

· 科学论坛 ·

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

汪 俊*

(合肥工业大学科学技术研究院,合肥 230009)

[摘要] 大科学时代,科学数据是具有重要价值的战略性资源,加强科学数据共享是提高科技投资效率、提升创新能力的重要手段,然而科学数据共享过程中面临着一些困境。文章首先对科学数据共享相关概念进行清晰界定,从经济学基础和法理基础两个方面,分析科学数据共享的内在逻辑机理。系统比较了美国 NSF 和 NIH 在科学数据共享的理念、内容和方式等方面差异,深度剖析了其在科学数据共享方面面临的困境及其解决方案,结合国家自然科学基金在科学数据共享方面的实践进展,提出了完善我国科学数据共享的对策建议。

[关键词] 科学数据共享;NSF;NIH;经验借鉴

科学数据共享的过程是反复实验和验证研究结论的过程,有助于改善研究方法、改进研究成果,让科学研究结论更具普适性。而且数据共享对形成良好的科学研究氛围、促进社会协作和提升科研投资效率等具有重要推动作用。同时,科学数据的共享和重复利用能进一步增加数据自身的价值,使得其他研究者能够以全新的视角去诠释数据、解读数据、分析数据,有助于产生新的研究成果。反之,重要科学数据信息的私人占有和封闭保守,会阻碍研究结果得到进一步的核查与验证,这与科学的基本原则背道而驰^[1]。但是随着研究目标、方法、使用仪器、涉及的研究群体等的不同,以及数据的类型、数据共享形式等的差异,使得人们对数据的理解产生巨大差异,而且出于数据共享成本、隐私保护和知识产权保护等的考虑,使得科学数据共享过程中存在诸多障碍。

随着科学技术复杂性和全球化科技活动的不断增强,导致全球范围内科学数据信息资源的深度交互。大科学时代,学科交融使得任何一个研究者和项目团队仅仅依靠自身的力量,只能通过有限的手段获取有限的时间、空间和专业范围内的科学数据,为了全面、客观地认识研究对象,均需要其他研究者

和研究项目的科学数据支持。如何让海量科学数据信息资源在全社会流动起来,规范信息数据在管理和社会流动中的行为,最大限度地发挥科学数据资源的作用,是科学数据共享的核心^[2]。

科学数据形式多样,收集方法和目的各异,一旦脱离了数据最初产生的背景则很难解释。科学数据共享包括以下4个理由:进行研究再现或验证;使公共资助研究的结果为公众所用;使其他人利用现有数据提出新的科学问题;提升研究和创新水平。不同的利益相关者动机和激励的差异,使得人们对这些理由的认识不同,因此科学数据共享是一个错综复杂的问题,什么数据应该被共享、向谁共享、在什么条件下共享、采取何种方式共享等,这些问题的探讨将贯穿于整个科学数据共享的理论研究和实践发展^[3]。科学数据共享的逻辑起点源于其资源属性,因此,本文首先对科学数据共享相关概念进行清晰界定,从经济学基础和法理基础两个方面,分析科学数据共享的内在逻辑机理,深入剖析了美国 NSF 和 NIH 在科学数据共享方面的实践做法,以为建立健全符合客观规律和现阶段我国国情的科学数据共享机制,构建高效、有序的科学数据共享体系提供决策依据。

收稿日期:2015-07-20;修回日期:2015-11-04

* 通信作者,Email: wangjunjun@hfut.edu.cn

1 科学数据共享的理论基础

1.1 科学数据共享的概念

科学数据是人类在从事科学研究活动中所产生的原始观测数据、探测数据、试验数据、实验数据、调查数据、考察数据、遥感数据、统计数据、研究数据以及相关的元数据和按照某种特定的需求系统加工的数据,是具有重要价值的战略性资源^[4]。科学数据是大数据时代最基本、最活跃、影响最广泛的科技创新资源,具有重要的科学价值、经济价值和社会价值。科学数据的科学价值体现为其既是科学研究的基础也是科学研究的牵引力和推动力。科学数据的经济价值表现为可直接或间接为数据创建者、数据使用者带来经济效益。而社会价值主要体现在满足科技进步与创新、经济增长、社会发展和国家安全等多种需求以及科普教育和全民素质的提高等方面。科学数据多重价值的实现不仅取决于社会对科学数据效用的需求,又与科学数据的可得性、准确性、系统性和可共享性等紧密相关^[5]。

