

科学数据资源开放共享体系研究

马宁 刘召

(北京航空航天大学, 国家科技资源共享服务工程技术研究中心, 北京 100191)

摘要: 本文在阐述科学数据内涵与功能的基础上, 对美国地球系统科学数据开放共享作法和我国地球系统科学数据共享平台建设现状进行对比分析, 指出我国科学数据资源开放共享体系构建所具有的组织模式、政策制定、标准规范、主体特点、技术服务等特征, 进而从政策法规、标准与管理规范、统一门户、公益性与市场化相结合、提升关键技术、知识产权保护等方面探讨我国科学数据资源开放共享体系建设的推进策略。

关键词: 科学数据; 科学数据资源共享; 地球科学数据平台; 国家科技基础条件平台

中图分类号: F204; G311; G321 文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2017.03.001

Research on Open Sharing System of Scientific Data Resources

MA Ning, LIU Zhao

(National Engineering Research Center for S&T Resources Sharing Service, Beihang University, Beijing 100191)

Abstract: This paper compares with the open sharing of earth system science data in the U.S. and China based on the connotation and function of scientific data, then points out that the establishment of scientific data resources sharing system in China is characterized by its organization mode, policies, standards, main features and technical services. Finally, strategies for the construction of scientific data sharing system in China is discussed from the aspects of policies and regulations, standards and managements, unified portal, the combination of public welfare and marketization, the promotion of key technologies, and the protection of intellectual property rights.

Keywords: scientific data, scientific data resource sharing, earth science data platform, national science & technology infrastructure platform

科学数据既是科技创新活动中最基本的要素, 也是科技创新、经济发展和国家安全的重要战略资源, 对国家科技发展有重要影响。近15年来, 科学数据资源开放共享体系建设已经引起了我国政府的高度重视。2004年, 科技部和财政部启动国家科技基础条件平台建设, 其中“国家科学数据共享工程”被纳入国家科技基础条件平台建设7个重点领域。2005年, 通过了《科学数

据共享工程发展规划(2005—2010年)》, 在资源环境、农业、人口与健康、基础与前沿等领域开展了科学数据共享工作。

本文在阐述科学数据内涵与功能、对比中美两国地球系统科学数据共享工作现状的基础上, 研究探讨了我国自然科学领域科学数据资源开放共享宏观体系的构建特征与推进策略。

作者简介: 马宁(1965—), 女, 北京航空航天大学国家科技资源共享服务工程技术研究中心主任助理, 研究方向: 科技管理与技术(通讯作者); 刘召(1985—), 男, 北京航空航天大学国家科技资源共享服务工程技术研究中心项目经理, 研究方向: 科技管理。

收稿时间: 2017年2月27日。

1 科学数据及其开放共享内涵

本文所讨论的科学数据,是指科技活动中所产生的各类基础性、观测性的数据资料和相关信息,是科研成果最重要的信息载体,具有客观性、多样性、基础性、资源性、传递性、共享性、增值性等特点^[1]。科学数据是科技活动的成果,也是科技创新的重要对象与条件,具有明显的潜在价值和开发价值,是信息时代最基本、最活跃且影响面最宽的科技资源^[2]。

科学数据既是科研创新的源泉又是科研创新的成果。特别是大数据时代,越来越多的政策制定、科技创新、产业发展和社会进步依赖于对科学数据的分析利用。同时,科学数据虽然格式多样,但是便于认知理解,为相同领域学者交流提供了极大便利,有助于科研工作者以更低的成本获得创新资源,多角度全方位挖掘利用数据,激发数据潜在价值,开展交叉研究。此外,科学数据对开展协同创新有良好的促进作用。科学数据资源开放是指通过多种形式公开发布科学研究数据并为他人所用,其逻辑起点源于科学数据的资源属性。目前,科学数据主要的共享模式是国际数据交换、国家主导、企业或者私有数据共享3种^[3]。

2 中美两国地球系统科学数据的开放共享比较

地球科学数据是最具代表性的科学数据类型之一,也是国际上最早开展开放与共享的科学数据领域之一。本文力图通过对地球科学数据共享情况的分析,以期得到科学数据共享体系建设的基本情况。

