

加强和规范我国科学数据管理的思考

王瑞丹¹ 杨静² 高孟绪¹ 王超¹

(1. 国家科技基础条件平台中心, 北京 100862; 2. 甘肃省分析测试中心, 甘肃兰州 730000)

摘要: 科学数据是国家科技创新发展和经济社会发展的重要基础性战略资源, 特别是在当前大数据时代, 科技创新越来越依赖于科学数据的综合分析支撑。分析和梳理了国际科学数据管理和共享的发展趋势, 对照我国科学数据管理现状, 研究提出了加强和规范我国科学数据管理和促进资源开放共享的思考。

关键词: 科学数据; 科技管理; 科学数据管理; 开放共享

中图分类号: G311

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.02.001

Thoughts on Strengthening and Standardizing the Management of Scientific Data in China

WANG Ruidan¹, YANG Jing², GAO Mengxu¹, WANG Chao¹

(1. National Science and Technology Infrastructure Management Center, Beijing 100862; 2. Gansu Analysis and Research Centre, Lanzhou 730000)

Abstract: Scientific data is an important basic strategic resources for national scientific and technological innovation and economic and social development. Especially in the era of big data, scientific and technological innovation is dependent on comprehensive analysis of scientific data increasingly. Based on the analysis and sorting out the international development trend of scientific data management and sharing, and comparing the current situation of scientific data management in China, this paper puts forward the thinking of strengthening and standardizing China's scientific data management and promoting the opening and sharing of resources.

Keywords: scientific data, science and technology management, scientific data management, open and share

0 引言

科学数据通常是指在自然科学、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, 是能够客观反映事物本质、特征和变化规律的原始性、基础性数据^[1]。科学数

据作为最基本、最活跃的一类科技资源, 具有客观性、共享性、时效性、分散性、多结构性、再创造性、非排他性、传递性等特点^[1], 既是科技创新活动的重要产出和动力源泉, 也是经济社会发展的基石和工具。

近年来, 我国科技创新能力快速提升, 科技投入强度不断增加, 通过各级科技计划、科研基地建设、国际科技合作以及科技应用类公益事

作者简介: 王瑞丹 (1965—), 女, 国家科技基础条件平台中心研究员, 研究方向: 科技资源管理与科技评价 (通讯作者); 杨静 (1982—), 女, 甘肃省分析测试中心副研究员, 研究方向: 科技资源管理; 高孟绪 (1982—), 男, 国家科技基础条件平台中心副研究员, 研究方向: 科技资源管理; 王超 (1987—), 男, 国家科技基础条件平台中心博士, 研究方向: 科技资源管理。

收稿时间: 2018年3月16日。

业等促进了科学数据的快速积累和发展。随着大数据时代的到来,科技创新对科学数据的依赖性愈发增强,国家对科学数据管理和开放共享重视程度进一步加强,科学数据的发展在我国国家战略层面得到前所未有的高度重视。积极开展科学数据共享利用,对于改变我国科学数据封闭独享的局面,在我国形成科学数据共享氛围和服务意识,有着重要的引导和推动作用。

1 国际发展趋势

1.1 在国家层面对科学数据进行管理和部署

发达国家普遍重视科学数据的积累和重用,在法律、政策等层面提出科学数据管理的原则,在国家科学数据管理总体框架与行业领域科学数据管理等方面形成布局和多样化的运行机制,在开放共享等方面建立数据政策,均取得较为显著的效益和国际影响。美国政府早在20世纪就确立了在国家层面上建设国有科学数据和信息全社会共享环境的战略部署;欧盟发布的《欧洲研究领域开放数据获取政策和策略》将科学数据开放存取以政策的形式加以规范;英国自2005年起实施信息自由法。这些数据政策为各国形成强大的国家级数据基础设施以及提升科技创新能力提供了重要支撑。

1.2 科学数据管理贯穿数据全生命周期

目前已开展科学数据管理的国家和机构多以全生命周期为主要轨迹进行科学数据管理。许多国家及机构都针对科学数据生命周期各环节制定出台了相关政策制度与标准规范,涉及数据收集、数据归档、数据认证、数据加工、数据保存、数据发布、数据共享等环节^[2]。2013年白宫颁布了《开放数据政策——将信息作为资产进行管理》,强调了对数据的管理应贯穿数据的全生命周期,以促进开放性和互操作性;英国的科学研究资助机构制定了科学数据管理规定,内容涉及数据生命周期的各个阶段,包括:项目介绍与背景;数据类型、格式、标准与数据采集方法;数据使用道德与知识产权;数据检索、共享与重用;短期保存与数据管理;数据长期保存;资金