本文将科学数据共享界定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。然而,不同科学数据的可用性、可信性和价值存在很大差异。科学数据共享的实质是实现科学数据资源的开放与共用。开放是前提,共用是目的。只有共用才能使科学数据资源通过广泛的社会服务,将其潜在的科学价值、经济价值和社会价值得到最大限度的发挥而增值,也只有共用才能减少国家的多轮重复投资,增加科研投资效率。

1.2 科学数据共享的经济学基础

科学数据有其特殊的投入产出规律。就投入而言,科学数据的形成主要存在两方面费用:极为昂贵的开发费用和非常廉价的复制费用。这意味着重复地开发将会使得科研总投资成倍地增加,而把科学数据复制给更多的人,并不需要追加太多的投资。就产出而言,科学数据的价值除了受数据本身的客观因素影响外,还与数据需求者、开发者和利用者的知识、能力和方法密切相关。举用一个形象的比喻,科学数据就像是带有“干细胞性质的种子”,如果掉进荒漠很快就会干瘪死去,而把它种在沃土良田,却能开花结果。不同的“适合土壤”就会结出不同的“果实”。而这里“适合”与否,除了受科学数据的质

量和可理解性影响外,很大程度上取决于使用者的知识水平和创新能力。在对科学数据的利用中,常常出现“小人物做出大事情”的情况^[6]。而且在科学数据的共享应用过程中会衍生出更高层次需求的新数据,体现为科学数据的螺旋上升发展。因此,国家创建一种有助于全社会实现科学数据共享的环境至关重要。

从公共经济学的视角而言,科学数据具有公共产品的特性,首先科学数据具有消费非排他性,一部分人对于某一科学数据的使用不会影响另一部分人。因而科学数据一旦共享之后,不同领域的学者可以从各自的角度去挖掘各自所需知识;其次,科学数据使用减损程度较低,不同于有形的物质资源,科学数据具有近乎可以无限制复制的特点,不会因为满足某人、某时的需求而影响他人对它的需求。科学数据的这些特性决定了公开共享将使其效用价值倍增。

1.3 科学数据共享的法理基础

科学数据作为一种重要的科技资源,在法律上是作为一种财产来加以处理的。现代财产制度是通过财产权利制度来加以保障,包括物权制度和知识产权制度。产权制度实行合法所有,遵循谁投资谁受益、谁创造谁受益的原则,而法律保护的基本原则是保护财产所有人和创造人的合法权益。尽管科学数据能否在法律上被视为一种受财产制度约束的完整意义上的“财产”还需要深入研究。但是,由国家投资产生的科学数据既具有“财产”的特性,又具有公共利益的特征。因此,实施科学数据共享与现代法治社会对财产保护的制度并不矛盾。

从法理的视角来看,遵循“取之于民,用之于民”的原则,决定了科学数据拥有者必须让数据进入公众领域,服务于国家的整体发展。科学数据共享是政府信息公开的一项重要内容,是政府应当对社会公众承担的一项法定职责。通过政府的合理引导和有效监管,在确保国家安全和相关知识产权的前提下,促进科学数据相关信息公开共享,进而更加有效、更大范围地服务于全社会,是社会发展的必然。

2 美国 NSF 与 NIH 科学数据共享的比较分析

美国 1996 年修订的《信息自由法》(Freedom of Information Act)是美国政府资助项目数据公开共享的起点。1999 年联邦政府预算管理办公室(OMB)依照《信息自由法》修订“OMB 公告 A-110”,对政府资助科学研究提出了数据共享的要求。在国

际范围内,美国国立卫生研究院(NIH)是倡导项目数据共享并开展早期实践的组织之一。2003年2月NIH《研究数据共享的最终声明》中关于数据共享目标的陈述被2004年OECD引为《数据开放宣言》的“推动数据访问和共享”目标^[7]。OECD于2012年制定了《公共资助科学数据存取的原则和指南》,提出了开放、灵活、透明、遵守法律、保护知识产权、承担责任、专业、互操作、质量、安全、效率、问责、可持续等原则^[8]。可见,美国开展“数据共享”的成熟经验可以成为未来我国科学基金相关实践的借鉴依据。

