、期刊等的数据共享、数据管理要求。如大气研究数据中心、生物学数据中

心、气候数据中心等,这种资源的目的是进一步确保研究人员所撰写的每个要素内容符合本学科数据库、期刊的实际共享要求。事先让研究人员了解日后可能存储数据的数据中心的要求,以帮助研究人员真正地作到数据的长期保存与共享。

1.4.3 服务评价

目前9所高校图书馆对于数据管理计划服务模块的评价尚未规范化。如杜克大学图书馆通过提供研究数据管理顾问的方式,对服务对象撰写的的数据管理计划进行实时评价与审查。约翰·霍普金斯大学图书馆和芝加哥大学图书馆通过审核研究人员撰写的的数据管理计划方式开展评价而通过提供标准化的文档帮助研究人员实现自我评价,是多数高校图书馆的主要方式。但这些评价活动都过于零散,且主要由服务对象自主选择,并不能达到预期的评价效果。

1.4.4 培训与指导

目前图书馆开展的与数据管理计划服务有关的培训与指导大体可以分3种,即数据管理需求评估、数据管理计划撰写实际问题解决、数据存储库的选择与使用。

(1) 数据管理需求评估。图书馆帮助研究人员评估项目、实验室及自身研究过程中的数据管理需求,通过需求的评估来指导具体的服务活动。

(2) 数据管理计划撰写实际问题解决。图书馆在这方面提供的培训与指导的内容较为广泛,不仅包括数据管理、文档组织、数据引用、数据共享等问题,还包括DMPTool等工具使用问题。但它并不是仅包含以上内容,用户还可以根据自己在实际中遇到的问题提出咨询。

(3) 数据存储库的选择与使用。部门高校图书馆会设置自己的数据存储库,以辅助研究人员的数据存储管理与共享,如斯坦福大学图书馆、杜克大学图书馆。所以高校图书馆还会提供对于数据存储库的使用方面的指导与培训,以解决在数据保存与共享方面遇到的问题。

以上3种培训与指导,主要是通过提供专门的数据管理人员邮箱咨询及定期小组研讨会方式进行。

1.4.5 总结与对比

9所高校图书馆数据管理计划服务内容可总结为相关服务模块、组成要素、资源链接、评估、培训与指导五大方面,现将各个高校在五个方面具体服务开展情况进行总结(见表2)。

表2 美国9所高校图书馆数据管理计划服务的服务内容

服务内容	具体内容	普林斯顿大学图书馆	哈佛大学图书馆	芝加哥大学图书馆	耶鲁大学图书馆	麻省理工学院图书馆	斯坦福大学图书馆	宾夕法尼亚大学图书馆	杜克大学图书馆	约翰霍普金斯大学
与数据管理计划相配合的服务模块	数据获取及类型	√					√			√
	数据记录与描述	√	√	√	√	√	√	√	√	√
	数据保存备份与数据安全	√	√	√	√	√	√	√	√	√
	数据共享与重用	√	√	√	√	√	√	√	√	√
	伦理与知识产权	√	√			√	√			√
	数据引用	√	√		√	√		√	√	
明确组成要素	数据长期归档	√	√		√	√			√	√
	数据管理计划组成要素	√	√	√		√	√		√	
数据管理计划撰写需要的资源链接	研究资助机构的数据管理要求	√	√	√	√	√	√	√	√	√
	数据存储库或期刊数据管理要求	√	√		√	√	√	√		√
	相关机构的数据管理计划参考样本	√	√	√	√	√	√		√	
数据管理计划评估	实时评估			√			√		√	√
	最终效果评估						√		√	√
培训与指导	明确数据管理需求		√		√	√	√	√		√
	要素撰写指导	√	√	√		√	√		√	√
	数据存储指导		√	√	√	√	√	√	√	√

对于服务内容,主要从数据管理计划配合服务建设及其具体内容建设两个角度进行总结与分析。首先,在提供与数据管理计划服务相配合的模块服务上,数据记录与描述、数据保存备份与数据安全、数据共享与重用3个主要方面成为9所高校图书馆数据管理计划主要关注的内容。而在数据获取及类型、数据引用、数据长期归档方面关注度较低,尤其是在数据获取及数据类型描述上极少提及。其次,在数据管理计划自身的服务内容上,在明确数据管理计划组成要素、提供数据管理计划撰写需要的资源链接两大服务内容上,9家图书馆的开展情况差距不大。就数据管理计划评价服务活动而言,可以明显看出开展情况较差,仅斯坦福大学、杜克大学、约翰·霍普金斯大学3所大学涉及。在相关的培训与指导工作上,都主要集中在要素的撰写及数据存储上,而对于用户的数据管理需求评估关注度较低。