与人力支持;监督与评估^[2]。

1.3 将国家科学数据中心建设作为科学数据管理的重要手段

近年来,随着科学数据在重大前沿科学问题研究及战略发展方面的重要性日益突显,美国、欧盟等国家以及相关国际组织认识到科学数据是最重要的公共信息资源,也逐渐认识到构建国家级科学数据中心体系的必要性与重要性,发达国家和地区多年持续投入大量人力物力,支持建设了一大批科学数据中心,如美国国家空间科学数据中心(NSSDC)、英国数据保存中心(DCC)、英国数据档案中心(UK Data Archive)以及澳大利亚国家数据服务中心(ANDS)等^[3]。这些科学数据中心在相关学科领域持续汇聚和整合本国乃至全球科学数据资源,成为科学数据资源汇集的高地,面向全世界提供服务,为本国科技、经济、社会发展带来了显著效益。有关研究表明,美国通过对科学数据的高效使用,产生了巨大的社会效益,在实施科学数据共享政策的10年间,美国平均年经济增长率后5年比前5年增长了1.1%,其中0.5%是由于数据和信息的流通和应用所产生的^[4]。

1.4 多重手段促进科学数据的汇聚和利用

大数据的开发应用模式给传统科研活动带来新的启发,欧美发达国家除了利用科学数据中心和政府资助的科研项目等传统手段进行数据收集外,还利用学术期刊和网络平台推动科学数据汇聚和利用^[5]。在科研项目管理中,英国的科研资助机构要求申请人按照规定提交科学数据管理计划,将研究过程中产生的科学数据汇交至相应数据中心;美国国家科学基金会(NSF)要求基金项目申请者必须在提交的项目申请书中包含一份“数据管理计划”,对项目产生的有关科学数据进行管理和传播共享。在学术期刊方面,《Nature》杂志在线发布和共享了80多种期刊论文和相关数据服务,每月全球有数以百万计的科研人员对其进行浏览和访问。此外,基于网络平台也出现了“众包众筹”的新方法来加速科研进程,如分布式志愿计算项目Foldit项目通过互联网发起了

大规模的协同研究，以数据为纽带联合数千名科研人员共同参与研究，进行联机计算，使得该项目能够以前所未有的速度得到推进。

1.5 重视数据信息安全与网络安全

欧美国家非常注重科学数据信息安全与网络安全管理，因为科学数据开放共享在带来科技创新成果和经济价值的同时，也存在风险，如数据被破坏、用户个人信息泄露、侵犯个人隐私、数据滥用等。在发达国家，政府、科研资助机构以及大学、研究所等制定的与科学数据开放共享相关的政策中，一般都包含对涉及个人隐私、国家安全和机密等数据的例外条款，以保护有关数据被恰当发布出去。美国先后调整了国家信息安全政策，以巩固其在国际的领先地位，促使数据安全在国家信息安全政策中的地位不断上升；欧盟则于2007年3月22日正式通过了《关于建立欧洲信息安全社会战略的决议》，将区域的信息安全提升到社会形态的高度，要求在全社会实现网络和信息系统的规制，以保障信息网络系统的安全。

1.6 多手段保障科学数据生产者、服务者权益

科学数据和信息是一种宝贵的资源，是科学数据产出者的辛勤劳动成果。欧盟数据库保护指令、英国布加勒斯特宣言和《信息自由法》等，在科学数据的产权归属、共享管理和开发利用等方面均作了明确规定。欧盟通过资助“欧盟知识产权帮助”项目，保护其科技创新成果及其自主知识产权。2015年12月，欧盟执委会通过了《一般数据保护条例》，以欧盟法规的形式确定了对个人数据的保护原则和监管方式。在这个问题上，美国在保障国家安全、政府政务和个人隐私的前提下，采取谁投资谁收益的原则。

2 我国取得的成效

2.1 科学数据政策法规体系日趋完善

2018年1月，中央深改领导小组第二次会议审议通过了《科学数据管理办法》，成为我国首个在国家层面出台的科学数据管理办法。该办法针对目前我国科学数据管理各环节中存在的突

出问题，首次从国家层面明确了科学数据管理的职责、原则、方式和机制，以全链条管理思维对科学数据的统筹协调、生产加工、汇交保存、共享利用和安全保密等各个方面工作进行了全面部署。结合已经发布的《中华人民共和国测绘法》《地图管理条例》《基础测绘条例》《中华人民共和国测绘成果管理条例》《中华人民共和国气象法》《气象资料共享管理办法》《气象信息服务管理办法》《海洋资料汇交管理暂行办法》《海洋资料使用申请审批管理暂行办法》等多项部门政策制度，基本形成了国家统筹与行业联动的科学数据管理政策制度体系。