在“项目申请和立项”阶段,NIH、NSF设置的项目管理机制分别适用于合同类项目和资助类项目。NIH提出《数据共享政策和实施指南》,一直将各项目类别中年度预算(直接成本)超过50万美元的项目申请单列出来,正式要求此类项目须公开共享项目数据,在项目申请时应根据数据共享政策的要求制定《数据共享计划》^[9]。NSF于2011年修订的《资助与管理指南》中明确提出了NSF对研究者公开共享项目数据的期望。要求凡是在2011年1月18日及其以后申请NSF资助的项目申请书必须附带一个两页的“数据管理计划”来说明数据管理的内容,详细描述如何遵守NSF的“传播与分享研究成果”政策,大致内容包括数据类型、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,而且这份数据管理计划同样接受同行评审。2011年12月,为了促使NSF完成其数据共享使命,促进科学研究和工程实践的发展,美国国家科学基金会理事会发布了《数字研究数据共享与管理》^[10],该文件提出了NSF资助而产生的数字科研数据的管理和共享的十大挑战和五项建议。

NSF的要求适用于任何规模、任何学科(部)的项目申请。虽然NSF要求的还仅是数据管理计划而不是数据共享计划,但是这些计划确实强烈鼓励共享而且也要接受同行评审。因此,NSF的要求远比NIH更为广泛,NIH资助项目的数据管理仅适用于大型资助,同时由研究者和项目管理者之间协商确定,不需要参加同行评审。系统比较NIH《数据共享计划》和NSF《数据管理计划》见表1。

NIH采用“要求数据公开共享”的方式,详细界定了数据共享的范围,并提出了关于数据共享可选方式、数据共享及时性、数据共享时长等方面的具体要求(见表2)。项目组申请NIH项目时,必须按要求提交《数据共享计划》。NSF采用“鼓励数据公开共享”的方式,没有提出项目“数据共享”的确切要求,而是通过“双向调适”方式确定项目的《数据管理计划》(见表2),并且NSF在项目申请评议时将《数据管理计划》将作为评价学术价值和广泛影响的依据之一,这会对申请者产生重要影响。

表 1 NIH《数据共享计划》和 NSF《数据管理计划》的内容比较

NIH《数据共享计划》	NSF《数据管理计划》
年度预算(直接成本)超过50万美元的项目 不需要参加同行评审	任何规模、任何学科部的项目申请 需要参加同行评审
1 计划数据共享的内容 2 所共享数据的使用群体 3 所共享数据的存储地点 4 数据公开共享的时机 5 研究者如何查询和访问数据	1 提出项目预期将共享的数据(可包括样本、实物资料等) 2 说明数据格式(包括数据和元数据的格式和标准) 3 提出关于“隐私信息、保密或安全信息,知识产权以及其他要求和权利保护”的数据访问和共享限制 4 提出数据再利用、传播和衍生使用的要求 5 提出数据、样本以及其他研究产出存档和保存的计划

表 2 NSF 与 NIH 数据共享做法的对比

NSF 数据共享的相关要求	NIH 数据共享的相关要求
“鼓励数据公开共享”(自主选择,但参加同行评审)	“要求数据公开共享”(大额资助项目强制要求共享)
1 项目申请者可界定申请项目中数据的涵义,并据此说明该项目将共享的数据	1 研究数据指用于证明研究发现的、科学界公认的真实数字化资料,不包括样本、实物资料、音频、视频等内容
2 项目申请者可界定项目的数据管理和访问方式	2 数据管理和访问的4类可选方式:①研究者自我提供数据共享;②提交数据库自由共享;③数据受限访问;④混合方式
3 项目申请者可说明项目数据的开放共享时机	3 研究发现公开发表后立即共享
4 项目申请者可提出项目数据存储和共享的期限	4 至4少保存和共享3年
5 项目申请者可界定项目关于保护知识产权或潜在商业价值的共享限制	5 除了小企业创新研究资助项目可在资助结束后保留数据专用权4年,NIH仅允许项目研究者延后公布数据30到60天