科学数据的开放共享涉及顶层设计与制度保障、资源整合与服务、数据管理与技术支持、知识产权保护等方面。美国由海洋与大气管理局(NOAA)主导,经多年建设形成了完善的地球观测数据开放共享体制和机制。而我国地球系统科学数据共享平台是国家科技部认定的6个科学数据共享平台之一,自2002年起承担着国家地

球系统科学数据共享工作。现从以下几个方面对比中美两个机构的共享服务特点。

(1) 顶层设计与制度保障

为保证共享工作的顺利开展,NOAA十分重视顶层设计,在实践过程中结合领域特点制定了为数不少的规则条例。2010年发布了《下一代战略规划》为数据管理体系建设提出了战略构想,同年又发布了NOAA 212-15号行政令为数据管理提供决策和行动指南,并授权成立了环境数据管理委员会负责数据管理核心工作。2013年3月,NOAA又发布了数据管理顶层框架文件^[4],作为其进行数据管理的依据。相比之下,我国地球系统科学数据共享领域还没形成系统的顶层设计管理制度与规范。

(2) 数据整合与服务

NOAA的数据主要来源于《民用地球观测国家计划》相关项目的观测数据,汇集后的数据将通过data.gov的开放数据库向社会提供服务。我国地球系统科学数据共享平台主要以中科院地球系统的科研院所、长期野外监测台站的观测数据为主要来源,同时联合了相关机构科学家形成了一个科学数据共享联盟,并积极吸纳科研项目参与数据共享^[5],整合后的数据通过国家地球系统科学数据共享平台网站开展对外服务,同时相关资源元数据汇入国家科技基础条件平台中心主持建设的中国科技资源共享网,由中国科技资源共享网对外提供导航服务。中美两个机构同时也十分注重与国际数据组织的合作,积极吸纳国际数据资源。

(3) 数据管理与技术保障

NOAA要求数据开放共享应遵循完整及开放获取、长期保存、信息质量控制、易于使用4个原则,将数据管理过程划分为14个流程(图1)进行详细规范,主要包括:管理层制定观测要求文件,明确需要得到何种数据,指导地球观测系统的开发与部署;根据数据管理规划指南,对于所观测或获取的数据如何处理及保存提前规划;建立对数据性质、来源及质量等参数进行说明的元数据体系;要求开放共享的数据首先要根据文

档指南进行质量控制；将数据通过标准化的在线服务予以公开，便于其他用户查询及获取；根据数据存档程序，将数据或元数据存储在下设的3个国家数据中心；根据数据引用指南，由数据中心为档案数据分配永久标识符ID；由数据中心为存档数据提供与初始数据相兼容的查询及获取等服务；采用数据管理“仪表盘”自动统计数据库中的元数据记录，并统计相关使用情况；用户可利用共享数据开发其他产品，如气象预报、决策咨询等；用户可以通过ID对使用的数据进行引用，便于第三方对于数据的准确性加以验证，NOAA也可统计数据的最终用途；用户可对数据质量及其他特性提出反馈意见。

我国地球系统科学数据共享平台的共享工作主要是通过“地球系统科学数据共享服务网”（以下简称“地球共享网”）开展。目前，地球共享网通过两种方式管理数据资源：有服务能力的数据生产者只需汇交元数据，具体的数据还存储在本地，由本地服务器对外共享服务；无服务能力的数据生产者可将数据备份到地球共享网，由地球共享网代为提供数据共享服务。数据汇交采用“六位一体”的模式，包含元数据、数据缩略图、数据文档、数据样例、数据实体、关联分类等描述信息，保证数据资源的完整^[6]。地球共享

