

美国国家标准与技术研究院科研数据 管理实践及启示*

陈晨

(天津外国语学院图书馆, 天津 300270)

摘要: 美国国家标准与技术研究院的科研数据管理服务是目前美国最成功的数据管理实践典范, 本文利用网络调查法对美国国家标准与技术研究院的科研数据管理实践开展背景、服务体系等方面进行调查分析, 总结其科研数据管理的成功经验, 以期为我国开展科研数据管理实践提供参考。

关键词: 科研数据; 数据管理; 数据服务; NIST

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2019.04.007

美国国家标准和技术研究院(National Institute of Standards and Technology, NIST)成立于1901年, 是美国高端的研究机构。其下有多个研究所和实验室, 涵盖纳米科学、化学、物理学等多种学科, 拥有研究人员3 000人, 合作研究员2 700人。2012年, NIST的科学家第五次荣获诺贝尔奖, 其纳米科学、计量科学等多项技术走在了世界前沿^[1]。NIST卓越的研究成就离不开成功的科研数据管理服务的支持与推动。NIST经过多年的探索和实践, 形成具有特色的科研数据管理体系, 是目前美国评价最好的科研数据管理实践典型案例。NIST图书馆获得美国国会图书馆颁发的2003年、2008年、2013年年度联邦图书馆奖, 2007年马里兰州卓越表现奖, 彰显了其取得的科研数据管理创新实践与成就。本文研究NIST科研数据管理实践经验, 以期为我国开展科研数据管理实践提供启示与参考。

1 NIST数据管理开展背景

NIST科研数据管理服务是在外部因素和内部需求的驱动下开展的。外部因素包括如下方面。①基金资助机构要求。2011年1月, 美国科学基金会(National

Science Foundation, NSF)推出新的科研数据共享政策, 要求提交到NSF的申报书中必须包括数据管理计划(Data Management Plan, DMP)^[2]。自此, 数据分析成为数据时代科研人员必须掌握的新技能。②美国政府经费投入。NIST的研究和运行经费主要由联邦政府投入, 列入国会预算。每笔拨款都有针对性和具体要求, 美国国会通过的《2014综合拨款法案》^[3], 明确要求NIST的财政拨款中要有0.98亿美元用于实验室科研数据的开放存取、数据库基础设施建设的专项基金。内部因素包括如下方面。①机构资源管理的需要。NIST内部的科学家、工程师和客座研究人员每年大约发表1 500篇论文^[4], 并在评估标准参考数据的过程中产生大量用以分析和支持学术界、工业界和其他政府机构使用的已发表研究成果的数据, 除了科学文献, 这些科研产出的数据同样是机构的重要资源。②研究人员的需要。研究人员在申报科研项目之初需要考虑数据管理问题, 对于研究人员产出的小型分散的科研数据, NIST自身的数据管理系统更能契合科研人员的需求。以上内外部因素, 促进了NIST科研数据管理服务的开展。

*本研究得到天津外国语学院青年科研项目“Web 2.0环境下外语院校在线信息素质合作教育模式研究”(编号: 15QN16)资助。

2 NIST科研数据管理服务体系

为有效地实施科研数据管理服务, NIST建立了专门的科学数据管理服务网站, 形成以数据管理平台、工具资源、人力资源、政策规范为服务基础, 以数据管理计划、数据组织、数据存储与共享、标准参考数据、数据管理培训为服务内容的科研数据管理体系, 发展成为较为完善的服务工作流。

2.1 数据管理平台

NIST研发的科研数据管理平台^[5], 主要包括学科数据知识库 (Disciplinary Repository)、图像廊 (Images)、企业数据清单 (Enterprise Data Inventory, EDI)、标准参考数据库 (Standard Reference Data, SRD) 和一站式数据公共访问平台 (Common Access Platform)。

