

· 开放科学大家谈 ·

国内外科学数据管理与共享政策分析及 对国家自然科学基金的启示

周文能¹ 刘云^{2*} 王刚波³

1. 广东工业大学 生态环境与资源学院, 广州 510006
2. 中国科学院大学 公共政策与管理学院, 北京 100049
3. 北京理工大学 管理与经济学院, 北京 100081

[摘要] 本文阐释了开放科学的理念、科学数据共享的理论与实践意义以及科学数据的发展趋势,系统梳理了欧盟框架计划、英国研究与创新署、美国国家科学基金会和国立卫生研究院等主要国家或区域组织科学数据管理与共享的政策演进及做法,分析了我国科学数据管理与共享的政策演进以及国家自然科学基金项目数据管理与共享的现状,按照《科学数据管理办法》的总体要求,借鉴国际经验,结合科学基金项目数据管理的实际和发展需求,提出了改进国家自然科学基金项目数据管理与共享的政策建议。

[关键词] 国家自然科学基金;科学数据;管理与共享;政策

科学数据是科学研究重要的知识产出成果,也是科学研究赖以发展的基础保障,在当今大数据时代,科学数据已成为国家重要的战略性科技资源。加强科学数据管理与开放共享是提高科技资源利用效率、提升科技创新能力的重要手段。科学数据管理和共享体系能力建设直接关系到整个国家科技创新能力和创新效率。近10多年来,主要国家高度重视国家科技计划项目的科学数据管理问题,制定相应政策,以加强科研项目的数据管理、促进开放共享。美国国家科学基金会(National Science Foundation, NSF)于2010年1月发布的项目管理指南(Award and Administration Guide, 2010)规定:从2011年1月18日开始,所有提交到NSF的项目申请书必须包含一份“数据管理计划”(Data Management Plan)附件,该计划应详细描述申请者如何管理和传播研究项目所产生的数据,包括:(1)数据的类型;(2)数据标准;(3)数据获取与共享政策;(4)数据存档与保存计划。NSF所属各科学部也制定了各自专门的数据管理计划规定。美国



刘云 中国科学院大学长聘教授,公共政策与管理学院副院长、博士生导师。主要研究方向:科技政策、科技评价、创新管理、知识产权管理、数据挖掘。主持国家自然科学基金重点项目、国家社科基金重大项目、国家重点研发计划专项、国家国际科技合作计划项目、国家973计划前期研究专项等国家级课题40多项,发表学术论文200多篇,出版专著13部,获省部级科技奖6项。作为专家组组长或成员参与《国家中长期科学和技术发展规划纲要(2006—2020年)》、国家基础研究、国家科研基地和条件保障能力建设、国家自然科学基金等多项国家科技战略规划的研究、起草和评估。



周文能 广东工业大学教授,曾任科学技术部基础研究司二级巡视员。主要从事基础研究战略与政策、生态系统对全球变化的响应、野外科学观测数据分析、生物多样性保护与利用等领域研究工作。组织和参与了国家重点研发计划(基础研究部分)、科学数据管理、重大科研设施与仪器开放共享相关管理办法等多项政策文件制订工作,负责推进国家(重点)实验室、国家科学数据共享工程、国家应用数学中心和全国野外科学观测研究台站等国家科研基地发展战略布局与管理,为我国基础科学研究提供了政策保障与支持。发表学术论文10多篇,出版专著2部。

收稿日期:2022-05-19;修回日期:2022-08-08

* 通信作者,Email: liuyun@ucas.ac.cn

本文受到国家自然科学基金重点国际(地区)合作研究项目(71810107004)、国家社科基金重大项目(21ZDA016)和盖茨基金会资助项目(NV-035442)的资助。

国立卫生研究院(National Institutes of Health, NIH)2003年制定了《数据共享计划》,要求从2003年10月1日开始,所有向NIH申请项目经费在50万美元以上的科研人员都须提交一份包括数据共享计划或者数据不共享的说明。在数据共享计划中,申请者应简要描述:(1)数据共享的预期进度;(2)最终数据集格式;(3)将要提交的文档;(4)是否会提供一些分析工具;(5)是否要求共享者签署数据共享协议;(6)数据共享方式等。英国七大研究理事会除了工程与自然科学研究理事会外,其他6个理事会均制定了相应的数据管理政策,如生物技术与生物科学研究理事会要求研究者提交数据共享计划,包含:数据领域与类型、标准和元数据、与其他数据的关系、二次利用、数据共享的方法、共享限制、数据发布的时间表以及数据格式等,研究者须将出版物存放在英国公共医学中心(UK PubMed Central)。我国科学技术部分别于2001年和2014年开始试点推行科学数据共享工程和国家科技报告制度,涉及到国家科技计划项目的数据汇交,但是,工作进展一直举步维艰。2018年4月,国务院办公厅印发了《科学数据管理办法》,这是我国首次在国家层面出台的科学数据管理办法,该办法对利用政府财政资金资助的各级科技计划(专项、基金等)项目所形成的科学数据提出强制性汇交的要求,由项目牵头单位汇交到相关科学数据中心,实行规范管理、长期保存和共享利用。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,科学技术部、财政部对原有国家科技基础条件平台开展了优化调整工作,于2019年认定了“国家材料腐蚀与防护科学数据中心”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。尽管近年来我国出台了有关科技计划项目科学数据汇交与管理的指导性文件,但是,迄今国家科技计划包括科学基金项目还未制定可操作性的科学数据管理办法,科技计划项目申请指南均没有明确要求申请者须提交科学数据管理计划,科技计划项目产出科学数据如何汇交到科学数据中心,科学数据中心如何开展规范管理、数据加工保存、数据开放共享等方面的管理制度尚未健全。为适应开放科学发展趋势,借鉴国外科学基金组织的科学数据管理经验,结合我国科学基金项目数据管理的现状及需求,研究提出国家自然科学基金