2.2 科学数据生产总量在多个领域呈现爆发式增长，信息化水平显著提高

近年来，随着我国重大科研基础设施和仪器设备的建设与更新、传感器网络的持续监测以及新数据处理技术和方法的推广与运用，我国科学数据资源数量呈爆发式增长。以高能物理、空间和天文、地球资源与环境、生命科学等领域的数据增长最为迅速。据不完全统计，截至2016年年底，我国有效管理与保存的科学数据资源达到64.17PB，数据记录达到43.8亿条。而随着我国科学各行业领域科学数据汇聚速度的不断加快和汇聚数量的不断增大，我国近90%的重大科技基础设施都已实现科研数据的自动化采集，70%以上实现了自动录入网络传输，科学数据管理信息化水平显著提高。

2.3 科学数据中心建设初具规模

多年来，我国在科学数据中心建设和发展方面进行了诸多探索和实践。2004年，科技部、财政部联合启动国家科技基础条件平台建设专项，先后在基础科学、农业、林业、海洋、气象、地震、地球系统科学、人口与健康8个领域支持建成了国家科技资源共享服务平台^[9]，基本覆盖了本领域科技资源优势单位，初步形成了一批资源优势明显的科学数据中心。国土资源、测绘部门和部分科研机构也根据部门及行业发展需要建成了一批科学数据中心，用以支撑数据资源的汇聚、管理与共享利用。此外，我国自1988年

被国际科学理事会(ICSU)下设的世界数据系统(WDS)接纳为正式会员后,就积极参与WDS建设。截止2016年年底,WDS共收录了19个国家和地区的70个普通数据中心(系统)。我国已经建立了9个数据中心。

2.4 科学数据管理逐步纳入国家科技计划项目管理

在国家科技计划数据管理方面,科技部先后出台了《国家重点基础研究发展计划资源环境领域项目数据汇交暂行办法》《科技基础性工作专项项目科学数据汇交管理办法》并将科学数据汇交纳入项目管理流程,并专门实施“十一五”科技计划项目课题形成的科技资源汇交,实现了一大批数据的汇交整合与开放共享^[7]。国家自然科学基金委员会通过“科学基金共享服务网”汇聚项目信息和数据,在共享航次计划、“黑河流域生态—水文过程集成研究”重大研究计划等重大科研计划实施过程中专门建立了数据汇交机制和数据中心。中国科学院通过科研信息化专项支撑科学数据汇交整合与管理,推进科技数据基础资源、海量存储与处理基础设施、数据集成与应用先进环境的建设与服务。气象、地震、海洋、测绘等部门在科学数据汇交管理方面已经取得了积极进展和成效,基本建立起行业科学数据管理政策制度和标准规范体系,实现了行业数据的持续汇交与积累。

2.5 注重科学数据获取和交换中的国际合作能力建设

在国际科学数据跨越国界共享的大环境中,我国在科学数据方面开展的国际合作日益密切,科学数据的国际开放度越来越高,推动我国的科学数据走出实验室,走向世界,同时将国际优质科学数据资源引入国内。科学数据的国际共享合作卓有成效,国际数据获取和交换能力得到长足发展。随着我国科学数据管理的信息化程度提高和新技术新方法的不断涌现,带动了科学数据资源获取和交换处理能力的快速增长。如与国际空间天气子午圈计划,将与俄罗斯、澳大利亚、加拿大、美国、巴西等国家的120°E+60°W子午线

附近的近百个监测台站进行合作,共同对子午线附近空间环境进行监测,实现日夜24小时、全球纬度的同时观测,开展太阳剧烈活动对地球空间环境影响的研究;蛋白质组学数据资源库iProX,与国际的蛋白质组数据中心PRIDE(欧洲)和PEPTIDE ATLAS(美国)进行数据交换,共同为国际学术界提供公共的数据资源。

3 思考与对策

3.1 准确把握大数据时代科学数据发展趋势

随着大数据时代到来,科技创新模式发生了革命性的变化,科学技术发展呈现出明显的大科学、量化研究特点,科技创新越来越依赖于大量、系统、高可信度的科学数据的全面收集和准确利用,政府决策也越来越多地依靠科学数据的科学分析和综合利用。在全球普遍重视大数据发展和科学数据管理的背景下,我国必须要紧跟国际发展步伐,充分利用国内外已有条件,加强与相关科研机构、重大科学数据组织的交流与合作,充分借鉴国内外先进经验和成熟做法,运用网络信息、大数据、云计算等技术手段,全方位提升我国科学数据工作水平。