(1) 学科数据知识库。主要包括医学、化学、物理学等自然科学科研数据, 尤其注重经验数据和事实数据的收集、归纳。NIST将实验室科研人员在科研过程中产生的大量权威数据源经评审、汇总, 开发成跨学科的多类型数据存储库。如Atomic Spectroscopy Database是用于研究辐射原子或离子能级领域的数据库, 包含99个元素的观察转变数据和89个元素的能量水平数据, 以及超过9万条的能级记录和18万条的线性数据记录, 其中7.3万条记录具有转换概率^[6], 实现跨学科科研数据的有效管理。

(2) 图像廊。图像廊提供了大量的视频和图像, 包含原美国国家标准局 (NBS) 的历史照片, 公共和商业事务部、研究所、NIST博物馆等收藏的技术标准和其他文物的信息和图像。用户可以检索各种新的技术、实验、反应、产品等图片信息。

(3) 企业数据清单。企业数据清单是具有用户友好前端的数据集目录, 包含元数据的信息索引, 以及关于数据访问的位置信息和访问路径。其功能类似于卡片目录, 可以链接到DMP, 将数据集审核和批准后, 分配数字对象标识符 (Digital Object Identifier, DOI) 并上传到网络云中的安全存储库, 方便用户轻松地查找和使用任何来源的数据集。

(4) 标准参考数据库。它是NIST计量服务部提供的标准参考数据, 受1968年的标准参考数据法案 (P.L.90-396) 保护。目前, 标准参考数据已有116个数

据库产品^[7], 基本涵盖所有的科学技术领域。

(5) 一站式数据公共访问平台。一站式数据公共访问平台是NIST提供科学数据公共访问的数据门户。它将机构内用户发布的数据以及机构外权威数据资源进行整合, 提供各类型数据存储库间的互操作, 实现对已发布数据或其他开放数据的导航和一站式检索。

为激发将NIST标准参考数据与移动设备合并的新方法, 2015年NIST联合新泽西州普林斯顿的Mero Apps公司、德克萨斯州的Zachary Ratliff、肯塔基州的Daniel Graham公司、弗吉尼亚州的MetroStar System等高新技术公司开发了25款APP, 由评审团专家评出前三名获奖的应用程序并推广使用, 实现用户通过移动设备快速访问NIST数据^[8]。

2.2 工具资源

NIST通过内部资源整合和外部资源联合的方式提供丰富的数据管理信息资源与实现工具, 帮助研究人员制定满足基金组织要求的数据管理计划, 实现科研数据的有效管理。主要包括以下方面。①基金要求的解读工具。推荐使用本机构开发的工具Minerva, 制定数据管理计划; 提供工具的使用指南及资助机构的模板和优秀案例供参考。②文件组织格式。NIST的数据存储库提供的文件组织格式主要有文本格式、图像格式、视频格式、文件命名与重命名格式等。③元数据标准。提供不同学科资源类型的元数据标准以及组成要素, 并在机构内构建各类元数据描述框架。④数据管理工具。NIST对各领域科研项目软件工具按照科研数据生命周期进行收集整理, 定制开发了开放软件下载平台。集成了18种数据分析软件工具, 并详细描述每种工具的版本、软件类型、开发者、系统要求, 用户可根据需求选择合适的管理软件。⑤名词术语。提供了50多个术语解释, 包括科研数据、DOI、最终出版物、开放获取等。

2.3 人力资源

2009年, NIST成立数据与信息学办公室 (Office of Data and Informatics, ODI), 与研究图书馆信息服务办公室 (Information Services Office, ISO) 通过实验室联络计划 (Lab Liaison Program) 合作开展科研数据管理服务^[9]。ODI由各实验室学科领域专家、企业数据架构与设计方面的专家组成。ODI和ISO之间不是

自上而下的行政隶属关系,而是横向的平行协作关系。这种平等的机构管理模式,使ISO和ODI各司其职,各施所长,作为一个整体合作参与开放获取计划,探索NIST科研数据管理问题的解决方案,集中各部门的人力、物力、财力,共同提升科研数据服务质量。

