



开放科学
(资源服务)
标识码
(OSID)

中国科学数据基础设施建设及发展对策研究

屈宝强

中国科学技术信息研究所 北京 100038

摘要: 科学数据基础设施建设是管理和保存科学数据资源、促进科学数据高效利用的重要条件。本文梳理了我国政府、科研机构层面等科学数据基础设施的建设情况,分析了当前科学数据基础设施建设的主要特征及差距,提出在落实《科学数据管理办法》要求下,我国科学数据基础设施建设的基本要求和发展建议。

关键词: 科学数据; 数据共享; 基础设施; 数据中心

中图分类号: G35

Construction and Development Strategy of Scientific Data Infrastructure in China

QU Baoqiang

Institute of Scientific & Technical Information of China, Beijing 100038, China

Abstract: The construction of scientific data infrastructure is an important condition for the management and preservation and efficient utilization of scientific data. This article analyzes the current situation of scientific data infrastructure construction which are led by the Chinese government, scientific research institutions, et al., summarizes the characteristics of scientific Data Infrastructure, and puts forward corresponding suggestions to improve the ability of data infrastructure construction, complied with the requirements of “Measures for scientific data management of China”.

Keywords: Scientific data; data sharing; infrastructure; data center

基金项目: 国家社会科学基金项目 (15BTQ053)。

作者简介: 屈宝强 (1980-), 博士, 研究员, 研究方向: 科学数据共享, E-mail: qubq@istic.ac.cn。

引言

科学数据基础设施 (data infrastructure), 是用于存储和管理、并帮助促进发现和高效使用科学数据的信息技术环境。科学数据基础设施是数字时代科研活动的重要基础, 是保障科学数据有效管理, 促进科学数据高速流通, 提升科学数据使用效益的重要支撑。本文主要梳理我国在科学数据基础设施建设方面的进展, 分析存在的主要问题, 提出未来发展的相关建议。

大数据时代, 世界各国都非常重视科学数据基础设施建设。在美国, 大多科学数据基础设施由国家科学基金会 (NSF)、博物馆及图书馆服务研究所 (IMLS) 和梅隆基金会组织资助, 希望各类数据仓储和图书馆建立协作网络^[1]。美国 NSF 发起的 DataNet 项目, 规划创建面向科研信息化的科学数据保存与共享模式。在此项目支持下, 建立一批新型数据保存与共享组织架构, 为长期的科学研究提供可靠的数据保存、访问、集成和分析能力^[2]。由英国多家机构联合开展的 DISC-UK 数据共享项目, 其总体目标是促进科学数据共享新模式、工作流程和工具的形成。英国研究理事会资助一批数据中心为不断增加的数据提供专业的数据服务, 包括 MIMAS、EDINA、AHDS、UK Data Archive 等^[3]。加拿大有 “Data Canada”, 致力于建设国家基础设施, 担负起发展和执行科学数据开放获取战略规划的任务^[4]。澳大利亚国家科学数据服务网络 (ANDS), 旨在全面整合全国科学数据资源, 实现长期保存和共享利用, 其重点领域包括: 数据采集基础设施、元数据存储基础设施、

自动数据发布公共基础设施和软件基础设施^[5]。此外, 欧洲数据基础设施 (EUDAT) 努力创建全欧洲电子数据共享的配套基础设施, 欧盟 DRIVER 项目主要目标是为数字存储库创建灵活的泛欧洲基础设施, EuroVO-AIDA 项目整合了天文学的数字数据收集, IMPACT 项目整合了来自 10 个主要蛋白质家族数据库的数据^[6]。日本 NII 启动建设 NII Research Data Cloud 研究数据云基础设施建设, 它由研究数据管理平台 (GakuNin RDM)、数据存储库平台 (WEKO3) 和数据发现平台 (CiNii research) 组成, 用于帮助开展科学数据全生命周期管理。