目前在数据管理计划服务的建设方面,约翰·霍普金斯大学、斯坦福大学的图书馆建设较好,不仅内容全面而且涉及数据管理计划的各个方面,值得我国高校图书馆学习与借鉴。

2 美国高校图书馆数据管理计划服务对我国的启示

国外高校图书馆数据管理计划服务的调查研究,对我国高校图书馆数据管理计划服务的开展具有重要意义。

2.1 明确服务提供主体

数据管理计划服务活动中,图书馆作为高校数据管理计划服务的主要提供者和主导者,应该采用多部门、多组织合作的组织形式,组建多元化的服务团队。开展数据管理计划服务不仅需要图书馆中的各部门配合,同时还需要与学校其他相关组织合作,如计算机中心、研究中心/研究办、信息安全中心、数据存储中心等。其中,计算机中心可以提供元数据描述、数据格式转化等技术咨询服务,研究中心可以在数据共享与重用方面提供协助。各个部门与组织共同参与科学数据管理计划服务中,通过资源共享、信息及技术支持,获得最优的协作服务效果,形成多职能和多元化的服务团队,以更好地满足服务对象各阶段的需求,构建富有生命力的团队。

2.2 内容框架

随着研究的进行及用户需求变化,数据管理计划服务的内容与标准也会相应变化,这是一个动态过程。从明确数据管理需求到计划的完成及评价,服务标准由低层次到高层次,服务方式从简单到复杂。就我国高校图书馆而言,笔者认为可与目前高校图书馆的科研服务相结合,将内容框架定位为四个方面:一是将数据管理计划服务纳入目前的科学数据管理服务体系;二是明确我国的数据管理计划组成要素;三是选择体系化、系统化的服务模式,保证各个要素之间的配合;四是数据管理计划的评价。

2.2.1 将数据管理计划服务纳入高校科学数据管理服务体系

目前国内高校图书馆已经明确开展与科学数据管理相关的系列服务,如北京大学图书馆、复旦大学图书馆、武汉大学图书馆、上海外国语大学图书馆,另有部分高校图书馆虽然在网站中没有明确标示,但会在科研支撑模块中提供一些数据支持服务^[17]。图书馆在后续的数据管理计划服务开展上,可与现有的科学支持服务结合,尝试用数据管理计划来串联现有的科研服务。但要特别关注用户对于数据管理计划的认识与了解,要帮助用户明晰数据管理计划的重要性。调查显示,目前国内高校图书馆虽然已经开始关注学生、研究人员及数据馆员^[18]的科学数据素养教育及培训,但在具体的实践活动方面还很缺乏,因此高校图书馆亟需提高学生及研究人员等在构建数据管理计划方面的积极性,提高其数据管理的能力与效率。

2.2.2 明确数据管理计划组成要素

根据目前国内高校图书馆科研数据管理活动开展情况和服务需求,将数据管理计划服务的基本组成要素设定为数据获取、数据描述、伦理与知识产权、数据共享与重用、数据引用、数据保存6个基本组成要素。6个要素紧紧围绕科研全周期,全面地涉及了科研活动的各个阶段,保证了数据管理计划服务的整体性、系统性。

(1) 数据获取。详细描述如何生成或获取数据,包括明确规定数据采集的步骤,并概述相关主体的角色和职责;数据的类型,包括观察数据、实验数据、模

拟数据或派生/编译数据。

(2) 数据描述。上下文细节记录与描述(元数据); 数据的处理和文件的组织, 包括数据处理的方式、版本控制、文件格式、文件命名及组织格式。

(3) 伦理与知识产权。数据保护、机密数据的保护、可共享数据的处理、数据知识产权与版权。

(4) 数据共享与重用。数据管理的相关责任主体; 自由的访问、重用、重新分配数据的说明; 文件的可用性, 包括文件的命名、特定软件包的支持; 敏感数据的处理; 数据重用说明与指导; 数据的知识产权说明。