金项目数据管理政策建议具有重要的现实意义。

1 关于开放科学与科学数据共享

1.1 开放科学运动的兴起

20世纪后期以来,学术界发起了一系列旨在克服传统封闭性科学弊端的开放科学运动,倡导“自由、开放、合作、共享”的理念,开放科学促使在知识创造、信息传播、成果共享、知识应用、乃至科研组织和绩效评价等方面出现了新模式、新规则和新机制。欧盟“促进欧洲研究迈向开放科学的培训”(Facilitate Open Science Training for European Research, FOSTER)项目开发了一个开放科学的分类系统,旨在覆盖整个开放科学领域^[1]。FOSTER研究报告提出,开放科学即开源软件、开放数据、开放获取、开放研究方法、开放同行评议、开放教育资源等一系列的总称,它们也可以称为开放科学在不同阶段的表现形式^[2]。其中,开放获取与开放数据代表着开放科学运动的两支重要力量。2015年10月,经济合作与发展组织(Organization for Economic Co-operation and Development, OECD)发布《使开放科学成为现实》报告指出,网络和在线平台为研究项目、科学文献和大型数据集的组织和发布提供了新的机遇;信息通信技术使得通过收集大规模数据和信息作为科学试验和研究的基础成为可能,使得科学越来越多地被数据所驱动;而在线存储为获取和利用科研信息成为可能。这些都加速了科研人员和领域之间的知识转移,开辟了合作和新研究方法的新途径,导致“开放科学”的快速发展。在全球科技合作日益广泛、各国科技创新组织相互渗透的今天,开放科学越来越上升为一项国家战略。芬兰教育文化部于2014年11月发布了《2014—2017年开放科学与研究路线图》(The Open Science and Research Roadmap 2014—2017),明确提出芬兰要在2017年成为全球开放科研的领先国家。欧盟Horizon 2020提出在欧洲的公共资助研究,要确保科学出版物的开放获取并且促进科学数据的开放获取,试图打造一个基于数字开放理念的科研创新体系。2013年1月,欧盟委员会发布《开放科学的开放基础设施:Horizon 2020咨询报告》,提出了三方面的行动:建设一个开放、互操作、面向科学数据的信息化基础设施;通过支持开放科学政策来组织欧盟数据空间;鼓励个人参与数字科学的开放团体^[3]。欧盟开始要求其资助科研项目的研究数据对公众开放,一些世界级的学术机构,比如欧洲

核子研究理事会开始构建开放数据存储库,学术出版商如 PLoS 和 Nature 也开始尝试科研试验数据出版。总之,当今科技正走向开放科学,主要国家、区域组织以及科学组织都在积极推进科技出版物的开放获取和科学数据开放共享。同时,人们也注意到,开放科学既带来机遇,也面临着一系列挑战^[4],如科学数据的开放共享就会涉及到数据所有权与知识产权、隐私、安全性等问题,开放科学政策应促进研究成果和科学数据的开发与再利用,同时应尊重知识产权的所有权,以促使它能够更加合理有效地实施。在任何一种开放科学的框架中,版权都必须得到保护和支持,推荐采用许可授权的方式扩大数字出版内容的使用。含有个人信息的数据集在医疗和社会科学的研究中大多都具有至关重要的作用,因而对个人数据的处理与保护应格外关注,从法律层面和技术支撑上都需要制定一个万全之策。通过信息安全网络系统保证机密信息、敏感数据的安全性,预防不明攻击与外来的有意攻击行为也十分重要。

1.2 科学数据共享的意义

科学数据是具有重要价值的战略性资源,加强科学数据的管理与开放共享是提高科技资源利用效率、提升科技创新能力的重要手段,是开放科学所提倡的重要科研行为,其对于科研个体及团体,甚至公众,均具有重要意义。早在 1985 年,美国科学院(National Academy of Sciences)就指出,科研数据共享能够强化开放科学需求,对原始结论加以验证与改进,在现有数据的基础上推进新研究,鼓励政策模拟和评估中更合理的使用经验数据,提升数据收集理论与方法,发展理论知识与分析技术,鼓励多角度分析,提供研究培训资源,预防错误数据,帮助研究者克服狭隘观点与自以为是的态度^[5]。

中美关于科学数据的定义存在一定的差异,但本质是相同的。中国科学技术部《国家科学数据中心建设技术规范》关于科学数据的界定是特指以公益性和基础性为主体的、具有科学研究应用价值的的数据资源,包括在观测、监测、调查、试验、实验以及研究等科学技术活动过程中产生的原始性数据,以及按照不同科技活动需求进行系统加工整理的各类数据。美国管理和预算办公室关于科学数据的界定是指科学界所接受并视为验证研究成果所必要的,已经记录下来的事实材料,但不包括以下内容:初步分析、科学论文草稿、未来研究计划、同行评议、与同事的沟通。上述记录下来的材料不包括实物(如实

验室标本)。

科学数据共享行为则表现为通过多种形式公开发布科学研究数据为他人所用。数据发布形式多样,基于需求的私下数据交换、公开收集和存储数据、将数据发布到公开网站、以期刊论文附件材料形式提供数据等,都可视为科学数据共享。合理的数据共享能够有效减少不同项目、不同主体间存在的重复性数据收集,避免整体层面科研基金不必要的浪费,提高科研效率与资金利用率。

从经济学角度看,在给定条件下,科学数据具备公共物品的性质—非竞争性与非排他性。即:在确定可以共享的科学数据中,科学数据共享行为的边际成本接近零,也就是说,科学数据共享行为不会影响他人对科学数据的使用中获得效用;科学数据共享行为不但不会排除他人使用该科学数据,反而会促进他人对该科学数据的使用。再加上科学数据具备的一些诸如一次性高成本投入,共享成本远低于再生产成本等特征,对科学数据的共享使用并不具备明显的负外部性,也不会随着共享使用程度的增加而造成公共资源的拥挤,因此,科学数据共享行为并不会出现公地悲剧的现象,实现科学数据共享能充分利用现有经费,减少无效劳动。同时,对于受限制的科学研究数据,也可以通过赋予产权的形式加以使用与交换,在一定程度上弥补了受限制科学数据所不具备的公共物品性质,通过产权交易的形式对原始数据生产者进行经济补偿,在尽量不损害原始数据生产者的前提下,推进科学数据共享。对于不能共享的数据,则需要相关法律的完善,做到有法可依,提高侵权成本,切实保障数据所有权持有者、知识产权持有者以及数据隐私涉及者的合法权利。