1 我国科学数据基础设施建设情况

机构存储库、领域和国家级别的数据中心以及所有其他存储和共享数据的方法都是使研究人员访问和利用数据的基础设施。

1.1 科学数据共享平台建设

科学数据共享平台建设是我国科学数据基础设施建设的核心内容和抓手。我国 20 世纪 80 年代末参加了世界数据中心 (WDC) 项目。从 2001 年开始, 在充分调研、吸收国外数据共享经验的基础上, 选择了九个单位开展科学数据共享试点, 2005 年开始试点单位进入国家科技条件平台进行建设, 2011 年科技部与财政部联合对建设单位进行全面评估, 选出了六个试点建设单位 (国家气象科学数据共享平台、农业科学数据共享平台、地震科学数据共享平台、林业科学数据共享平台、人口与健康科学数据共享平台、地球系统科学数据共享平

台)转入运行服务阶段,科学数据共享平台也陆续由项目支持转为运行服务支持。目前初步构建了由主体数据库、科学数据中心或科学数据网、门户网站所构成的三级结构数据管理与共享服务体系,形成了一批资源优势明显的科学数据中心,实现了一批数据的汇交整合与开放共享。

在国家科学数据中心建设方面,我国为落实《科学数据管理办法》和《国家科技资源共享服务平台管理办法》的要求,规范管理国家科技资源共享服务平台,完善科技资源共享服务体系,推动科技资源向社会开放共享,通过部门推荐和专家咨询,经研究共形成“国家高能物理科学数据中心”等20个国家科学数据中心、“国家重要野生植物种质资源库”等30个国家生物种质与实验材料资源库^[7]。在领域数据共享平台和数据库建设方面,截止目前,已形成地球系统科学、人口与健康科学、气象科学、地震科学、海洋科学、农业科学、林业科学等共享服务平台,建设有大气、陆地表层、遥感反演产品、人类尿蛋白质组、气象数据和产品、极地、固体地球、陆地水圈、高能物理、空间科学、天文科学等主题数据库。在区域科学数据共享方面,各省、市、自治区都建设了科学数据共享平台,服务本地科研活动。在利用共享平台开展科学数据统一揭示和发布方面,2009年9月中国科技资源共享网正式开通,运用信息、网络等现代技术,从国家层面将科学数据共享平台建设成果开放,把科学数据以及大型科学仪器设备和研究实验基地、自然资源、科技文献等相关科技资源统一发布和开放共享,实现了科学数据资源目

录级开放共享。

1.2 机构数据仓储建设

从事研发活动的机构开始探讨建设机构数据仓储存储和管理产生的研究数据,建立提供对研究数据集的访问,并提供不同研究数据服务。许多大学图书馆担当了该机构科学数据基础设施建设的职责,例如康乃尔大学图书馆开发了数据仓储 DataStar 系统支持在研究过程中的协作和数据共享,普度大学图书馆建立“分布式数据管理中心”(D2C2),满足校园研究数据管理需求,明尼苏达大学图书馆建立了 e-Science 和数据服务合作计划,实施校园研究基础设施联盟,探索数据服务模式^[8]。许多机构知识库开始向机构数据仓储升级,例如 RADAR 是一个通用的保存和发布数据的数据仓储库,确保服务可以在学术机构和高校实现,可持续管理和发布的研究数据与数字对象标识符(DOI)的分配是由德国国家图书馆(TIB)提供^[9]。

机构层面的科学数据仓储是我国科学数据基础设施建设的重要内容。我国的科研机构也不断探索将科学数据管理内容融如机构知识库和图书馆发展当中,例如武汉大学图书馆基于 DSpace 建立科学数据管理平台,复旦大学基于 Dataverse 开展社会科学数据的共享与交换服务,北大开放研究数据平台为研究人员提供研究数据的管理、发布、存储和使用追踪服务,促进研究数据的传播、重用和规范引用。复旦大学图书馆等9家单位于2014年10月共同成立了中国高校研究数据管理推进工作组,主要目标在于宣传推动研究数据的科学管理,推进研究