(5) 数据引用。引用格式的选择; 提供永久标识符; 尽可能在最详细的级别引用数据集, 并提供适当的

版本; 尽量保证所引用的数据集与论文的链接。

(6) 数据保存与数据安全。需要短期保存的数据的方式、存储周期; 长期保存方式(存储库的选择)、存储周期; 数据安全(敏感数据的安全性); 支持数据使用的软件或工具的存档。

2.2.3 构建体系化的数据管理计划服务模式

构建体系化的服务模式也能帮助活动中的参与主体明确自身的责任与任务, 防止角色混乱, 保障服务的顺利开展。笔者结合对9所高校图书馆的数据管理计划服务的案例研究, 构建数据管理计划服务模式(见图2)。

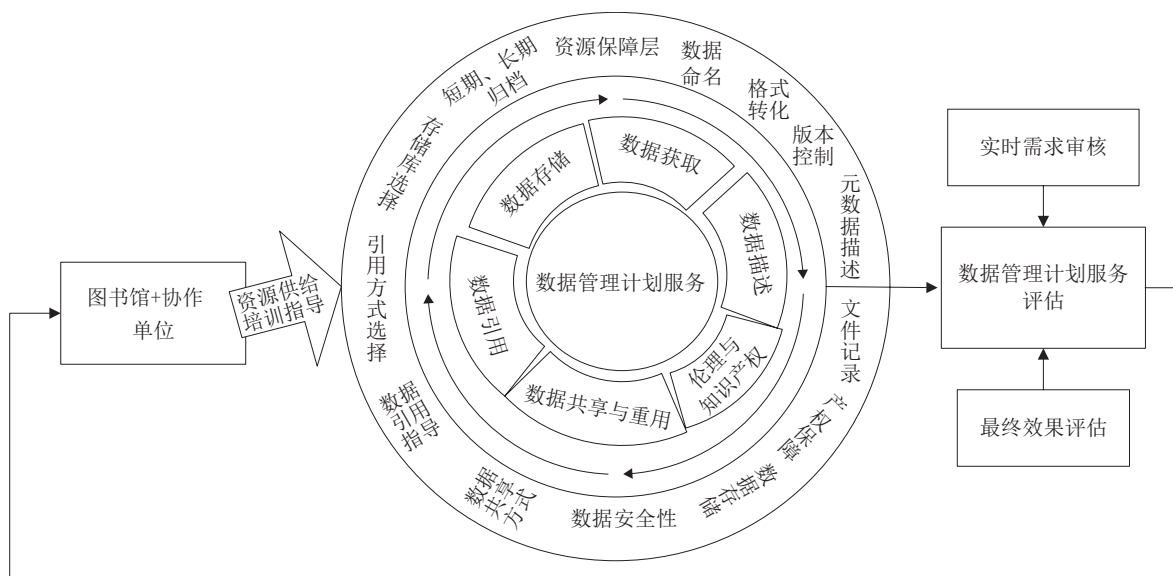


图2 数据管理计划服务模式

在数据管理计划服务部分, 笔者将服务模式设定为两个层次, 中心层为数据管理计划的6个主要组成部分, 与其对应的是相关资源保障层, 主要目的是保障6个要素阶段的顺利完成。其中可能包含的资源支持有数据获取阶段的数据命名、格式转化、版本控制等; 数据描述阶段的元数据描述、文件记录等; 伦理与知识产权阶段的产权保障、数据安全等; 数据共享与重用阶段的数据存储、数据共享方式等; 数据引用阶段的引用指导、引用方式选择等; 数据保存阶段的存储库选择、短期及长期归档等。

2.2.4 重点关注数据管理的评估体系

本文特别针对基础服务层构建了计划评估体系,

具体分为两个环节, 第一环节是数据管理计划各个阶段的实时评估, 目的是明确服务对象在各个阶段的需求, 使服务更有针对性, 进而保障数据管理计划的顺利完成; 第二环节是终期管理计划效果的评估, 特指所有用户已经完成的数据管理计划的评估, 目的是通过完成的数据管理计划文档来进一步识别数据管理计划整体服务的缺失, 并发现用户在数据管理方面尚存问题, 为图书馆后续的培训及指导提供支持。