1.3 科学数据的发展趋势

随着人工智能、大数据、云计算等新一代信息技术的迅猛发展以及在科学研究领域的广泛应用,促使科学研究的范式发生深刻变革,科学数据在科学研究中的作用凸显,其发展呈现以下趋势:

(1) 数据密集型科研成为科学发现和研究工作的新特征

当今,许多前沿科学领域取得的重大突破和发现都与海量科学数据的全面收集、深入分析挖掘利用密切相关。继实验型科研、理论型科研、计算型科研之后,数据密集型科研已逐渐成为科学发现的重要方式^[6],该研究方式强调以数据作为科学发现的基础,呈现出以对海量数据处理和分析来发现新知识的特征。在迈向科技大国的过程中,我国已积累

了海量科学数据,然而共享与利用不足已成掣肘,其潜在价值未能得到充分挖掘,科学数据管理与共享是推进数据密集型科研和催生科学发现的的重要保障。

(2) 科学数据已成为国家重要的战略性资源

科学数据是联合国教科文组织界定的三种开放存取信息资源(科学出版物、教育资源、科学数据)之一。在数字化科研和大数据背景下,全球科学研究将越来越依赖于海量数据的开放共享与充分利用,科学数据已成为重要的科学资产和公共信息资源^[7]。世界各国对数据的依赖快速上升,国家竞争焦点已经从资本、土地、人口、资源的争夺扩展到对数据的争夺。数字主权将成为继边防、海防、空防之后另一个大国博弈的空间。

(3) 科学数据管理与共享服务体系是培育科技创新增长点的重要支撑

在一些重要科学研究领域,外部技术变革推动一些基本科学问题的新突破,其结果可能会深刻影响人类的生产和生活方式、社会的发展进程,以及国家的国际竞争优势和地位。中国在从科技大国向科技强国转变的过程中,要力争在重要科技领域实现跨越发展,跟上甚至引领世界科技发展新方向,掌握新一轮全球科技竞争的战略主动。依托于布局合理、高效运行的科学数据中心体系,将为国家引领前沿领域科学研究和培育新的创新增长点提供必要的条件与保障。

(4) 人工智能、大数据技术为有效发挥科技数据的支撑引领作用提供重要技术手段

随着新型信息储存材料、网络通信等技术的不断突破发展,使总量巨大且增长迅速的科技数据的存储、传输成为现实。随着一些国内外大型基础研究设施在数据管理方面的逐步开放,促使其成为数据有效汇集、传播以及复用的重要载体。这些来源于不同研究机构提供的共享科学数据经由智能优化算法、数据挖掘及可视化等人工智能、大数据技术手段分析,在支持科学研究、指导科学数据生产应用等方面承担的角色越来越重,也推动实现了科学数据从发布、传输、访问、利用的全生命周期有效管理。

2 主要国家和区域组织科学数据管理与共享政策分析

2.1 欧盟科学数据管理与共享政策

2011年12月,欧盟委员会发布“开放数据战略(Open Data Strategy for Europe)”,提出促进公共

资助项目产出科研数据的获取和重复使用^[8]。遵循该战略,2012年欧盟启动开放数据门户网站(European Data Portal),该网站所提供的开放数据中就包含有公共资助科研项目的数据。2016年欧盟委员会公布《Horizon 2020 框架下的 FAIR 数据管理指南》^[9],提出了科学数据管理的指导原则是“可查找、可访问,可互操作和可重用”(简称 FAIR 原则)。随后,欧盟借助第七框架和地平线 2020 计划持续高强度投入“开放科学云计划”^[10]等科学数据管理与共享关键技术研发和相关基础设施建设,强力推动科技数据的开放使用。

欧盟委员会在 2020 计划框架下的数据管理指南规定,凡参加“开放研究数据先导计划”的所有项目都必须提交数据管理计划(Data Management Plan, DMP),DMP 要求包含数据管理的关键要素,并描述数据的全生命周期,即数据产生、处理及管理的过程。初始的 DMP 应在项目开始前 6 个月提交,但随着项目的推进,DMP 的相关内容须进行相应的更新,如新的数据、项目联盟政策的变化、联盟成员和外部因素的变化等。为帮助科研人员制定 DMP,指南提供了一个 DMP 模版,并列出了 DMP 需要解决的问题摘要,主要包括:数据概述、数据的 FAIR 原则、资源配置、数据安全、数据伦理等。

为提高科研数据的使用效率,以吸引更多的科研人员参与到数据开放工作,欧盟委员会还专门为科研数据开放申请预算。如在科学出版物与科研数据开放获取管理指南规定^[11],符合开放数据补贴政策的,可申请相应的经费用于补贴在开放科学信息(包括已发表在学术期刊同行评审的成果论文和科研数据)过程中产生的费用成本。欧盟通过上述措施,极大地促进了科研人员知识发现和科研创新以及后续的知识集成和重新利用,有效推动并实现了各成员国间科技数据的汇集、共享和利用,提升了科技数据治理的安全水平和访问可靠程度,同时还发挥了吸聚外部优质科学数据资源的作用。

2.2 英国研究与创新署数据管理与共享政策

英国研究与创新署近年来发布了一系列数据管理政策。2011年发布《研究数据管理最佳实践指南》(2015年修订)^[12],提出了在数据管理政策方面的七大原则,包括数据来源、数据标准、数据存储、共享和开放获取权限等;2013年发布《通过拨款资助支持研究数据管理成本》^[13],对项目经费中支出的研究数据管理费做了明确规定;2016年发布《开放研究数据协议》^[14],确定开放研究数据的十项原则,