数据的创建、收集、存储、发布、应用、重用、出版及共享,为高校图书馆的研究数据管理建设提供最佳实践案例等^[10]。

数据管理和共享需要长期支持,这是单个机构和项目无法单独提供的。领域数据仓储则面向学科/领域的共同标准、政策和实践,促进领域科学数据基础设施发展,它一般由研究资助者规划、倡导和投资支持。例如生物医学领域 Dryad、政治和社会科学领域 ICPSR、地球与环境科学 PANGAEA、心理健康领域 Samhda、DNA 序列领域 Genbank 等,另外还有一些可以用于通用领域的数据仓储,例如 figshare。我国也开始这方面的探索,由中国科学院微生物研究所(国家微生物科学数据中心)发起,包括我国微生物领域学会、期刊、相关科研机构和主要测序机构的微生物科学数据共享联盟 2019 年 6 月成立,但此类互联的数据共享体系还较少。

1.3 新型数据出版体系建设

科学数据出版体系是我国在新时代科学数据基础设施建设的新探索。新型科学数据出版体系既是促进科学数据权益保护的重要举措,同时也是不断完善科学数据基础设施支撑高水平研究的相关机制。高质量的学术论文需要高水平的科学数据支撑,我国不断加强以利用各类数据仓储和共享平台发布数据的同时,当前积极探索将数据与数据论文关联出版的新模式,积极开展数据论文出版,数据论文是描述特定数据集的可搜索元数据学术出版物文档^[11],目前已经有一些期刊开始出版数据论文,并与相关数据中心的原始数据关联出版,例如《中国

科学数据》、《地理学报增刊》、《图书馆杂志》等,出现了一批针对科学数据出版模式、标准、技术的研究成果^[12-15]及领域科学数据出版问题的探讨^[16-17]。

1.4 科学数据政策与标准体系

科学数据政策与标准体系是我国科学数据建设的重要保障。在数据政策层面,2004 年,中国签署了《OEOD 获取公共资助科学数据宣言》。2015 年国务院印发《促进大数据发展行动纲要》,指出大力发展科学大数据,积极推动有国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享,实现对国家重要科技数据的权威汇聚、长期保存、集成管理和全面共享。2016 年中共中央、国务院印发《国家创新驱动发展战略》,指出要适应大科学时代创新活动特点,加快建设包括数据资源在内容科技基础条件平台。2018 年 3 月,国务院颁布《科学数据管理办法》,要求按照“开放为常态、不开放为例外”的共享理念,推动科学数据管理。

在标准体系建设层面,目前已经通过全国科技平台标准化技术委员会制定发布一批国家标准,重点对科学数据资源信息的统一发布和互联互通进行规范。各领域/行业已逐步形成一批科学数据标准规范,包括国家标准、行业标准以及各类技术规范,涉及科学数据采集生产、加工整理、共享服务全流程。2018 年《科学数据引用》发布,规定了科学数据引用元素描述方法、引用元素详细说明、引用格式等方面的内容。这些相关的政策内容和标准体系对于科学数据资源的管理、利用与服务等起到了

重要的作用。

2 我国科学数据基础设施问题剖析

从新世纪以来,我国充分发挥集中力量办大事的制度优势,初步形成了数据基础设施体系,同时倡导科学数据开放共享,对科技创新起到了重要支撑作用。但是与我国科技创新发展的要求相比,还存在一些不足。

(1) 数据基础设施的投入结构有待优化

Gabella C^[18]总结了科学数据资源和设施的12种资助模式,根据收入来源,这12个模型可分为三大类。大多数科学数据基础设施建设依赖国家预算的资金,并根据各种规则和条件在申请人、机构或服务机构之间重新分配。第二类包括依赖于面向各类用户收费的模式,还有依靠自愿捐赠和参与或第三方资助的模式。在我国,科学数据中心和数据资源以及基础设施建设主要以公共财政投入为主,各类数据基础设施的投入模式仍需优化,企业和社会资金投入引导性政策不足,制约了更大程度创新数据服务模式,提升科学数据服务经济社会发展的能力和水平。