NIH 要求大额资助项目开展“数据公开共享”，其原因在于 NIH 认为“年度预算(直接成本)超过 50 万美元”的资助强度已经可以吸引项目组在要求“数据共享”的前提下申请此类项目。此外，NSF 和 NIH 都十分重视项目实施数据公开共享的成本。NSF 允许研究者在一定成本限度下提供合理期限的数据公开共享，并期望受资助单位为数据公开共享提供帮助；NIH 提供的 4 类数据管理和访问的可选方式(见表 2)中，允许研究者选择合适的共享渠道降低数据共享成本，例如研究者可以选择“自主提供数据”的方式简单共享少量的、访问量不大的数据，选择将数据“提交公共数据库”的共享方式实现“访问需求量/或数据量较大”的数据共享。

2014 年 8 月 27 日，NIH 发布了“基因组数据共享”(Genomic Data Sharing, GDS)政策。该项政策旨在促进基因组数据共享，并加快数据向知识、产品和流程的转化，同时保护相关研究参与者的隐私。NIH 要求自 2015 年 1 月 25 日开始对申报的项目申请执行 GDS 政策，这项政策适用于所有 NIH 资助的大规模人类和非人类的基因组科研项目，包括受 NIH 拨款和合同支持的研究，以及在 NIH 院内研究计划框架内进行的研究。这项政策的实施将成为推动 NIH 的医学发现的关键^[11]。

3 美国 NSF 与 NIH 针对科学数据共享可能面临困境的解决方案

联邦政府预算管理办公室(OMB)依照《信息自由法》修订“OMB 公告 A-110”，在要求政府资助项目共享数据之前，进行了两次公开的意见征集，以获知科学界对项目数据共享的意见。调查发现：科学界对项目数据公开共享存在的疑虑或反对意见主要集中于三方面：一是对开展数据共享所增加的项目成本的担心；二是对知识产权保护、参与研究人员隐私保护等方面的顾虑，三是对过早公布科学数据的担忧^[12]。针对上述意见，NIH 和 NSF 分别提供了相关的促进数据共享的做法。

首先，当项目组公开项目数据和存储数据提供长期共享的成本显著时，NIH 允许研究者申请新的资助。NSF 也提出可以为数据的清洗、存档、传播、存储等活动提供资金支持。

其次，受资助单位担心数据公开共享危害其获取知识产权。由于《拜杜法案》将成果的产权授予受资助单位，NIH 和 NSF 支持项目获得知识产权，以

激励受资助单位和研究者开放、应用和推广其研究发现。在数据共享方面，NSF 和 NIH 都根据“OMB 公告 A-110”要求，提出项目数据公开不包括“商业秘密、信息、未公布研究结论的相关数据、或类似的受法律保护的信息”^[13]。此外，NSF 还强调研究者可以在《数据共享计划》中界定数据商业化使用的条件，以保护具有潜在商业价值的项目数据。

为了隐私保护，NSF 和 NIH 都根据“OMB 公告 A-110”的要求，提出公开共享的项目数据可不包括人事、医疗信息等能够用于明确研究参与人员身份的信息，防止个人隐私被侵犯。而 NIH 由于主要资助医学研究，更加关注隐私保护，提出项目组可以在不影响数据用于验证研究结论和二次分析的前提下，利用统计方法调整或隐去部分数据信息；或在《数据共享计划》中界定“数据共享协议”(data sharing agreement)，明确数据访问的条件(criteria for data access)、数据研究使用的条件(conditions for research use)等要求。

最后，针对研究者担心过早公布数据会不利于研究者开展后续研究和研究成果发表。NIH 认同项目研究者对于项目数据(尤其是一手数据)的独占使用权(exclusive use of data)。由于不同项目研究的数据分析时长不同，NIH 不以“数据获得”时间为依据提出数据公开共享的时限要求，而是以“研究发现或成果的公开发布”为数据公开共享的前提条件。项目组可以选择延后发布研究发现或成果，以享有数据的独占使用权；NSF 也认同研究者对项目数据的优先使用权，但与 NIH 不同的是，NSF 允许项目申请者在《数据管理计划》中提出该项目认为合理的数据延期公开共享的时长，并通过项目立项评议审核确定。