3 结语

随着我国对科学数据共享、重用和科研创新能力日益重视, 高校有必要拓展图书馆服务的功能和形式, 构建体系化、多元化的科学数据管理服务。本文对美国

高校图书馆数据管理计划服务的服务概况、服务内容、具体流程进行探讨,希望对我国高校图书馆开展科学数据管理计划服务有一定的启示。

本文虽然对数据管理计划的基本组成要素进行了设定,但因为各个高校学科体系不同,导致各个专业的科学数据具有差异性,由于数据管理计划的基本组成要素可能存在不同,因此高校图书馆要结合自身服务用户特点,构建切合用户需求的数据管理计划支持服务。

参考文献

- [1] CALVERT P. Big Data, little data: Scholarship in the networked world [J]. The Electronic Library, 2017, 35 (2): 391.
- [2] DIEKEMA A R, WESOLEK A, WALTERS C. The NSF/NIH effect: Surveying the effect of data management requirements on faculty, sponsored programs, and institutional repositories [J]. The Journal of Academic Librarianship, 2014, 40 (3/4): 322-331.
- [3] NSF ENG Data Management Plan Requirements [EB/OL]. [2018-08-20]. <https://www.nsf.gov/eng/general/dmp.jsp>.
- [4] NIH. Federal funding accountability and transparency act [EB/OL]. [2018-08-20]. https://grants.nih.gov/grants/public_accountability/ffata.htm.
- [5] SPARC Europe. Data management. Learning from the innovators: 2 leading funders compared [EB/OL]. [2018-09-06]. <https://sparc europe.org/who-we-are/about-us/>.
- [6] 王丹丹. 数据管理计划在图书馆科学数据管理服务中的应用 [J]. 图书情报工作, 2018, 62 (1): 76-81.
- [7] Best Colleges [EB/OL]. [2018-09-16]. <https://www.usnews.com/education>.
- [8] Research Data Management at Princeton [EB/OL]. [2018-09-15]. <https://libguides.princeton.edu/c.php?g=102546&p=665863>.
- [9] Research Data Management at Harvard [EB/OL]. [2018-10-05]. <https://guides.library.harvard.edu/dmp.edu/dmp>.
- [10] MIT Libraries [EB/OL]. [2018-09-18]. <https://libraries.mit.edu/data-management/>.
- [11] The University of CHICAGO Library [EB/OL]. [2018-09-15]. <http://guides.lib.uchicago.edu/c.php?g=298312&p=1989781>.
- [12] Yale University Library [EB/OL]. [2018-10-02]. <https://guides.library.yale.edu/datamanagement?hs=a>.
- [13] Stanford Libraries [EB/OL]. [2018-09-20]. <http://library.stanford.edu/research/data-management-services>.
- [14] Research Data Management at Duke [EB/OL]. [2018-10-02]. <https://library.duke.edu/data/data-management/>.
- [15] University of Pennsylvania [EB/OL]. [2018-09-27]. <https://guides.library.upenn.edu/data-management>.
- [16] Johns Hopkins Data Management Services [EB/OL]. [2018-10-06]. <http://dms.data.jhu.edu/data-management-resources/plan-research/>.
- [17] 周力虹, 段欣余, 宋雅倩. 我国高校图书馆科研数据管理服务调查与分析 [J]. 图书情报工作, 2017, 61 (20): 77-86.
- [18] 魏来, 高希然. 大数据背景下高校数据馆员的角色定位 [J]. 情报资料工作, 2015 (5): 90-94.

作者简介

王思明, 女, 1989年生, 硕士研究生, 研究方向: 信息服务, E-mail: 1310141065@qq.com。

Data Management Planning Service of American University Libraries and Its Enlightenment

WANG SiMing

(School of Information Science and Technology, Northeast Normal University, Changchun 130117, China)

Abstract: Investigate the top ten university library data management plan services in the United States. On the basis of a comprehensive understanding of the status quo of the service, analyze the motivation of its service development, and summarize the contents of the service activity process, service object, collaboration organization, activity content, and effect evaluation. Finally, the suggestions for the data management plan service of university libraries in China are put forward.

Keywords: Data Sharing; Research Data Management; Data Management Plan; University Library

(收稿日期: 2018-11-30)