(2) 各级基础设施的互联互通整体水平不高

数据密集型科研范式下,越来越多的学术期刊开始尝试与科学数据进行关联,为用户提供集文献和数据于一体的信息服务,实现科学数据的增值再利用。科学数据基础设施与科学数据管理以及其他科研管理体系的互联是其发挥重要作用的基础。英国FISHnet、I2S2、IDMB、MaDAM等研究数据管理基础设施建

设项目同研究数据管理规划项目(如DMBI、DMP-ESRC)、研究数据引用关联和集成项目(如Dryad UK、SageCite),以及研究数据管理培训项目(如CAiRO、DataTrain)等形成良好的合作关系^[19]。

我国目前建设的科学数据基础设施,一方面国家、地方、机构层级数据基础设施之间互联互通不足,同领域数据基础设施之间连通不足,管理的标准体系不统一,资源建设分散。另一方面,科技数据基础设施与相关的科技论文、科研项目、科技成果等其他基础设施和系统关联不足,这从客观上导致在学术交流过程中对科研活动的整体性认识不足,科技数据支撑创新的能力没有完全凸显。

(3) 科学数据基础设施管理和运行水平仍存在差距

经过多年建设,欧美发达国家已支持形成了一批高水平科学数据中心,如美国国立生物信息技术中心建立和维护的GenBank,目前已是世界上最权威的基因序列登记数据库之一。而我国领域/行业在科学数据生产、管理、共享、利用方面的工具体系、标准体系、技术体系尚未形成,另一方面,科学数据基础设施国际性、权威性不足,支撑新时期“三跑并存”科技创新局面的整体效能不高。世界知名数据仓储目录re3data.org目前共索引全球2479个数据中心/平台,其中,中国自主建设(35)和参与建设(8)的数据中心只有43个。科睿唯安DCI数据库索引的中国数据仓储仅有兰州寒区旱区科学数据中心、蛋白质乙酰化数据库、全球变化数据出版平台、北京大学开放研究数据平台、GigaDB等数十个,数量和比例相对较低。另外,

虽然已经开展包括科学数据的科技基础资源调查,但是尚没有从国家层面开展科学数据仓储和资源的注册与认证体系。另外,从促进科学数据基础设施建设的标准来讲,国际性、引领性的标准还严重匮乏。科学数据中心建设标准、认证标准等尚未形成,而国际上已有数据资产框架(Data Asset Framework, DAF)、Data Seal of Approval 数据仓储认证、Digital Repository Audit Methodology Based on Risk Assessment (DRAMBORA) 数据仓储能力评估等可用于数据仓储质量评价的相关方法^[20]。

(4) 面向共享和利用的数据服务生态尚未形成

虽然我国科学数据基础设施建设取得了重要进展,但是融入科研过程的数据服务体系尚未完全建立,“重建设、轻应用”的现象仍然存在。一方面,面向科研人员便捷使用科学数据的资源发现平台服务能力和水平有待提升,现有科学数据共享平台在促进共享的服务意识、权益保护意识还有待加强。另一方面,在促进科学数据开放共享方面的权益保护和利益实现机制尚未达成共识,科学数据工作尚未完全得到学术共同体认可。

3 我国科学数据基础设施建设对策

未来我国科学数据基础设施建设要以大数据、云计算和人工智能等信息技术发展为基础,通过科学数据资源的集聚、融合和利用,形成面向科研人员可用、易用、好用的科学数据资源服务体系。其总体要求是适应全方位开放新格局和科研国际化需求,与国际重要科学数据