4 国家自然科学基金在科学数据共享方面的实践进展

为了增加国家自然科学基金资助工作的透明度，促进基础研究学术信息资源的共享和利用，全面反映科学基金资助绩效，加强监督和道德学风建设，国家自然科学基金委于 2006 年 5 月 16 日正式开通国家自然科学基金资助项目信息共享服务网站简称“科学基金共享服务网”(http://npd.nsf.gov.cn)。在科学基金共享服务网开通之际，公布了一批结题项目的基本信息和取得的学术研究结果，包括公开发表的论文、公开出版的著作、会议论文及获得奖励

情况等。2012年2月,基金委开始启动科学基金共享服务网系统的重新建设以及数据更新工作,2012年进一步更新公布2008~2011年结题项目的基本信息和相关成果信息。新版服务网于2013年7月建设完成,利用先进技术和资源,针对用户需求,重新设计了一套资源整合和共享服务方案。新网站从资源整合利用和提供科研服务的角度,对网站的构建方式和理念进行了重新定义。根据不同的用户需求,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”等不同的系统功能模块^[14]。科学基金共享服务网从2013年开始逐步实现年度结题项目成果信息的发布,截至2015年11月3日,共收录国家自然科学基金结题项目116608个,项目成果1669928个,对13万多份论文进行了全文共享。

另外,关于科学基金研究型数据的共享,相关数据采集的主要实践集中于地球科学领域。由于地球科学的研究依赖于海量科学数据,通过多年的发展,科学基金构建了多个数据中心:1993年3月,基金委地球科学部指导成立“国家自然科学基金委员会地球科学部南京信息工程大气资料服务中心”(以下简称南京大气中心),负责接收、管理大气科学领域科学基金研究项目的重要资料并提供资料共享服务。2003年初,基金委委托国家海洋局第一海洋研究所筹建“青岛海洋科学资料共享服务中心”(以下简称青岛海洋中心),青岛海洋中心在基金委委托和授权下收集和管理科学基金海洋科学项目资料,建立数据和资料共享服务体系。2005年,为了在科学基金重大研究计划“中国西部环境与生态科学研究计划”(以下简称西部计划)内部共享西部环境与生态数据,基金委依托中国科学院寒区旱区环境与工程研究所建设“中国西部环境与生态科学数据中心(以下简称西部数据中心)”。

南京大气中心和青岛海洋中心主要从资助类项目中收集研究数据;西部中心主要服务于“西部计划”,在合同类项目内部共享数据。“西部数据中心”运转良好。经过“西部计划”数据采集、分析、集成和共享的多年实践,“西部数据中心”(http://westdc.westgis.ac.cn)逐步建成功能完善的数据共享平台,并更名为“寒区、旱区科学数据中心”,扩大了数据采集范围。

西部数据中心构建了相对完善的管理机制,明确了数据使用者参与的机制。(1)西部数据中心建

立了用户参与的数据评审制度,允许用户参与评审数据作者提供的元数据。西部数据中心认为元数据的质量会关系到用户在检索和使用数据时对数据集的理解程度,进而影响数据共享平台的数据发现功能。用户参与评审元数据时,重点从元数据信息的完整性角度审核元数据,以保障用户顺利使用数据;(2)西部数据中心还允许用户在数据使用后反馈对数据的评价。用户评价信息将为其他用户使用该数据提供更多的信息,包括数据的正确性、时效性、利用价值等信息。数据用户参与的元数据评审和数据使用后评价,为西部数据中心评价数据特征和价值、获知数据的使用需求提供了依据^[15]。

5 美国科学数据共享的经验借鉴

首先,美国政府首先从制定政策和法律入手。政府以原有的《信息自由法》和《版权法》为法律基础,规定政府应该主动告知禁止公开内容以外的信息。特别是NSF《数据管理计划》的要求适用于任何规模、任何学科部的项目申请,而且《数据管理计划》的内容将接受同行评审,因此研究者为了提高项目评审通过率,在项目申请时需要清晰地表达项目将产生何种数据、如何管理这些数据、通过何种方式共享这些数据、不共享的理由等。若无法清晰的阐述这些内容将会影响项目的资助。因此,NSF《数据管理计划》纳入同行评议的过程,增进了众多资助项目中的利益相关者之间关于数据共享的交流与沟通。因此,建议参照NSF的做法,适时在国家自然科学基金的申请书中增加“数据管理计划”的内容。