网分别按主体数据库、学科分类、数据格式、专题类型等对数据资源进行分类，用户可根据实际需求查询数据^[5]。地球共享网建设过程中还率先提出了地球系统科学数据共享标准的参考模型，研制了可扩展的地球系统科学数据编码、描述、集成与服务等关键标准^[7]。数据联盟内，采用统一的技术标准与软件，实现了数据发布、实时收割与一站式共享服务^[5]。地球共享网由一个总中心、认证中心和若干分中心构成。总中心和分中心分别维护着对方的相关信息。通过在分中心部署信息更新 Web 服务的方式实现了总中心与分中心数据的同步更新。采用了统一的用户注册和权限认证体系，能够为用户提供“一站式”的数据共享服务^[8]。

(4) 管理机构设置

NOAA的数据管理核心工作由环境数据管理委员会负责，该机构由NOAA观测系统委员会与NOAA首席信息管理办公室共同领导，并向NOAA管理层报告。环境数据管理委员会下设数据管理团队，分散在整个NOAA的所有执行项目中，负责数据管理的指导工作。此外，NOAA科学顾问委员会还下设数据获取与存档工作组，从专业角度为信息系统建设提出指导意见^[4]。

我国地球系统科学数据共享服务平台下设理

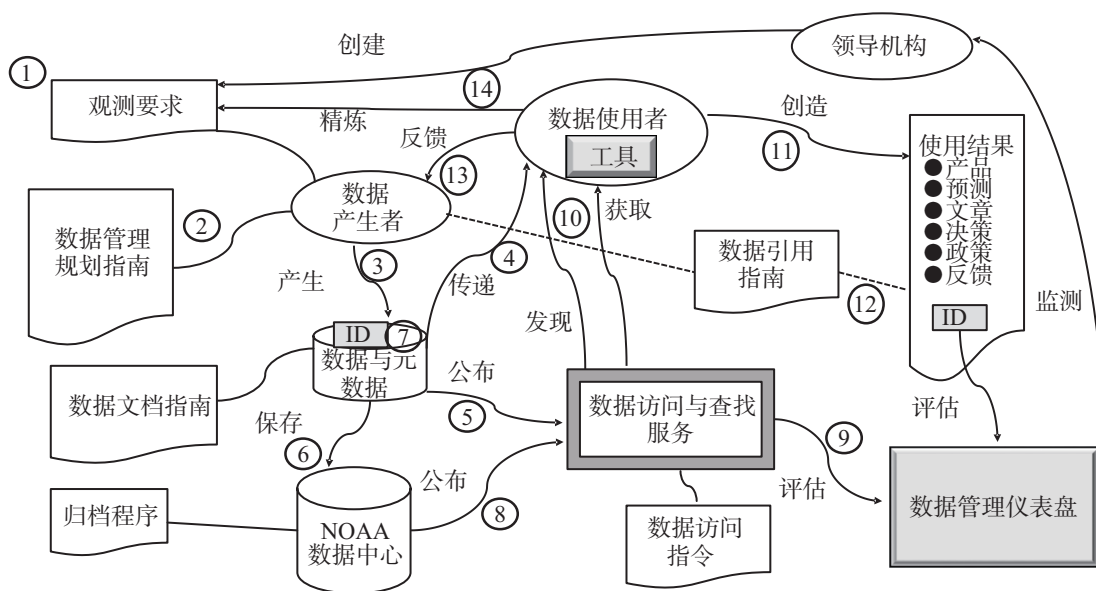


图1 NOAA数据管理流程框图

事会、办公室、业务组以及分平台数据中心，其核心机构——业务组负责平台发展的战略研究、标准规范制定、数据建设、软件研发以及用户服务，各分平台数据中心是总平台的数据提供和保存机构，按照总平台的标准和要求开展数据的整理、储存和共享工作。

(5) 知识产权保护

由于地球系统科学数据主要来源于国家财政投入项目，这些数据是典型的公共产品，因此无论NOAA还是我国地球系统科学数据共享服务平台对数据开放共享过程中的知识产权问题关注较少，只是在数据的溯源追踪、有序引用等方面提出了相关规定。