基础设施做好技术、标准、工具体系的对接和互联,充分集聚和融合全球科学数据资源,支撑颠覆性技术研发和“无人区”领域的探索。同时服务国民经济主战场,通过科学数据基础设施建设推进科技创新资源的优化配置,促进科技贡献率提升,通过新技术、新手段的探索新的服务模式。根据上述分析,结合我国《科学数据管理办法》的要求,今后一段时期,我国加强科学数据基础设施建设可以从以下几方面开展。

3.1 做好统筹规划,协同全国科学数据基础设施一体化发展

在已有工作基础上,统筹规划国家数据基础设施工作,做好国家科学数据中心的战略保障功能,引领示范功能。充分利用已有科技文献管理和服务的建设和经验,实现科学数据管理从不规范到体系化管理、从无关联到关联数据、从数据不能被研究者发现到可以很容易被发现、从数据单一使用到可重复使用。在国家科学数据基础设施整体推进过程中,既要强调对科学数据基础设施宏观上总体领域、区域布局的合理性,同时要注重每个数据仓储在科学数据获取、存储、质量控制、安全等方面的科学性和系统性。在已有科技资源调查基础上,梳理形成国家科学数据目录体系,并根据科研活动需求不断充实完善,形成科学数据检索系统建设和开发的重要基础,同时根据调查结果,组织建设全国科学数据仓储目录注册系统,帮助及时掌握各领域数据仓储和数据中心建设情况、运行服务和开放共享情况。

表 1 我国科学数据设施建设的内容与要求

序号	内容	科学数据管理办法要求	现状	建议和实施路径
1	组织管理体系	国务院科技行政部门；国务院相关部门、省级人民政府相关部门；科研院所、高校、企业为责任主体的三级管理体系	职责不清、权限不明晰	形成位阶不同、相互衔接的数据政策体系，强调支撑科学数据基础设施建设的标准体系建设
		法人单位应建立科学数据保存制度，配备数据存储、管理、服务和安全等必要设施	部分法人单位没有相关基础设施	加强整合与建设，加强数据互联互通，完善和优化科学数据全流程管理过程
2	建设国家科学数据中心	加强统筹布局，优化整合形成国家科学数据中心	已经认定第一批20个国家数据中心，30个数据库	数据中心认证，国际接轨
3	科学数据汇聚	科学数据汇交作为项目验收的必备要求	科技计划项目产生的数据汇交尚无具体细则	强化数据管理规划，明确汇交的时间、方式和监督
		依托行业部门管理体系进行长期积累和保存	行业部门数据积累较为成熟	—
		国内外学术论文数据汇交备份	学术论文数据汇交备份的基础设施和管理机制尚未建立	研究机构做好论文发表的数据汇交审查和接收工作
4	数据公开与开放	社会资金资助形成的涉及国家秘密、国家安全和社会公共利益的科学数据必须按照有关规定予以汇交	尚无相关实施细则	与国家财政科技计划项目产生的科学数据不断融合汇聚
		编制科学数据资源目录	尚无统一的科学数据资源目录	在科技资源调查基础上形成科学数据目录体系，并不断更新完善
		依托国家科学数据网络管理平台统一发布科学数据资源信息发布。	尚无国家科学数据网络管理平台统一，科学数据平台之间的互联互通不足	加强资源统一发现门户建设，制定元数据互操作标准，加大不同层级科学数据中心间的互联互通
		在线下载、离线共享或定制服务等方式向社会开放共享	开放共享不足	依托各类数据中心，加大开放力度，提升数据服务效果
		开展科学数据分析挖掘基础上的增值服务	部分领域开展了科学数据增值服务，数据产权问题不明确	明细数据产权，将科学数据及在其基础上研究成果推广至其他领域
		发展科学数据出版。发表产权清晰、准确完整、共享价值高的科学数据	部分数据中心、数据共享平台探索与学术出版关联的数据出版	认定和建设一批国际权威的数据期刊，扩大数据论文出版力度
		强调数据引用，在论文发表、专利申请、专著出版等工作中注明所使用和参考引用的科学数据	科学数据引用标准发布，但是远未引起重视，实施效果不佳	宣传贯彻数据引用标准，制定在论文及其他场景著录的细则
5	科学数据共享与利用	服务政府决策、公共安全、国防建设、环境保护、防灾减灾、公益性科学研究等	服务效果不突出	以重大工程和项目为突破口，提升服务水平
		服务经营性活动，签订有偿服务合同，明确双方的权利和义务	权利义务不明晰	将科学数据看作重要科研产出，按科技成果转化要求实施
		科学数据科普服务	科普服务几乎没有	开展宣传推广，特别是国家重点实验室等