在实验室联络计划中,数据馆员作为联络员被分派到NIST实验室科研项目中,成为各实验室和ODI的联系枢纽,数据馆员与指派组织中的管理者及科学家建立并保持密切的工作关系,详细了解各实验室的数据管理需求,为他们提供个性化定制服务。ODI工作人员还积极参与外部数据计划,如国家数据服务(the National Data Service)和研究数据联盟(the Research Data Alliance)。这种横向协作的机构成员组织模式,拓展了科研数据管理服务的深度和广度,取得了很好的服务效果。

2.4 政策规范

2013年,在白宫科技政策办公室(OSTP)发布的一系列备忘录和行政命令基础上,联邦机构制定了“完全与开放”数据共享政策,以增强政府资助科学研究成果的获取^[10]。为了贯彻落实白宫科技政策,NIST领导层通过审视本机构研究人员当前出版和管理数据的实践,制定了开放获取政策和系统实施方案,具体措施有:①完善NIST的各级组织单位数据访问流程,持续更新和评估开放获取政策,以保持数据长期保存和访问的有效性和相关性,减少相关的成本和管理负担;②与利益相关方及其他研究资助组织建立合作伙伴关系,提高出版物数据的兼容性和获取途径,探索和改进NIST开放获取数据管理政策的新方法;③经同行评审的科研论文和科研数据在发表12个月内可免费下载;④出版物所有权归属于作者和原始出版商;⑤为所有NIST资助的科研项目所产生的科研数据制定有效的数据管理规划;⑥为NIST员工和NIST资助的校外研究人员提供科研数据管理教育培训,以帮助他们遵守NIST政策。除了上述实施措施外,NIST还制定了评估开放获取政策指标,每年对政策的实施效果进行评价,以不断完善和解决科研数据管理过程中出现的问题^[11]。

2.5 服务内容

(1) 数据管理计划服务。数据管理计划服务是

NIST科研数据管理中开展率最高的一项服务。作为申请美国联邦资助科研项目的一部分,为满足科研人员的需求,NIST数据与信息学办公室开发了一款名为Minerva的网络数据管理工具。Minerva内置了美国科学基金会和主要项目资助机构的数据管理计划内容需求、模板文档及优秀案例。用户可按其提供的模式及流程分4步创建数据管理计划,即选择科研资助机构DMP模板、填写科研项目数据计划描述信息、完成DMP细节描述、生成科研项目数据管理计划。Minerva比类似的应用程序功能更为广泛,它囊括了一份完整的数据管理计划相关信息,通常包括活动描述、数据类型创建、保存和存储信息、公共访问级别4个领域^[12]。2014年,ODI工作人员对Minerva进行了Alpha可用性测试,并对工具的功能布局和语言使用提出修改建议。2015年春季,NIST正式发布MinervaV2版本,迄今为止,已有超过1 000位研究人员提交了2 000多条数据DMP记录。

(2) 数据组织服务。为了确保科研人员在未来科研工作中能够方便查找、获取、管理和共享数据,数据馆员提供了正式的元数据标准、数据格式、文件命名规范、文件结构,以便对不同属性数据资源进行有序化地组织、分类、存储、处理。如在NIST材料基因组计划(Materials Genome Initiative)项目中,图书馆信息服务办公室首先对特定存储库系统(SIdora)进行评估,审查研究人员工作流程,推荐最佳数据组织标准,并协助开发材料软件模型和代码目录,确定开始此协作的最佳位置是为目录开发元数据标准框架。此架构中使用的一些元数据字段映射到Dublin Core,但大多数字段特定于描述一般软件的属性(codelanguage和operating system),涉及与材料科学相关的软件(如Scale)或与软件相关的法律问题(如exportControls),扩展字段包括opSystemName、opSystemVersion等,ISO工作人员已使用此架构描述超过75种产品来填充目录的初始版本,并通过Datacite共享已发布科研数据^[3]。