3 我国现有科学数据资源开放共享体系的构建特征

经过近十几年的规划发展，我国已初步建成三层架构的科学数据共享体系（图2），累计整

合农业、气象、地球、人口健康等领域71大类、超过1.6PB的科学数据^[9]，形成了以下特征。

(1) 国家主导的组织模式

我国的科学数据共享工作主要由国家主导推进。2001年我国启动科学数据共享工程，2003年又启动国家科技基础条件平台建设。科学数据共享是国家科技基础条件平台建设的重要内容之一。各地方政府也纷纷组织实施了地方科技平台，如首都科技条件平台、上海研发公共服务平台、黑龙江科技创新创业共享服务平台等。同时，各高校及科研院所也搭建了具有行业特色的共享服务网络，如中国科学院科学数据库、中国工程科技知识中心等。目前，在我国形成了科学数据共享多头并进、全面建设的局面。

(2) 专项性的政策法规

我国制定了许多法规政策来促进相关科学数据的开放共享，但是这些法规并不是以促进科学数据开放共享为主题。如《中华人民共和国科

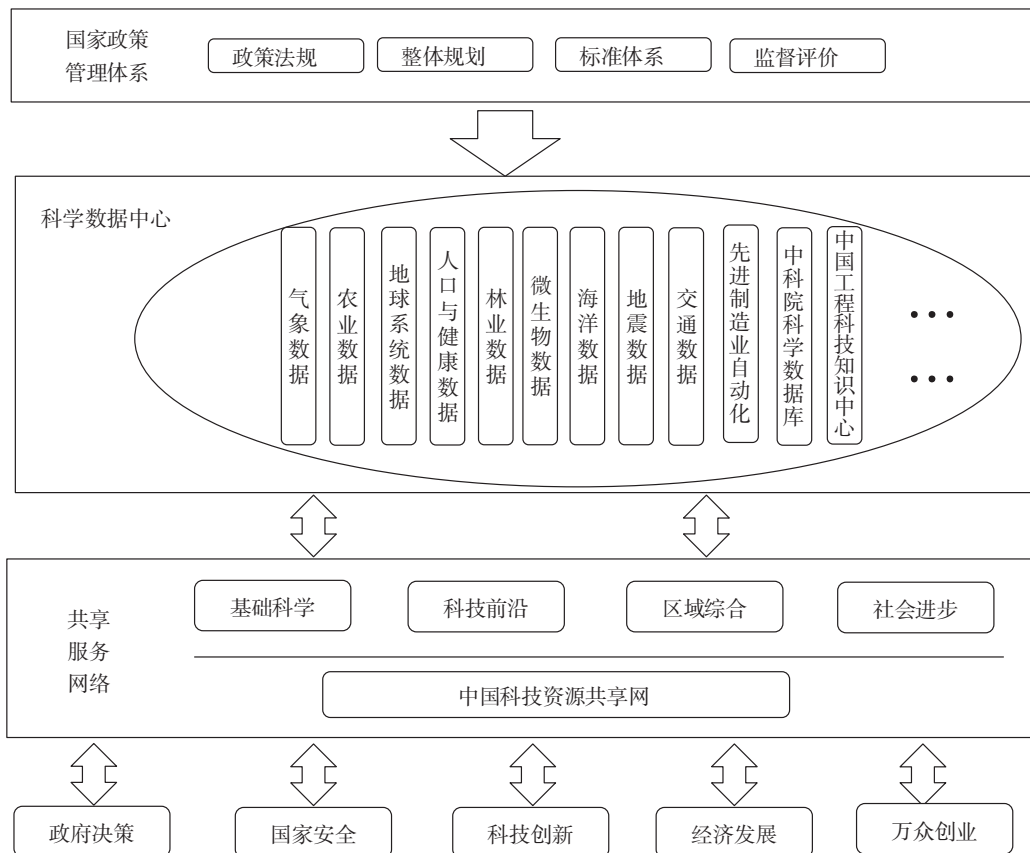


图2 科学数据共享三层体系架构

学技术进步法》虽然要求科学技术行政管理部门应当会同有关主管部门开展工作促进科学数据共享，但是该法规的主题是为了促进科学技术成果向现实生产力转化。许多公共科研资助部门也从自身需求出发制定了一些相关的政策规定，要求项目承担单位在立项之初就承诺研究成果的开放共享。如：《国家重点基础研究发展计划（973计划）管理办法》就明确建立规范、健全的项目科学数据和科技报告档案，并建立973计划项目和成果数据库；科技部《国家高技术研究发展计划（863计划）管理办法》、国家自然科学基金委员会《国家自然科学基金管理条例》等，对相关成果的开放共享也做了明确要求。但是长期以来的观念及知识产权保护缺位造成相当数量的单位和科技工作者将相关成果看作“私有”，共享积极性不高。