(续表)

序号	内容	科学数据管理办法要求	现状	建议和实施路径
6	数据安全	严格制定国家有关保密规定	国家有关保密规定需要修订	将科技数据安全纳入整体科技安全
		做好保密与安全管理	保密与安全管理意识不强	做好安全防范
		建立数据共享和对外交流的安全审查机制	数据定密和审查缺失,大量数据流失	加强科技数据安全保密认定
		加强科学数据全生命周期安全管理	科学数据全生命周期管理标准化和规范化不足	研发各类科学数据管理工具体系,运用到全生命周期各个流程,强化科学数据质量
		建立网络安全保障体系	现有网络安全保障体系较为成熟	重视网络数据传输特别是跨境流动安全
7	人才队伍建设和激励体系	做好应急管理 with 容灾备份	尚无国家科学数据应急管理 with 容灾备份机制	启动国家科学数据保障体系建设
		在岗位设置、绩效收入、职称评定等方面建立激励机制	尚无相关政策,实施和操作难度大	以“三评”改革、去“四唯”为抓手,将科学数据工作纳入科研评价体系

以当前科学数据管理的实践为基础,研发一批领域通用的科学数据管理工具、平台和技术方法,强化科学数据全生命周期管理。研制中文科学数据唯一标识技术体系,包括标识体系、解析机制、与其他标识的关联机制,形成科学数据记录、数据集、数据库、数据产品之间的关联系统和溯源体系,并与科技论文的出版过程紧密耦合,落实好科学数据管理办法中关于发表论文必须提交科学数据到相应存储机构的要求。

构建一体化科学数据基础设施管理与服务机制,一方面,通过元数据互联互通,构建全国一体化科学数据服务网络,实现从单个数据存储机构(数据中心、机构知识库、共享平台等)服务到协同服务网络过渡。另一方面实现从其他科技信息资源管理系统的工作流中抽取数据,将科学数据与研究过程中的科研人员、科研机构、科技项目等实体、产出、活动关联。

3.2 强化科学数据统一揭示和发现,提升科学数据基础设施服务效能

日本 NII 建设了面向科研领域服务的云研究数据基础设施 (Research Data Cloud), 由研究数据仓储、研究数据管理、研究数据发现三部分组成,帮助科研人员发现科学数据管理和共享。澳大利亚 ANDS 建立研究数据注册与发现门户 (Research Data Australia), 帮助数据发布、引用和发现^[21]。我国应在科技资源共享网建设基础上,以全国科学数据目录体系为基准,汇聚全国主要科学数据中心和数据库资源元数据,进行统一资源揭示,实现从目录级开放逐步走向元数据和数据内容的开放,让科研人员实现一站式全国科学数据发现、获取和利用。

同时建立科学数据与科技人才资源、大型仪器设施、以及其他材料试剂类科技资源的数据化统一描述,将各类元数据资源进行统一逻辑抽取和整体融合,形成统一的科技创新资源

展示、检索、查询、选择、推荐、利用的门户。实现科技创新资源服务与科技计划管理、科研过程管理等的深度融合，支撑整体科研活动与科技创新全链条，实现各类科技创新数据资源在创新活动中的协同服务。