适用于公共资助的所有研究领域。

依据上述政策要求,各专业研究理事会等机构均制定了相关数据管理政策,规定由其全额或部分资助的项目,在提交正式项目申请时必须制定完善的数据管理和共享计划,且作为同行评议的内容之一。数据管理计划应涵盖科研项目可能产生的所有数据,包括原始数据和发布数据。数据管理计划内容包括:项目的数据来源、分析现在可能利用的数据与项目研究所需要数据存在的差距、研究项目将产生的数据相关信息、数据质量保证及数据备份计划、数据共享困难及应采取的措施、数据保密性与数据使用伦理、数据知识产权、项目组成员的数据管理职责等。

科学数据公开要充分考虑到数据创造人员的优先使用权,以及相关数据公开需要准备时间,通常允许科学数据公开有一定的时滞。科学数据公布的时间节点应根据其本身的性质和类型、研究方法和学科领域,一般不应晚于主要研究发现发布之日。如果资助方有特殊要求或某一学科领域已有惯例,科研人员及科研机构应予以遵守。

2.3 美国国家科学基金会(NSF)的数据管理与共享政策

美国 NSF 在 2011 年度《申请与资助政策指南》^[15] 中首次新增了科学数据管理条款(Data Management Plan Requirements),要求项目申请书必须附带一个“数据管理计划”,详细描述如何遵守 NSF 的“传播与分享研究成果”政策,主要内容包包括:数据类型、数据与元数据形式及内容标准、获取与分享政策、再利用规定、存储数据计划等,并且这份数据管理计划同样要接受同行评审。同时,NSF 要求各科学部可以根据其负责学科领域特点研究制定更为详细的数据管理计划规范与要求。如生物学部依据学科发展特点及数据特征,已三次更新相关科研数据管理要求,当前使用的是 2018 年 2 月修订的数据管理指南。

NSF 要求数据管理计划应提供如下内容:项目预期产生数据与元数据的类型与标准;授权期结束后用于储存和保护数据的机构设施;关于个人隐私、保密或安全信息、知识产权以及其他权利的数据公开共享政策;数据再利用、再分配以及衍生利用的相关规定;获取数据、样本、研究成果的规定以及资助结束后的储存计划;授权到期后各方在数据管理中应当承担的责任等。申请书中的数据管理计划要接受同行评审。

NSF 指定各学科的科学数据汇交平台,强调科学数据汇交的及时性,要求在数据完成后即可汇交,资助周期较长的研究项目鼓励逐年汇交。科学数据管理计划实施采用项目负责人(Principle Investigator, PI)负责制。NSF 资助项目的年度报告(Annual Report)与最终报告(The Final Report)中应当包含数据管理计划的实施情况,并将其作为项目验收评审的一部分。其中,年度报告应包含本年度研究成果的数据管理与共享情况;最终报告应包含数据管理计划的总体实施情况以及最终数据管理计划与原定数据管理计划的变更情况等。

2.4 美国国立卫生研究院(NIH)数据管理与共享政策

相较于 NSF 的数据管理计划,NIH 的数据共享政策要早很多。1999 年,NIH 发布了《NIH 研究工具政策》(NIH Research Tools Policy),从政策层面保障政府资助研究中特殊数据的需求与扩散。2002 年,NIH 发布了《共享研究数据的草案声明》(NIH Announces Draft Statement on Sharing Research Data)期望并支持 NIH 资助研究的最终研究数据能够及时发布与共享。2003 年,NIH 发布了《关于共享研究数据的最终声明》(Final NIH Statement on Sharing Research Data)与《NIH 数据共享政策》(NIH Data Sharing Policy)。2004 年,针对数据共享政策的常见问题,NIH 发布了《数据共享常见问题》(Frequently Asked Questions of Data Sharing)。之后,NIH 又相继发布了《NIH 公共知情权政策》(NIH Public Access Policy)、《数据标准与寻常数据元素资源指南》(Data Standards and Common Data Elements Resource Guide)、《数据存储资源指南》(Data Repositories Resource Guide)、《NIH 资助医学实验信息扩散政策》(Policy on Dissemination of NIH-Funded Clinical Trial Information)等一系列数据共享政策指导文件。

NIH 于 2003 年发布的《关于共享研究数据的最终声明》和《NIH 数据共享政策》,明确要求:自 2003 年 10 月 1 日及以后,在申请 NIH 资助的项目申请中,研究期内任一年度直接经费在 50 万美元及以上的项目申请书必须包含一份关于最终研究数据出于研究目的的共享计划。如果该研究的最终数据无法共享,则列明原因。数据共享计划主要包括:数据共享的预期日程、最终数据集的格式、将要提供的文件编制、是否提供分析工具、是否要求共享协议以及数据共享方法等。

针对数据存放问题,NIH支持了各研究领域和数据类型的众多数据存储库,包括PDB、GenBank等在内的82个数据中心,为研究人员存储其数据(即dbGaP)指定相关的汇交数据中心。针对参与研究的人类对象的权利与隐私保护问题,NIH规定:保护人类对象的权利与隐私是调研者、其评议委员会及其所属机构的责任;此类数据应该采取数据编修,去身份识别,匿名化等有效手段,将有可能直接泄露或者演绎性泄露个人信息的数据进行清洗,或采取数据包等限制性方法进行数据共享,将未授权的个人信息公开泄露风险降到最低。针对数据所有权问题,NIH也做出规定:小企业可以在资助结束后4年内仍然保留数据所有权;当机构出于保护知识产权目的而延期披露研究发现时,NIH允许该机构的披露延期30~60天。