（3）行业性的标准规范

科学数据涉及领域众多、数据类型多样，怎么抽取其间的共性关系并制定标准，对推动整个科学数据共享工作有十分重要的意义。各科学数据共享项目及其建设单位在共享实践过程中逐渐认识到标准化工作的重要性，积极围绕科学数据的汇交、整理、保存、共享等环节开展了标准化工作。截至2014年底，科技平台标准体系已经初步建立，9项国家标准已经正式发布并实施，20个平台国家标准已立项^[10]。

虽然国内科学数据共享标准规范很多，但是各标准规范只适用于本行业或本单位。在核心元数据定义、数据汇交与保存等方面，虽然国家科技基础条件平台中心牵头制定并发布了一些通用性标准规范，部分国家科技平台也制定了一些行业标准规范，但是相关标准规范尚未在全国范围内得到普遍性推广应用。

（4）公益性的共享服务

我国科学数据开放共享多由政府推动开展，数据来源多是国家财政支持的项目产出，某种程度都带有公益性质，但不代表科学数据的共享必须是公益性的。国内科学工作者多是根据单位或项目要求被动参与共享工作，很难从共享中得

益，因而变相阻碍了共享的推动。如何调动科研机构、科技工作者的共享积极性成为推进科学数据共享进展必须要解决的问题。

（5）各具特色的服务技术和软件工具

为适应大数据时代的变化，科学数据在来源、方式手段、作用、理念和质量要求等方面都发生了质的变化，用户对科学数据质量和实效性的要求越来越高，如何提高数据的质量和时效性成为共享工作开展的重要问题。为此，各科技资源所属单位纷纷购置并开发各种软件系统，包括数据库、操作系统、应用系统、Web网站、软件工具等。但是很多单位的软件开发及服务经常采取外包方式，导致软件开发能力和科技资源管理水平参差不齐，极大地影响了共享服务的效率和水平，更不利于构建全国统一的科学数据共享平台。此外，各科学数据共享平台单独维护其数据资源，建设与服务成本很高，既造成了浪费，又与共享宗旨相悖。

4 结语与策略建议

科学数据作为最重要的科技资源类型之一，是信息时代最活跃的资源要素，只有在不断的共享流通中才能实现其价值增值，体现其经济、社会效益。自国家科技基础条件平台建设以来，我国科学数据资源共享服务取得了很大发展，然而在实践中还有很多的工作要做，迫切需要国家进一步完善顶层设计和法规制定，督促科技资源拥有单位边建设边共享、边总结边改进，不断扩展共享领域和范围。同时需进一步理顺共享服务链条和工作机制，加强相关标准、技术、工具的研发，提高各参与方对共享数据的整合效率和挖掘利用能力，以有效支撑我国科技创新和社会发展。

现针对我国科学数据共享工作现状及特征，为构建完善的全国科学数据资源开放共享体系提出如下推进策略建议。

（1）系统制定完善的政策法规

科学数据共享是一项综合性、全局性的工程，必须有国家的指导与调控。完善的法规体系

是国家引导、推动科学数据共享工作最直接的途径和方法。我国先后发布了《2004—2010年国家科技基础条件平台建设纲要》《国家科技计划项目科学数据汇交暂行办法(草案)》《国家重点基础发展计划资源环境领域项目数据库汇交暂行办法》等,对重大科研项目的科学数据汇交、整理与共享工作进行了规定。科技部于2002年开展了“科学数据共享政策法规体系框架的研究”项目,形成了《中华人民共和国科学数据共享条例(建议稿)》。但是与美国等数据共享发达的国家相比,我国的相关法规制定还相对滞后,体系也不够健全。因此,很有必要制定系统性、普适性、国家层面的政策法规,对科学数据的信息自由与公开、基础设施、技术平台、数据保护、数据汇交、数据开放利用、网络等进行指导^[11]。同时,通过相关法律的制定营造科学数据共享氛围,帮助各科研机构和科研人员树立科学数据共享的理念。