(3) 数据存储与共享服务。为协助研究人员将科研数据长期保存于存储设施,便于今后的数据共享,NIST建立了公共访问存档系统NIST Digital Archive,存储机构内产生的元数据和经过同行评审的手稿与出版物,该系统将文档使用当前存储架构中最先进技术兼容的格式进行存储,支持文件中包含的图形、表格、数据文件或补充信息等任何文件类型。对于NIST的机构外数据,美国联邦计划官员(Federal Program

Officer, FPO) 可以选择通过NIST的数据管理基础设施(即MIDAS, 相关数据存储库和NIST科研数据门户) 存储此类数据, 此类数据必须符合存储条件: ①项目申请者是个人或研究小组; ②资助协议赋予NIST公开数据的权利; ③数据具有很高的重复使用潜力, 其持续可用性符合NIST的利益; ④FPO和部门主管对数据质量审核评估^[13]。资助接受者需要将他们的数据稿件副本转发给联邦计划官员或指定人员, 后者将该论文上传到NIST的编辑审查系统。NIST还与美国国立卫生研究院(NIH) 合作, 利用现有的PubMed Central (PMC) 存储库系统, 作为NIST公共访问存档系统, 通过PMC公共访问存储系统的NIST接口界面, 用户可自动检索和批量下载文章子集以及PMC存档中的所有科研数据, 实现NIST科研数据的永久保存和长期可访问性服务。

(4) 标准参考数据服务。标准参考数据服务是由数据与信息学办公室开发的国家标准参考数据系统的数据管理项目, 以支持NIST范围的标准参考数据开发、管理、维护和传播, 确保科学家、工程师和公众可以轻松获取可靠的参考数据。1968年美国国会通过的《标准参考数据法》将标准参考数据(SRD) 定义为可靠的、经过评价的数值数据。目前, NIST共开发了116个SRD数据库产品, 其中41种是收费SRD产品, 75种免费SRD可通过Web界面获得。SRD按编号列出所有目录, 并按主题区域分组, 以方便用户浏览和定位。不同的SRD有不同的使用人群和不同的营销需求, 营销方案分为3个主题: ①利用在线营销技术跟踪客户足迹, 并建立在线SRD社区; ②开发成功案例, 采用线上、线下相结合的方式, 向公众推广SRD产品; ③编辑出版有关标准参考数据的出版物。根据标准参考数据计划, 2015年底数据与信息学办公室启动评价SRD项目的工作, 成立项目审查委员会, 以评价SRD项目的现代化进程, 包括网络界面再设计、应用程序编程接口和技术内容。SRD每个数据产品的形成都凸显了数据收集、数据评估、数据传播和扩散三个阶段过程, 集合了数据专家和实验室高级研究员的集体智慧。作为NIST的品牌性数据产品, SRD在科研数据管理服务领域获得较高声誉^[14]。

(5) 数据管理培训。NIST的科研数据管理培训主要包括两部分, 一是针对数据馆员的教育培训, 二是针对用户的数据素养教育。对数据馆员的培训主要有数据课程、交流研讨会、科研项目合作等方式。NIST与华盛顿大学签订合作协议, 定期选派优秀数据馆员到华

盛顿大学接受短期的数据课程培训, 学习内容包括数据工程和管理、机器学习和应用数据科学等。数据馆员也主动加入专业工作组与实验室研究人员进行合作, 为科研项目提供嵌入式数据管理服务。对用户的数据素养教育主要以信息服务办公室与信息技术实验室合作开设的“数据科学”课程形式。课程内容包括制定数据管理计划、文件格式与转换、数据产权与共享, 以及最终把数据解决方案应用到具体科研项目和学科领域的方法, 以实际案例让用户了解数据处理生态系统。课程采用传统面授和在线学习相结合的教学形式, 用户可根据自身的实际情况, 任意选择学习方式, 极大地增强了课程的可获取性。