3 我国科学数据管理与共享的政策演进及 自然科学基金项目数据管理与共享的现状

3.1 我国科学数据管理与共享的政策演进

我国的科学数据管理与共享工作始于科学技术部的科技基础性工作。1999年,科学技术部在科技基础性工作和社会公益研究专项中,启动了科技基础数据库建设。2001年,科学技术部《社会发展科技工作要点(2001—2005年)》首次提出制定可持续发展数据规范与标准,建立数据共享管理办法,形成完整的数据共享支持系统,为我国可持续发展网络(SDN)建设创造条件。同年,科学技术部成立“科学数据共享调研组”,在基础性、公益性行业领域开展科学数据资源调研、共享机制研究,提出我国科学数据共享工程实施意见。2001年12月,科学技术部启动了第一个共享试点——气象科学数据共享试点。

2001年,科学技术部、教育部、中国科学院、中国工程院、国家自然科学基金委员会(以下简称“自然科学基金委”)联合印发《关于进一步增强原始性创新能力的意见》提出要建立重要科研设备和科学数据资料共享机制,实施科学数据共享工程,建设国家科学数据管理中心,建立国家科技计划项目数据资料汇交制度,对科学数据资料实行分时段、分级、分类的共享。2002年12月,科学技术部启动8个领域的科学数据共享试点:农业、林业、水文水资源、地震、测绘、地球系统科学、可持续发展、农村现代科技。

2003年,科学技术部、财政部联合设立科技基

础条件平台建设专项,科学数据共享工程作为重要组成部分纳入其中。同年,科学技术部印发了《关于加强国家科技计划成果管理的暂行规定》,要求将项目实施取得的实验报告、数据手稿、图纸、声像及其他形式的科学数据进行收集整理,建立档案,并作为项目通过验收的基本条件。2005年,科学技术部发布《国家科学数据中心建设技术规范》,明确了科学数据管理与共享的概念、管理方法等。2008年,科学技术部启动973计划资源环境领域项目数据汇交工作。

2018年,国务院办公厅印发了《科学数据管理办法》^[16],首次在国家层面对科学数据管理的范围、职责、采集、汇交、保存、共享、利用、保密与安全等做了明确的规定,主要措施包括五个方面:一是明确主管部门和法人单位的职责,强化了法人单位的主体责任,体现“谁拥有、谁负责”“谁开放、谁受益”,同时,对科学数据的生产者和使用者提出了明确要求;二是加强科学数据交流和利用的监管,明确主管部门和法人单位依法确定科学数据的密级及开放条件;三是加强知识产权保护,对科学数据使用者的行为进行了规范,体现对科学数据知识产权的尊重,同时,对科学数据生产者也做出了约束,如出现数据造假等行为,将受到相应惩罚;四是加强数据积累,促进开放共享,要求科技计划项目产生的科学数据进行强制性汇交,并通过科学数据中心规范管理和长期保存;五是加强科学数据管理能力建设,提出法人单位要在岗位设置、绩效收入、职称评定等方面建立激励机制,将科学数据工作情况作为重要的考核内容。

3.2 国家自然科学基金项目数据管理与共享现状

多年来,自然科学基金委在推进基金项目成果开放共享方面做出了积极的努力,主要表现为建立基金项目科研报告制度和项目成果信息开放共享平台两个方面。但是,自然科学基金委至今尚没有出台统一的科学基金资助项目科学数据管理与共享政策,科学数据管理与共享机制尚不健全。

3.2.1 科学基金项目的科研报告制度

自然科学基金委自1986年成立以来,借鉴发达国家成熟经验,根据科学基金项目管理工作的需要,形成了一套要求明确且内容规范的项目科研报告制度,为科学基金项目管理的规范高效奠定了实践基础。特别自国务院2007年颁布《国家自然科学基金管理条例》(以下简称“《条例》”)以来,从顶层设计上将包括项目科研报告制度在内的科学基金相关制度

以法律的形式加以确立和规范,为项目科研报告和科学基金信息公开工作提供了有力的法律保障。目前,自然科学基金委将各种类型科学基金项目全过程管理中的项目进展报告、结题报告、成果报告等相关信息,全部纳入科学基金项目科研报告制度和科学基金网络信息系统,进一步增强了科学基金管理工作的透明度。

经过多年的经验积累,自然科学基金委逐步形成了围绕着科学基金项目管理体系而设立的多层级主体共同负责的科研报告制度,责任主体涉及项目负责人、依托单位、自然科学基金委相关部门及人员等。其中,项目负责人依据自然科学基金委要求,在规定的时间内按规定的格式和内容,向依托单位提交项目进展报告、结题报告、成果报告等科学基金项目所需报告材料;依托单位负责审核项目负责人所提交报告的真实性和合规性,每年按时向自然科学基金委统一提交所在单位的项目相关报告,并说明本单位的报告完成和提交情况;自然科学基金委的管理人员收到这些报告后,根据项目主任、科学处、综合处和科学部的职责要求,分级审核科学基金项目科研报告的相关内容,并向项目负责人和依托单位反馈相关意见。除了《条例》对科学基金项目科研报告制度相关主体的责任进行了详细规定之外,自然科学基金委每年按照项目管理流程制定的主要工作进程也对何人何时提交或受理或审核何种材料或报告的时间节点加以细化,保证了责任清晰、分级管理的科研报告制度运行高效规范,增强了科研人员和管理人员对财政资金使用的责任感和使命感。

科学基金项目科研报告制度不仅是科学基金的管理手段,也是科学基金服务社会的重要载体。为了更好地保存这些珍贵资料,并使其发挥更大作用,便于进一步总结经验,提高科研管理水平,自然科学基金委自2002年开始就启动了项目档案资料数字化工程,对2002年及以前的7万多个项目档案完成数字化处理。由于2003年起科学基金项目实行网上提交电子化项目申请书、计划书、进展报告和结题报告等相关科研报告,目前,自然科学基金委保存了自成立以来所有受资助项目资料电子版,可以反映过去30多年我国科学家对各学科发展态势的认识和判断,以及科学基金管理制度逐渐完善和我国基础研究水平不断提升的真实轨迹。