在我国,科学数据共享还是一个相对较新的理念,具体操作过程中会遇到很多困难与挫折,需要制定与法规配套的实施细则,给实施单位更多的指导,以保证相关工作的顺利开展。

(2) 制定共性的管理规范与技术标准

在大数据时代,如何实现数据汇交、整理、加工、利用的快速无缝对接,同时尽可能扩大数据整合领域和范围成为推动共享工作必须要解决的问题。标准化是开展相关工作的基础和前提。通过制定统一规范的标准,将提高我国科学数据共享工作的科学化、合理化和工程化,降低不必要浪费,提高共享效率。

此外,由于科学数据的特性,其共享多是通过建设数据库、利用网络进行的。只有对建库标准进行科学的规定并在全国普及,才能有效扩大科学数据的整合效率。同时,科学数据的整合过程中只有运用先进的计算机网络技术、软件工具和数据挖掘技术才能实现对数据价值的深度利用。因此,很有必要制定相关技术标准、开发统一软件系统。

(3) 搭建统一服务平台门户

科学数据广泛地分布在不同的机构中,如何快速寻找或发现所需的资源是打通共享服务的重要环节。因此,很有必要建设一个类似美国 Science.gov 的全局性统一服务平台门户,对这些汇集的科学数据进行导航检索服务。中国科技资源共享网作为国家科技基础条件平台共享服务的门户网站,经过近几年的建设,目前已整合了23家国家科技平台的核心元数据资源,基本满足了用户“一站式”查找科学数据的需求^[12]。但是,用户对科学数据的需求复杂多样,精度有高有低,目前已有的资源信息远远不能满足需要,因此很有必要进一步扩大资源整合的领域,打通中国科技资源共享网与国家科技平台、地方平台、行业平台数据库的连接,在全国形成统一门户、多个支撑中心的合理布局。国家级门户致力于扩大整合范围,提供统一数据管理标准与导航,传播共享理念;各支撑中心专注于提高资源整合深度和数据挖掘精度,并加强与国家级门户的互通互联,及时将数据更新信息进行备份、传递。

(4) 探索公益性与市场化相结合的道路

目前,我国科学数据开放共享工作主要是国家主导下的公益行为,因为这些科学数据大多是国家财政投入产出的成果。但是很多研究机构、企业和科研人员也积累了大量的非财政投入产出的科学数据,这些“私有”数据总量巨大,同样蕴含了极大的生产力。然而,由于多种原因,这些机构、企业和个人的数据共享意愿较低,形成了新的“信息孤岛”,而大数据时代数据更新很快,数据一旦沉淀就很容易过时,造成资源浪费。如何推动“私有”数据的共享成为当前科学数据共享工作中不得不思考的问题。美国政府也认识到“私有”数据的极大潜力,正大力推进公私合作的数据创新生态系统建设,在开源程序平台启动了在线开放数据工程^[13],并取得了一定的效果。

科学数据作为重要的生产力要素,其经济价值、社会价值很有必要通过市场化进行体现。在我国现有条件下,借鉴美国等在科学数据开放共享市场化方面的经验,探索科技资源公益性与市

市场化相结合的道路就显得格外必要。通过市场化服务的探索，鼓励各单位将科学数据转化为经济效益，必将调动其共享积极性。对于公益性的服务，国家可以将共享服务效果纳入对单位的项目申请和评定考核的指标中，并对共享效果突出的单位给予一定的补贴奖励。