3 NIST科研数据管理实践成功经验分析及启示

通过分析NIST科研数据管理实践内容, 笔者认为NIST所积累的成功经验值得国内机构深思。

3.1 完善的政策规范是开展科研数据管理实践的重要依据

调研发现, NIST非常重视科研数据管理法律、法规和规章制度的引进和建设, 从引进美国联邦政府相关法律、法规到机构内制定各种政策规范, 发布了一系列的数据管理政策。如2015年6月25日发布的《管理公众对联邦资助项目研究成果的访问》^[15]政策, 2017年11月13日制定的《公开科研数据的保存和维护》^[16]政策, 2017年11月29日制定的《资助科学研究提交数据管理计划审查指南》^[17], 这些政策从数据管理平台、科研人员、科研数据、服务内容方面来规范和指导科研数据管理行为, 值得我们学习和借鉴。我国的科研机构、出版机构要借鉴NIST优秀经验, 制定出理论与实践相结合的数据管理政策, 为推动我国科研数据开放共享提供指导方针。

3.2 完整的运行机制是开展科研数据实践的重要保障

科研数据管理是一项系统工程, 必须构建一个切实可行的运行机制, 保障科研数据管理实践的顺利开展。NIST构建的科研数据管理运行机制, 涉及整个

科研数据生命周期,每个阶段都有特色的管理策略。

①管理平台不仅充分利用机构内已有设施,还与美国国立卫生研究院合作,使用Pub Central存储生物医学数据;②机构成员来源广泛,职责分工明确,保障了业务的正常开展;③资源工具通过自主开发和联合开发为科研人员提供丰富的数据管理资源,供其参考使用;④政策法规是数据管理服务全过程的指导方针;⑤数据管理培训涵盖科研用户和数据馆员,教育形式主要有在线课程、讲座、专题研讨会等;⑥服务内容是科研用户需求的数据管理计划、数据存储与共享,以及标准参考数据服务等方面。总之,NIST构建的全面系统运行机制是成功开展科研数据管理实践的重要保障。我国科研机构要借鉴NIST成熟的运行机制,结合自身实际,构建特色的数据管理实践的运行框架;在具体的服务实践过程中,要积极采纳用户提出的新需求或新问题并不断进行改进,保证科研数据实践的持久性。

3.3 全方位合作是开展科研数据管理实践的重要途径

NIST在科研数据管理实践过程中开展了全方位、多层次、多元化的合作方式。一是机构内部合作。NIST的图书馆、信息技术中心和实验室在数据管理平台设计、软件工具开发、基础设施建设、教育培训等方面进行了广泛合作。各部门发挥自身优势,增强了机构内数据管理水平及对科研的支撑能力。二是机构间的合作。NIST积极参与外部数据服务,如国家数据服务和科研数据联盟等,与70多个联盟成员和工作组合作探索科研数据共享的发展,数据库的互操作机制,以及其他部门和机构的数据存储、可发现和数据重用,努力为遍布全球的NIST社区提供示范资源,推动NIST使命。三是与高新技术公司的合作。通过开展竞赛的形式,对全国各地开发商提交的25个APP与6个常用SRD子集的应用程序进行评审,评审团成员包括NIST数据和信息学办公室主任、商务部首席数据官、谷歌副总裁兼首席互联网传播者等7位高新技术专家。由以上分析可知,寻求多方合作是NIST推进科研数据管理进程的重要途径。我国在开展科研数据管理服务时要借鉴NIST的合作精神,在人力、物力、资源方面联合多个部门,组建专业服务团队,共同开发技术平台,丰富服务内容,拓展科研数据管理实践途径。

综上,制定完善的数据管理政策,构建全面系统的服务体系,寻求多方位的合作方式,重视培养高素质的数据管理人才,是NIST科研数据管理实践成功的关键因素。目前我国科研数据管理还处于探索阶段,尚未形成良好的实践成果和合作机制,NIST科研数据管理的成功做法与实践经验对我国开展科研数据管理具有很好的借鉴意义。