其次,保障投资者获得利益是美国科学数据共享机制制定的基本原则。在保障国家安全、政府政务和个人隐私的基础上,谁投资谁受益。在上述原则的指导下,美国在科学数据管理中,严格区分保密性管理机制、“完全与开放”管理机制和市场管理机制等3种不同的运行机制,而区分的主要标准和依据是投资来源和数据的产权性质。其中:对于有可能危及国家安全、有可能影响政府政务、有可能涉及个人隐私的数据和信息均纳入保密性运行机制中管理;对国家所有和国家投资产生的、不会危及国家安全、不会涉及个人隐私的全部数据和信息纳入“完全与开放”的运行机制管理;对由私营公司投资开发的数据,采取市场管理机制,并通过对开发证的批准、税收、反经济垄断等渠道来加强管理。这3种运行机制的优先顺序是保密性管理机制优先,此后是“完

表3 NSF的科学数据类型划分和相关特征

数据类型	数据使用	保存期限	数据标准	经费来源
研究型数据	仅供项目研究使用	项目实施和结束后一定期间内保存	数据一般只进行有限地管理和维护,不需要遵循特定的数据标准	经费支持源于项目拨款
资源型数据	服务于特定研究群体	数据中期或长期保存	数据格式依从该研究群体认可的统一数据规范	经费来自研究群体或机构的投入
参考型数据	服务于整个科学界	数据以永久保存为目的	服从公认、稳定、完善的数据标准	经费通过多元化的渠道获得

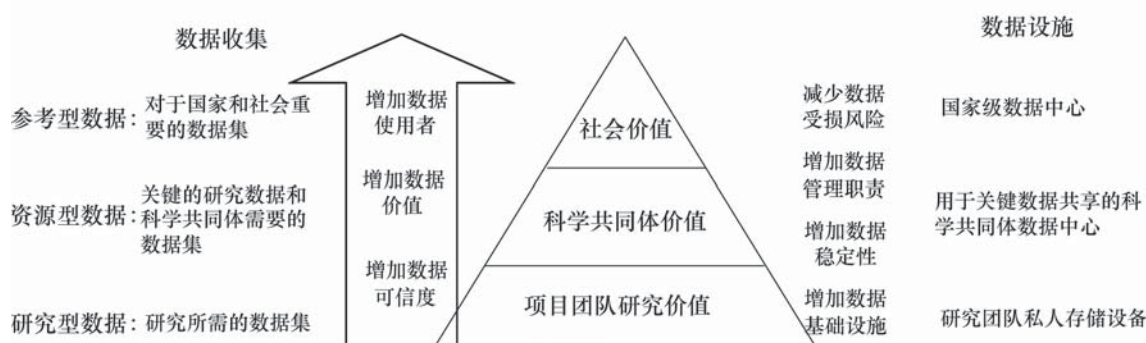


图1 科学数据的价值增值过程

全与开放”机制和市场机制并列。为了保障这套机制的顺利实施,国家还建立了配套的强制性、鼓励性和奖励机制^[3]。

再次,科学数据的分类管理和系统集成。2005年9月NSF发布了关于科学数据共享的研究报告《推动21世纪研究与教育的长期数据库》^[16],将科学数据分为研究型数据(research data)、资源型数据(resource data)、参考型数据(reference data),3种类型数据的特征比较见表3。Berman将从“研究型数据”到“资源型或参考型数据”的数据集成过程描述为科学数据增值的过程^[17](图1)。通过合理的科学数据分类管理和系统集成,有力地提升了科学数据的价值增值^[18]。

最后,为了消除研究者对于科学数据共享所引发的一系列担忧:诸如成本增加、隐私保护、知识产权保护以及对于后续科研成果发表的影响等,NSF和NIH通过成本补偿、产权保护等一系列措施,具体包括:美国NSF和NIH可以很好地弥补成果共享成本,可以促使项目组有动机投入数据获取、存储、公开、传播等活动;保护项目数据知识产权和隐私内容、满足项目组对数据的优先使用要求,能够消减项目组对于产出项目数据成果、提供数据公开共享的担忧。