2003年科学基金网络信息系统面向全社会推广使用之后,自然科学基金委逐步实现了科学基金项目申请、项目评审、项目中后期管理、项目成果登

记和项目信息发布的网络化业务管理,作为科学基金项目科研报告组成部分的项目进展报告和结题报告等全部实现了电子化联网提交和查询,既便于科学基金管理人员全面便捷地掌握项目资助情况,也进一步推动了项目数字化管理和信息公开制度建设,促进科学基金管理实行更加公开透明和公正合理的运行机制。

3.2.2 科学基金项目成果信息共享平台

自然科学基金委积极推动科学基金项目成果信息共享服务网系统建设。2013年上线的新版科学基金共享服务网设置了“资助项目检索”“结题项目检索”“成果检索”“导航”以及“用户反馈”等不同功能板块;根据“按申请领域统计”“按项目类型统计”“按成果类型统计”以及“按结题年份统计”等主题提供分析统计功能和数据图表展示;提供智能化检索服务,通过历年项目资助信息统计、项目指南链接、成果信息发布,向科研人员提供相关学科领域发展动态。截至2014年,该网共公布科学基金结题项目11.7万个,项目成果信息167万条,并全文发布部分项目的结题报告,为广大科研人员、科研管理人员和社会公众提供了涉及科学基金项目科研报告及其相关成果等知识关联与信息获取的一站式共享服务,极大地促进了科研报告信息的广泛公开与利用。

2019年12月,科学基金共享服务网(科技成果信息系统 <http://output.nsf.gov.cn/>)对结题项目与项目成果数据进行了更新,共计更新结题项目36483项,其中面上项目15107项、重点项目563项、重大项目114项、重大研究计划463项、创新研究群体科学基金41项、专项基金项目86项、联合基金项目674项、青年科学基金项目16163项、地区科学基金项目2741项、海外及港澳学者合作研究基金134项、优秀青年科学基金项目397项;更新项目成果520100条,其中期刊论文376770条,会议论文52844条,著作5524条,奖励13359条,专利71603条。科学基金共享服务网(科技成果信息系统)已公开截至2018年度结题项目343577项,项目成果3942710条。

3.2.3 科学基金项目数据管理与共享的相关政策

目前,自然科学基金委针对科学基金项目数据管理与共享的政策缺失,但与此相关的政策包括:2008年11月发布的《国家自然科学基金委员会信息公开管理办法》、2014年5月发布的《国家自然科学基金委关于受资助项目科研论文实行开放获取的

政策声明》、2015年5月发布的《国家自然科学基金委员会基础研究知识库开放获取政策实施细则》、2015年9月发布的《国家自然科学基金资助项目研究成果管理办法》等，相关政策涉及科学基金项目数据管理与共享的要点如下：

2008年《国家自然科学基金委员会信息公开管理办法》第九条 自然科学基金委应当确定主动公开的信息的具体内容，其中，包括：基金资助项目的结题报告摘要、申请摘要和研究成果报告。

2014年《国家自然科学基金委员会关于受资助项目科研论文实行开放获取的政策声明》要求国家自然科学基金全部或部分资助的科研项目投稿并在学术期刊上发表研究论文的作者应在论文发表时，将同行评议后录用的最终审定稿，存储到国家自然科学基金委员会的知识库，不晚于发表后12个月开放获取。如果出版社允许提前开放获取，应予提前；如果论文是开放出版的，或出版社允许存储最终出版PDF版的，应存储论文出版PDF版，并立即开放获取。并授权自然科学基金委相关部门制定受资助项目科研论文开放获取细则，建立开放获取知识库，按照国际通用规范，支持公众通过网络检索开放获取内容。同时，要求相关部门积极与国家相关部门和公共教育科研机构合作，推动国家各类科技计划和公共机构资助项目已发表论文的开放获取；要求相关部门积极与国际科技界合作，共同促进公共资助项目科研论文在全球的开放共享。

2015年《国家自然科学基金委员会基础研究知识库开放获取政策实施细则》第四部分“提交”中要求基金项目负责人通过“自然科学基金会项目管理系统”提交与项目研究相关的论文全文和元数据，将经过相关处理后转入“基础研究知识库”进行存储，并提供开放获取。第五部分“使用”中要求“基础研究知识库”将严格遵守《中华人民共和国著作权法》《信息网络传播权保护条例》及其他相关法律法规的规定，切实保护作者及相关权利人的合法权益。其中，针对元数据使用，“基础研究知识库”允许用户为个人学习、研究或者欣赏，免费获取和使用研究论文元数据；用户在使用研究论文元数据时必须给出原始元数据记录的超链接或DOI标示符。第三方在超出许可条件下使用作品，需得到著作权人的许可。

2015年《国家自然科学基金资助项目研究成果管理办法》涉及资助项目科技成果及管理、使用及开放的有关规定。

由上述政策可见，自然科学基金委关于科技成果信息公开、资助项目科研论文开放获取、资助项目成果管理与共享服务等出台了一系列政策和管理办法，从资助项目的科学数据共享看，仅重点针对发表论文及其相关元数据的开放共享有一定的规定，但是，尚没有出台专门的类似美国NSF数据管理计划、NIH数据共享计划的科学数据管理与共享政策，对资助项目科研过程中产生的科学数据的数据范围、数据保存、数据汇交、数据开放、数据保护等均没有明确规定。按照国务院办公厅《科学数据管理办法》的要求，自然科学基金委亟需建立健全科学基金项目数据管理与共享的制度。