在支撑统一发现的技术体系方面，建议充分借鉴目前主要开源数据仓储系统 DSpace、Fedora 和 iRODS 的成熟经验，从技术架构上考虑互联互通要求，研发具有自主知识产权、用于科学数据仓储运行管理的信息系统。完善科学数据领域的元数据标准体系，力争形成助力资源统一发现的科学数据核心元数据集，形成通用元+领域+特征三层次的元数据结构框架。在此基础上建立多源异构科学数据统一揭示和发现的技术框架和标准体系，从硬件和操作系统层面解决不同数据仓储的系统异质性、从元数据模型和框架协议方面解决各类科学数据描述语法、结构、和语义方面的一致性，推进科学数据元数据、本体和词表等相关内容的研究，通过模式映射、语义注释、唯一识别等实现适用不同场景数据融合的方法。

在科学数据发现门户建设方面，建设方便用户查找、访问使用数据集的统一搜索引擎，优化领域数据仓储资源发现门户，实现全国科学数据仓储的关联和协作。通过建立领域、区域以及国家层面集中的科学数据门户，将数据的发现、获取、集成、安全和管理等统一实施，并做好分布式数据联合存储基础设施，使得各个数据仓储之间建立相互联系的合作关系。强化对科学数据、科技论文、科技报告等多类型数据资源集成发现系统的研发部署，并同科研信息系统互联互通，形成科技资源、研发活动、

研究产出三位一体的科研信息发现网络，从而形成全国性的数据基础设施。

在高质量科学数据发现方面，建立科学数据索引数据库，通过数据仓储、数据共享平台、数据机构已有科学数据使用情况的度量作为数据发现的基础，可帮助科学家更好地进行数据评价、发现、获取、利用/重用数据。

3.3 强调科学数据基础设施建设中技术与政策、文化等的协同推进

科学数据基础设施不仅是用于存储、共享和利用数据的基础条件，更需要成为结合科技人文、科研诚信、科技评价建设为一体的组合，全面解决科学数据管理和利用政策、法律、文化和权益（支持数据引用）问题。

在政策体系建设方面，在落实《科学数据管理办法》的政策要求下，鼓励各级科技管理部门、期刊出版部门、研究机构等完善相关数据政策，形成位阶不同、相互衔接的数据政策体系。另外，一方面注重现有国际性科学数据标准的本地化，同时强调科学数据基础设施建设相关的通用技术标准、数据互操作标准、数据使用和管理标准等导性、前瞻性标准的研制。

在科学数据权益保护方面，突出对科学数据工作的认可。以《科学数据引用标准》推广为突破口，推动数据合理有效利用，全面强制性实施科学数据引用著录，并在科技期刊论文发表中试行科研活动全要素著录的探索。

在促进科学数据评价方面，按照《关于深化项目评审、人才评价、机构评估改革的意见》《关于破除科技评价中“唯论文”不良导向的

若干措施(试行)》的要求,针对科技活动评价中“唯论文”不良导向,敢于“破”、破得坚决,更要善于“立”、立好新规。立则可以建立对科学数据生产、制作、管理、服务人员相关工作成果和绩效纳入科研评价体系。各类科研单位可结合本单位实际,完善制度、细化措施、加强管理,在职称评审、职务晋升中体现对科学数据工作的认可。

在强化科学数据管理意识方面,鼓励各级各类数据中心、仓储编制相关科学数据基础设施使用指南,帮助研究人员和研究数据管理人员培养创建、管理和共享高质量研究数据的必要技能。结合中国科协等《关于组织实施中国科技期刊卓越行动计划》的工作要求,部署和建设一批适应国际发展趋势的示范性数据期刊。

4 结束语

我国是科学数据大国,对于科学数据共享的发展起步也比较早。但是近年来系统化推进科学数据资源建设、管理和利用的措施和手段有待提升。2018年《科学数据管理办法》出台,对未来一段时期科学数据发展提出了相关要求,但是后续实施细则尚未推出,各领域、各地方、各类机构在可操作性方面的方案还不完善,需要至上而下的统筹和自下而上的自律相结合,共同推进科学数据基础设施建设,为助力和支撑科技创新做好基础。