(5) 提升开放共享各环节关键技术

大数据、大科学时代，科学数据呈现海量化、复杂化的特性，这些数据的质量控制、整合管理、开放服务和挖掘利用需要强大的技术支撑。构建我国科学数据共享体系的重要环节之一就是开发一系列的关键应用技术，特别是海量数据检索技术、挖掘技术、展示技术、质量管理技术等。在国家科技基础条件平台中心指导下，国家科技资源共享服务工程技术研究中心正在加强各环节关键技术的研发，采用最新共享服务架构、快速数据检索技术、海量数据挖掘技术和在线服务技术，以“中国科技资源共享网”为载体，在一定程度上提高了我国科学数据共享服务成效。

建议今后将借鉴美国等成功经验，联合国内相关领军企业，开发更具创新性的统一数据共享技术和数据分享架构，为科学数据共享服务提供坚实的技术保障。

(6) 加强开放共享过程中的知识产权保护

科学数据的开放共享和知识产权保护是相互统一而又相互对立的矛盾统一体^[13]，必要的知识产权保护有助于保护数据生产者的合法利益，过多的知识产权保护又与开放共享精神相悖。

首先，建议国家出台相关法规条例进行规范管理，对不同性质的科学数据区别对待。由财政投入产生的科学数据要突出其公益性，对于科技工作者在科研活动中产生的科学数据和二次加工汇编形成的数据集（库）要加强知识产权保护。

其次，探索建立科学数据开放授权许可机制。国际常用的知识共享许可CC和开放数据公用许可ODC都能提供弹性化的许可方式，在保留部分知识产权同时，鼓励第三方用户最大限度地自由使用科学数据。我国可根据纳入国家共

享体系的科学数据特点与实际情况，借鉴国际通行做法，建立科学数据开放授权许可制度与授权模式。

最后，完善与提高科学数据知识产权保护与共享许可的技术措施。充分利用计算机与网络发展最新成果，在保护与共享许可的不同环节合理使用防火墙、数据与软件加密、数字水印、认证、访问控制等技术，实现科学数据分层级的保护与共享。

参考文献

- [1] 孙九林, 黄鼎成, 李晓波. 我国科技数据管理和共享服务的新进展[J]. 世界科技研究与发展, 2002, 24(5): 15-19.
- [2] 张红. 我国科技资源共享的现状及其分析[J]. 科技与法律, 2007(2): 18-24.
- [3] 左建安, 陈雅. 大数据时代的科学数据共享模式研究[J]. 新世纪图书馆, 2014(3): 32-35.
- [4] NOAA. NOAA environmental data management(EDM) framework[EB/OL].[2017-01-05]. <https://nosc.noaa.gov/EDMC/framework.php>.
- [5] 刘润达, 诸云强. 科学数据共享关键问题探索: 以地球系统科学数据共享网为例[J]. 地理科学进展, 2007, 26(5): 118-126.
- [6] 苗茹, 宋佳, 王卷乐, 等. 地球科学数据共享平台建设思路与实践[J]. 中国科技资源导刊, 2014(4): 64-72. DOI:10.3772/j.issn.1674-1544.2014.04.011.
- [7] 孙九林, 诸云强, 闫国年, 等. 地球系统科学数据共享国家平台构建、关键技术与应用服务[J]. 中国科技成果, 2015(2): 70-72.
- [8] 诸云强, 刘润达, 冯敏, 等. 分布式地球系统科学数据共享平台研究[J]. 计算机工程与应用, 2009, 45(1): 245-248.
- [9] 科技部. 科技部工作[EB/OL].[2017-02-23]. <http://www.most.gov.cn/xinwzx/xwzx/twzb/fbh17022202/index.htm>.
- [10] 本刊通讯员. 2015年全国科技平台标准化技术委员会工作会议在京召开[J]. 内江科技, 2015(2): 66.
- [11] 刘润达, 彭洁. 我国科学数据共享政策法规建设现状与展望[J]. 科技管理研究, 2010, 30(13): 40-43.
- [12] 国家科技基础条件平台中心. 中国科技资源共享网[EB/OL].[2017-01-18]. <http://www.escience.gov.cn/>.
- [13] 翁扬水, 李丽亚. 科学数据共享中知识产权保护问题探析[C]// 先进制造与数据共享国际研讨会, 2007.