4 对国家自然科学基金项目数据管理与共享的启示及政策建议

鉴于科学数据管理在支撑国家科技进步与创新中的关键作用，2018年3月，国务院办公厅印发了《科学数据管理办法》^[16]，这是我国首次在国家层面出台的《科学数据管理办法》。作为科学数据重要来源之一，自然科学基金委积极探索科研数据管理方式^[17]，重点围绕资助项目发表论文及相关数据的开放共享，建立了项目成果信息开放共享平台。但与国外科学基金资助机构相比，对资助项目产出科学数据的范围、保存、汇交、开放、保护等方面均未明确规定^[18]。因此，根据《科学数据管理办法》要求，借鉴美国NSF、NIH数据管理与共享政策的经验^[19, 20]，结合科学基金项目数据管理的实际和需求，提出国家自然科学基金项目数据管理政策建议如下：

(1) 建议研究制定《国家自然科学基金项目数据管理办法》

根据《科学数据管理办法》的总体要求，研究制定《国家自然科学基金项目数据管理办法》，用以指导和落实项目全周期、全要素的科学数据管理工作，规范数据管理的要求、流程、共享和利用机制、知识产权保护、评估监督以及各相关主体的职责，提高基金项目数据管理与共享的水平，有效发挥科学基金项目科学数据在国家创新体系中的独特作用。

(2) 启动实施科学数据管理计划

建议将科学数据管理计划纳入到项目申请书中，数据管理计划作为项目申请所需材料中的必须材料，缺少数据管理计划的项目申请不给予受理。自然科学基金委制定统一的项目申请书中的数据管

理计划要求,各科学部可以结合本学部相关学科领域的科学数据特点,制定体现本科学部相关学科特点的项目申请数据管理计划的特殊要求。在数据管理计划中,申请人应陈述两个核心问题:一是所申请项目实施过程中将会产生何种数据以及何种成果;二是申请人将如何管理这些数据与成果。数据管理计划主要包括:1)项目预期产生的数据与元数据的类型与标准;2)授权期结束后用于储存和保护数据的机构设施;3)关于个人隐私、保密或安全信息、知识产权以及其他权利的数据公开共享政策;4)数据再利用、再分配以及衍生利用的相关规定;5)获取数据、样本以及研究成果的规定以及资助结束后的储存计划;6)授权到期后各方在数据管理中应当承担的责任。同时,要建立先汇交项目科学数据、再项目结题或验收的机制,明确将科学数据管理计划执行的情况作为项目结题评审的重要考核指标,并把基于数据全生命周期的科学数据管理计划执行情况作为申请新项目的资质条件。在项目评审中,将申请项目的数据管理计划列为评审的重要内容之一。

(3) 对项目科学数据管理计划实施动态跟踪管理

鉴于科学研究具有不确定性的特点,科学数据的管理也需要随着项目的研究进展、研究目标和研究内容的变化而进行动态调整与更新。项目负责人需要对因研究工作调整而形成的数据变化情况进行实时更新,建议借鉴美国 NSF 经验由项目负责人以提交年度报告形式说明科学数据动态调整与更新等相关内容。现有的科学基金项目成果信息共享服务网继续担负科学基金项目产生的论文、专利、奖励等成果信息汇交与共享服务的职能。明确要求项目负责人在科学数据管理计划中提出科学数据的质量标准、潜在安全风险及预防措施,在数据汇交和提交结题报告时提交科学数据可用性说明、公开权限与范围,既保证科学数据的质量,又防范科学数据的滥用行为。明确科学数据所有权的归属,对研究人员或项目转移后科学数据所有权的移交与转让做出规定。

(4) 合理布局项目科学数据汇交与管理的科学数据中心

科学基金项目数据管理采取依托单位和自然科学基金委指定的科学数据中心双重管理模式,项目依托单位负责对本单位科学基金项目数据实施有效管理,科学基金项目负责人在项目结题时或结题后按照任务书要求向自然科学基金委指定的科学数据

中心汇交项目产出的科学数据,科学数据中心对基金项目科学数据实施集中管理与共享服务。充分利用现有的 20 个国家科学数据中心和 30 个国家生物种质和实验材料资源库,结合资助项目科学数据的特点,重点布局 and 选择国家科学数据平台作为各相关领域科学基金项目科学数据汇交的科学数据中心^[21],对于现有国家科学数据平台不能满足特定领域科学基金项目科学数据汇交需要的,可选择具有数据资源平台和服务优势的依托单位建立相关领域的科学数据中心,并授权其承担特定领域科学基金项目数据汇交的数据中心的职能。现有的科学基金项目成果信息共享服务网继续担负基金项目产生的论文、专利、奖励等成果信息汇交与共享服务的职能。科学基金应研究制定科学数据汇交质量评价指南,指导各学科科学数据中心开展汇交质量评价工作。

(5) 制定科学基金项目数据开放共享及保护的管理制度

科学基金项目负责人将项目研究过程中产生的实验计划、实验方法、实验流程记录、测试分析数据、元数据、软件源代码等各类数据汇交到科学数据中心,并对各类数据开放共享的时间、授权第三方使用的权限、数据安全、隐私保护、数据产权保护等作出说明。自然科学基金委各科学部数据管理与共享专家组对其进行评议,并将评议意见反馈给项目负责人。科学数据中心对数据质量进行评估,与项目依托单位签订科学数据开放共享协议,明确各方在数据的开放、共享、使用及保护等方面的权益,科学数据中心依据协议实施科学数据的管理、开放共享及保护。数据共享应保障数据安全,对数据重要程度与安全等级进行分级,设立必要的紧急情况应对机制,并建立完善的保障与备份服务、系统与机构,保障数据不会因人为因素或不可抗力因素遭到破坏。设定数据在项目期间与项目结束后可以提供的数据共享的密级分类、共享范围,以及提供数据共享的方式,如数据库在线提供,线下样本提供等。同时,针对不同密级的数据,应在不涉及泄密的前提下,提供诸如数据包、内部网络等不同的数据共享服务。对于不能共享的数据,也应加以解释。

此外,数据共享应考虑到知识产权的归属问题以及与之相关的成果归属问题。制定与知识产权与成果保护相关的规章制度,拟定于知识产权与成果保护相关的用户协议,通过自我约束、用户约束,降低知识产权与成果保护风险的同时,为科学成果顺