参考文献

[1] Anna Gold. Data Curation and Libraries: Short-

Term Developments, Long-Term Prospects. Data Curation and Libraries[EB/OL]. [2019-11-22]. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?referer=http://xueshu.baidu.com/usercenter/paper/show?paperid=746e1603383029cbb4ab3db52bc94783&site=xueshu_se&hitarticle=1&httpsredir=1&article=1027&context=lib_dean

[2] Kim J H. Data sharing and its implications for academic libraries[J]. New Library World, 2013, 114(11/12):494-506

[3] Lyon E. Dealing with data: Roles, rights, responsibilities and relationships [EB/OL]. [2019-12-23]. <http://core.kmi.open.ac.uk/download/pdf/1573430.pdf>

[4] Sabourin M, Dumouchel B. Canadian national consultation on access to scientific research data[J]. Data Science Journal, 2007(6): 26-35.

[5] Dong J L, Stvilia B. Practices of research data curation in institutional repositories: A qualitative view from repository staff[J]. PLoS One, 2017, 12(3): e0173987

[6] Sünje D T. Connecting Data Publication to the Research Workflow: A Preliminary Analysis[J]. The International Journal of Digital Curation,2017, 12(1):88-105

[7] 科技部, 财政部. 科技部财政部关于发布国家科技资源共享服务平台优化调整名单的通知 [EB/OL]. (2019-06-10)[2019-11-22]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2019/201906/t20190610_147031.htm

[8] Gold A. Data Curation and Libraries: Short-Term Developments, Long-Term Prospects[C]. AGU Fall Meeting Abstracts, 2009

[9] Kraft A. RADAR - A repository for long tail data[EB/OL]. [2019-12-23] <http://docs.lib.purdue.edu/iatul/2015/mrd/1/>

[10] 中国高校研究数据管理推进工作组. 中国高校研究数据管理推进工作组简介 [EB/OL]. [2019-11-22]. <http://society.library.sh.cn/node/2611>

[11] Chavan V, Penev L. The data paper: A mechanism to incentivize data publishing in biodiversity science[J].

- BMC Bioinformatics, 2011, 12 (Suppl 15):S2.
- [12] 马瀚青, 杨小梅, 侯春梅, 等. 数据论文联合出版模式及数据论文出版[J]. 中国科技期刊研究, 2018, 29(7):698-703
- [13] 涂志芳. 科学数据出版的基础问题综述与关键问题识别[J]. 图书馆, 2018(6):1-10.
- [14] 徐雷, 潘珺. 科学出版物语义数据及其应用研究[J]. 中国科技期刊研究, 2018, 29(7):704-710.
- [15] 吴立宗, 涂勇, 王亮绪, 等. 浅谈科学数据出版中的数字对象唯一标识符[J]. 中国科技资源导刊, 2010, 42(5):22-29
- [16] 高雅, 翁彦琴, 董文杰. ENCODE 项目科学数据出版模式创新[J]. 中国科技期刊研究, 2015, 26(8):808-812
- [17] 赵华, 王健. 科学数据出版现状及对中国农业科学数据出版的启示[J]. 农业展望, 2016(8):53-57.
- [18] Gabella C, Durinx C, Appel R. Funding knowledge bases: Towards a sustainable funding model for the UniProt use case[J]. F1000Research, 2018, 6(ELIXIR):2051
- [19] Andrew C. The 2014 DAF Survey at the University of Sheffield.International[J]. Journal of Digital Curation, 2015, 10(1):210-229
- [20] Alex B. Tools for Research Data Management (version 1.0)[EB/OL]. [2019-11-11]. REDm-MED Project Document. <http://opus.bath.ac.uk/29189>
- [21] Faniel, I M, Jacobsen T E. Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data[J]. Computer Supported Cooperative Work (CSCW), 2010, 19(3-4):355-375.