3.2 加强科学数据共享监管,首要保证数据安全

在对科学数据进行汇聚和利用的过程中,各部门及法人单位必须要严格按照国家有关法律法规,做好科学数据监管和保密工作。科学没有国界,但科学数据是有国界的,平衡、适度的开放,维护科学数据开放的合法界限,建立数据共享和对外交流的安全审查机制也是科学数据开放共享的一个重要问题。因此,主管部门和法人单位有责任制定并严格执行科学数据安全保护措施,加强对数据下载、认证、授权等防护管理,防止数据被恶意使用。在制作、审核、登记、拷贝、传输、销毁等环节中,对科学数据进行严格管理。同时,也应按照国家网络安全管理规定,建立网络安全保障系统以及应急管理系统,确保科学数据安全放在首要位置。

3.3 完善布局,加强国家科学数据中心建设

突出重大问题和需求导向,紧密衔接重大

科技创新活动，完善科学数据中心的布局和顶层设计，加快资源集聚，集中力量打造一批具有领域、行业优势的权威性科学数据中心。数据中心的建设要与当前数据驱动科学研究范式充分结合，在科技创新链条中找准科学数据中心定位，发挥科学数据中心在数据长期保存、高效利用和驱动创新中的关键作用。依托科学数据中心进行科学数据的分级分类、加工整理和分析挖掘，有效整合各领域各学科的科学数据基础资源，构建符合我国发展需求的科学数据服务体系。

3.4 强化科学数据开放共享与综合利用

科学数据开放共享应充分遵循“开放为常态、不开放为例外”的基本原则，任何部门、单位都不应该以任何理由推卸科学数据开放共享责任和义务。在科学数据共享服务过程中，要紧密对接重大科技创新研发活动需求，面向国家科技重大专项、国家重点研发计划等科研需求，完善科学数据共享服务机制，整合集成相关科学数据，开展相关学科领域科学数据的综合交叉与分析应用。加大科学数据分析挖掘技术方法研究，特别是围绕重大需求，以互联网+为契机和手段，综合运用关联、聚类等分析方法，提升科学数据利用效率和水平。

3.5 多手段促进科学数据持续整合

科研项目是科学数据资源积累的重要来源，建立健全国家公共财政支持科研项目产生的科学数据的汇聚与积累，建立健全公益性科研活动获取和产生的科学数据收集与整合工作机制，是促进科学数据持续积累的重要手段。在科研项目立项阶段，应在项目申请书中明确数据管理计划以及数据开放共享方式和范围，在科研项目研究阶段加强科学数据的生产与整理，并将完成数据汇交作为项目验收的重要前提和基础。对于学术论文的相关支撑数据，也应该建立健全相应的科学数据汇交与管理机制，防止数据流失。同时，要

鼓励和引导通过社会资本支持形成的科学数据向相关领域科学数据中心进行汇交并开展数据共享服务。

3.6 加强知识产权保护，保障科学数据生产者权益

科学数据与科技文献一样，既是科学研究活动中的基础，也是科学研究活动的产物。然而，科学数据的生产、管理和共享却还不能像科技文献的发表和创作那样得到产权利益保护，这也直接影响了相应的评价机制和激励机制。在科学数据共享过程中，要对科学数据使用者的行为进行规范，科学数据使用者应遵守知识产权相关规定，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，体现对科学数据知识产权的尊重。对于侵犯知识产权的行为，将受到相应惩罚。

参考文献

- [1] 国家科技基础条件平台中心.国家科学数据资源发展报告2016[M].北京:科学技术文献出版社,2016: 12.
- [2] 魏悦,刘桂锋.基于数据生命周期的国外高校科学数据管理与共享政策分析[J].情报杂志,2017,36(5): 153-158.
- [3] 杨行,屈宝强,赫运涛,等.世界主要国家科学数据资源共享和管理的对比分析和启示[J].中国科技资源导刊,2016(6):18-25. DOI:10.3772/j.issn.1674-1544.2016.06.003.
- [4] 科学数据共享大循环[EB/OL].[2014-04-24].http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2014.
- [5] 吴蓉,顾立平,刘晶晶.国外学术期刊数据政策的调研与分析[J].图书情报工作,2015,59(7):99-105.
- [6] 叶玉江.加强科技平台工作推进科技资源管理[J].中国科技资源导刊,2015,47(2):1-6. DOI:10.3772/j.issn.1674-1544.2015.02.001.
- [7] 杨杰,宋佳,诸云强,等.科技基础性工作专项数据汇交共享平台建设[J].中国科技资源导刊,2017(5):52-59. DOI:10.3772/j.issn.1674-1544.2017.05.007.