参考文献

- [1] NIST About NIST [EB/OL]. [2018-08-20]. <https://www.nist.gov/about-nist>.
- [2] NSF. ENG Guidance on Data Management Plans [EB/OL]. [2018-09-01]. <http://www.nsf.gov/eng/general/dmp.jsp>.
- [3] H.R.3547-Consolidated Appropriations Act, 2014 [EB/OL]. [2018-10-18]. <https://www.Congress.gov/bill/113th-congress/house-bill/3547>.
- [4] MEDINA-SMITH A TRYKAK K M, SILCOX B P, et al. Librarians and scientists partner to address data management taking collaboration to the next level [J]. Digital Library Perspectives, 2016, 32 (3): 142-152.
- [5] NIST Data Management Platform [EB/OL]. [2018-09-21]. <https://www.nist.gov/data>.
- [6] Atomic Spectroscopy Database [EB/OL]. [2018-09-21]. <https://www.nist.gov/pml/atomic-spectroscopy-databases>.
- [7] KAISER D L, HANISCH R J. Technical Review of NIST's Free Standard Reference Data Products [EB/OL]. [2018-04-22]. <http://doi.org/10.6028/NIST.SP.1223>.
- [8] MetroStar Systems' ChemBook App Wins Third Place in National Institute of Standards and Technology Reference Data Challenge [EB/OL]. (2015-12-01) [2018-10-18]. <http://www.prnewswire.com/news-releases>.
- [9] Lab Liaison Program [EB/OL]. [2018-10-18]. <https://www.nist.gov/programs-projects/lab-liaison-program>.
- [10] Office of Science and Technology Policy. open data Policy- Managing Information as an Asset [EB/OL]. [2018-10-21]. <https://project-open-dntn.cio.gov/pdicy-memo/>.
- [11] The NIST Plan for Providing Public Access to Results of Federally Funded Research [EB/OL]. (2015-09-20) [2018-11-21]. <http://dx.doi.org/10.6028/NIST.IR.8084>.
- [12] NIST Data Management Tool [EB/OL]. [2018-11-21]. <https://midas.nist.gov/>.

- [13] Guidance for Making Non-NIST Data Available Through NIST's Data Management Infrastructure [EB/OL]. [2018-04-25]. https://www.nist.gov/sites/default/files/documents/2018/06/19/final_g_5702_01.pdf.
- [14] KAISER D, ALDEROTY N, CAVANAGH R, et al. Standard Reference Data Workshop Report [R/OL]. [2018-06-28]. <http://doi.org/10.6028/NIST.SP.1225>.
- [15] Managing Public Access to Results of Federally Funded Research Policy [EB/OL]. (2018-06-19) [2018-06-28]. https://www.nist.gov/sites/default/files/documents/2018/06/19/final_p_5700.pdf.
- [16] Preservation and Maintenance of Published Research Data [EB/OL]. (2018-06-19) [2018-06-28]. https://www.nist.gov/sites/default/files/documents/2018/06/19/final_o_5702_1.pdf.
- [17] Guidance for Review of Data Management Plans Submitted with Applications and Proposals for NIST Funding for Scientific Research [EB/OL]. (2018-06-19) [2018-06-26]. https://www.nist.gov/sites/default/files/documents/2018/06/19/final_g_5701.pdf.

作者简介

陈晨，女，1986年生，硕士，助理馆员，研究方向：信息服务、数字资源管理，E-mail: chenchen022@126.com。

Practice and Enlightenment of Research Data Management in American National Institute of Standards and Technology

CHEN Chen

(Library of Tianjin Foreign Studies University, Tianjin 300270, China)

Abstract: The National Institute of Standards and Technology research data management service is currently the most successful example of data management practice in the United States. This paper uses the the network survey method to conduct in-depth analysis on the background, service system of the NIST data management and summarizes the successful experience of NIST research data management, in order to provide inspiration and reference for the development of research data management services in China.

Keywords: Research Data; Data Management; Data Service; NIST

(收稿日期: 2019-03-14)