利发表提供积极的支持与坚实的支撑。针对数据本身,数据分配与再分配,数据利用与再利用等过程中可能产生的知识产权的形式、内容,在此过程中产生的知识产权的相关规定,以及可能通过签署协议等方式明确知识产权与成果所有权范围的协议简述。项目结束后,项目获得的数据仍然可以由项目负责人保留一段时间,该时段应在申请书予以明确。对于未能按照申请书中明确的时间内将数据共享的情况,需要提出说明,并明确可以共享的时间。

(6) 加强对科学基金项目数据管理与共享的评估与监督

建议组建国家自然科学基金项目科学数据管理与共享专家委员会,负责对科学基金项目数据管理工作进行宏观指导、审议相关科学管理原则及重要实施计划,并对科学数据管理和共享服务的绩效进行定期评估。建议各科学部组建学科领域科学数据管理与共享专家工作组,负责对本学部科学基金项目数据管理与共享标准制定,审议资助项目的科学数据管理计划、对项目组及依托单位提交的科学数据的完整性和质量进行评估。

参 考 文 献

- [1] The Community Research and Development Information Service (CORDIS). Fact Sheet—FOSTER (Facilitate Open Science Training for European Research). (2022-05-25)/[2022-06-22]. <https://cordis.europa.eu/project/id/612425/en?format=pdf>.
- [2] The Community Research and Development Information Service (CORDIS). Final Report—FOSTER (Facilitate Open Science Training for European Research). (2017-01-11)/[2022-06-22]. <https://cordis.europa.eu/docs/results/612/612425/final1-foster-final-report.pdf>.
- [3] Hudson R, Pires C. Open Infrastructures for Open Science Horizon 2020. (2016-01-28)/[2022-06-22]. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.422.6056&rep=rep1&type=pdf>.
- [4] 温亮明,李洋,郭蕾.国内外开放科学的实践进展与未来探索.图书情报工作,2021,65(24):109—122.
- [5] National Academies of Sciences, Engineering, and Medicine. Open Science by Design: Realizing a vision for 21st century research. Washington: The National Academies Press, 2018.
- [6] 赵秋红,李元睿,邓修权,等.科学基金资助机构视角下的科学数据管理研究.中国科学院院刊,2021,36(12):1456—1462.
- [7] 王卷乐,王明明,石蕾,等.科学数据管理态势及其对我国地球科学领域的启示.地球科学进展,2019,34(3):306—315.
- [8] European Commission. Open Data Strategy for Europe. (2011-12-12)/[2021-07-26]. https://ec.europa.eu/commission/presscorner/detail/en/IP_11_1524.
- [9] European Commission. Guidelines on FAIR Data Management in Horizon 2020. (2016-07-26)/[2021-07-26]. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- [10] European Commission. European Open Science Cloud (EOSC). (2019-05-15)/[2021-07-26]. <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>.
- [11] European Commission. Guidelines on Open Access to Scientific Publications and Research Data. (2018-09-21)/[2021-07-26]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- [12] Research Councils UK. RCUK Common Principles on Data Policy. (2018-09-21)/[2021-07-26]. <https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/>.
- [13] Research Councils UK. Supporting research data management costs through funding. (2013-07-09)/[2021-07-26]. <https://www.ukri.org/funding/information-for-award-holders/data-policy/supporting-research-data-management-costs-through-funding/>.
- [14] HEFCE, UKRI, Universities UK and the Wellcome Trust. The Concordat on Open Research Data. (2016-07-28)/[2021-07-26]. <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>.
- [15] National Science Foundation (NSF). Proposal & Award Policies & Procedures Guide. (2018-10-30)/[2021-07-26]. https://www.nsf.gov/publications/pub_summ.jsp?ods_key=pappg.
- [16] 国务院办公厅.科学数据管理办法.(2018-04-02)/[2021-10-30]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
- [17] 李东,于笑丰,杜一,等.国家自然科学基金资助成果开放共享平台:现状与展望.中国科学基金,2021,35(5):808—814.
- [18] 阿儒涵,吴丛,李晓轩.科研数据开放的国际实践及对我国的启示.中国科学院院刊,2020,35(1):11—18.
- [19] 谢艳秋,钱鹏.国外科学数据共享政策的发展研究.新世纪图书馆,2014(1):67—71.
- [20] 汪俊.美国科学数据共享的经验借鉴及其对我国科学基金启示:以NSF和NIH为例.中国科学基金,2016,30(1):69—75.
- [21] 王瑞丹,杨静,高孟绪,等.加强和规范我国科学数据管理的思考.中国科技资源导刊,2018,50(2):1—5.

Policy Analysis of Domestic and Foreign Scientific Data Management and Sharing and its Enlightenment to National Natural Science Foundation of China

Wenneng Zhou¹ Yun Liu^{2*} Gangbo Wang³

1. *School of Ecology, Environment and Resources, Guangdong University of Technology, Guangzhou 510006*

2. *School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049*

3. *School of Management and Economics, Beijing Institute of Technology, Beijing 100081*

Abstract This paper explains the concept of open science, the theoretical and practical significances of scientific data sharing and the development trend of scientific data; systematically arranges the policy evolution and practice of scientific data management and sharing for major national or regional organizations such as the European Union Framework Programme, the UK Research and Innovation Agency, the National Science Foundation of the United States, and the National Institutes of Health and so on; analyzes the policy evolution of scientific data management and sharing in China and the current status of data management and sharing for National Natural Science Fund of China. According to the general requirements of scientific data management methods and international experience, combined with the actual and development needs of data management for science fund, policy recommendations for improving the data management and sharing for National Natural Science Fund are put forward.

Keywords National Natural Science Foundation of China; scientific data; management and sharing; policy

(责任编辑 崔国增 张强)

* Corresponding Author, Email: liuyun@ucas.ac.cn