6 讨论

科学数据共享是美国科技与综合国力领先战略

的重要保障,而我国国家投资科研项目产生数据的共享管理问题是制约我国科技发展的一个重要历史遗留问题。由于科技数据共享涉及数据开发者个人、数据开发者所在单位、管理与共享数据单位以及投资者等多重关系,在个人、单位和国家公众利益之间找到平衡点,是目前我国当前科学数据共享体系设计中亟需解决的关键问题。此外,为了保证国家投入产生的科学数据不流失,科学数据和规范管理,应该参照美国NSF和NIH的经验,在科学基金项目申请书中增加数据共享管理内容,包括数据汇集内容、时间、方式等,纳入项目立项同行评审,从而以合同的方式保证科研项目数据得到有序、有效、规范的管理,并构建完善的科学数据分类分级管理机制。以期建立健全符合客观规律和现阶段科学基金管理现状的科学数据共享机制,构建高效、有序的科学数据共享体系提供决策依据。

参 考 文 献

- [1] 杨敏 编译. 学术研究数据共享? 不共享? 中国社会科学报, 2012-06-13.
- [2] 黄鼎成. 科学数据共享的理论基础与共享机制. 中国基础科学, 2003, 5(2): 22—27.
- [3] Borgman CL (著), 青秀玲(译). 科学数据共享的挑战. 现代图书情报技术, 2013, (5): 1—20.
- [4] 程津培. 科学数据的共享管理创建共享新秩序. 中国基础科学, 2003, 5(1): 10—12.
- [5] 路鹏, 苗良田, 李志雄, 等. 我国科学数据共享现状的调查与分析. 地震, 2007, 27(3): 125—130.

- [6] 刘闯. 中国科学数据共享学术研讨会专辑. 中国基础科学, 2003, 5(1): 34—39.
- [7] National Institutes of Health. Final NIH Statement on Sharing Research Data. <http://grants.nih.gov/grants/guide/notice-files/NOTOD-03-032.html> [2015-05-20].
- [8] OECD. Principles and Guidelines for Access to Research Data from Public Funding [2012-10-21]. <http://www.oecd.org/science/sci-tech/38500813.pdf> [2015-06-20].
- [9] National Institutes of Health. NIH Data Sharing Policy and Implementation Guidance. https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm [2015-07-18].
- [10] Digital research data sharing and management. <http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf> [2015-07-15].
- [11] 宋凌霄. 美国国立卫生研究院发布基因组数据共享政策. 科研信息化技术与应用, 2014, 5(6): 95—96.
- [12] American Association for the Advancement of Science. AAAS Policy Brief: Access to Data. <http://www.aaas.org/spp/cstc/briefs/accesstodata/index.shtml> [2015-07-19].
- [13] NSF Education & Human Resources Directorate (EHR). Data Management for NSF EHR Directorate. <http://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf> [2015-07-19].
- [14] 王岩, 李建军, 王东鹏. 科学基金共享服务网的建设及思路. 中国科学基金, 2014, 28(3): 190—194.
- [15] 李新, 南卓铜, 吴立宗等. 中国西部环境与生态科学数据中心: 面向西部环境与生态科学的数据集成与共享. 地球科学进展, 2008, 23(6): 628—637.
- [16] Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf> [2015-06-25].
- [17] Berman F. Got data?: a guide to data preservation in the information age. Communications of the ACM, 2008, 51(12): 50—56.
- [18] 王东鹏. 科学基金资助与管理的关键问题研究. 合肥工业大学博士学位论文, 2014.12.

American scientific data sharing experience and its implications for NSFC: case study of NSF and NIH

Wang Jun

(School of Science and Technology of Hefei University of Technology, Hefei 230009)

Abstract In the era of big science, scientific data is a strategic resource of great value. Strengthening scientific data sharing is the important method to improve the efficiency of science and technology investment, promote the innovation ability, however, the scientific data sharing process is facing some difficulties. The article clearly defined the concept of scientific data sharing, and analyzed the internal logic mechanism of scientific data sharing from two aspects of economics and legal basis. Then we compared the concepts, content, methods and other differences of scientific data sharing between NSF and NIH. The dilemma of sharing scientific data and their solutions were deeply analyzed combined with the scientific data sharing practice of NSFC. At last we put forward the countermeasures and suggestions for improving scientific data sharing in China.

Key words scientific data sharing; National Science Foundation (NSF); National Institutes of Health (NIH); experience reference