

# 科学数据共享大循环

□ 中国工程院院士 孙九林

当前,经济发展的全球化和全球化科技活动不断增强,导致全球范围内对科学数据信息资源的交流、互通和深度使用的强烈需求和高度信赖。各国在迈向信息社会的进程中,由于国情不同、认识不同等原因,所实施的政策和措施不尽相同,从而产生不同的效果。

## 一、美国的数据共享“大循环”做法

如何让越来越多的数据信息资源在全社会流动起来,最大限度地发挥数据信息作为资源的作用,同时规范信息数据在管理和社会流动中的行为,是科学数据共享的核心。美国之所以在科技领域占主导地位,是与其所采取的“大循环”的科学数据共享做法分不开的。

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。

20世纪90年代提出的“完全与开放”的数据共享政策,成为美国联邦政府在信息时代的一项基本国策。从1990年到1999年的十年中,分两步实施数据共享计划。首先由美国航空航天局启动“分布式的最活跃数据档案中心群”,重点是搞好数据共享和国家级数据中心的建设;第二步,由白宫启动总统长期专项,建设国家级数据

信息共享网络,用十年的时间,即在20世纪末完成建立世界上最强大的科学数据共享体系的任务。

“完全与开放”的数据共享政策极大地刺激了美国经济的发展,数据共享实施的十年间,1995—1999年平均每年经济增长率比1991—1995年多增长1.1个百分点。据美国经济学家计算,其中0.5个百分点是由于数据和信息的传输和应用所产生的效益。

美国是一个市场经济起主导作用的国家,最终在信息时代核心的数据信息开发利用上,却选择了一条与市场方式相“背离”的“完全、开放与无偿共享”的道路。这是因为他们把数据信息资源的特征研究的很透,并找到了开发利用最好的途径。

让每一位公民在数据“信息”知识“理论”决策“效益”的各个环节上发挥才能,把数据信息流动和应用过程中的各种价值充分挖掘出来;国家则创造好环境并做好服务,利用税收制度收回资金,达到国富民强,这就是美国联邦政府选择的数据信息共享的“大循环”道路。

## 二、我们的数据共享经验

在国家科学数据共享工程和国家科技基础条件平台的牵引和领导下,我们于2002年率先建立的地球系统科学数据共享网(现为

国家地球系统科学数据共享平台)为实现中国科学数据的全面共享积累了宝贵经验。

地球系统科学数据共享的对象主要是分散的科学研究数据,没有成熟的经验可供借鉴。我们在建设过程中,坚持边建设、边服务、边探索、边积累。经过多年的实践,获得了13条初步的认识和经验:(1)制定平台发展顶层规划,(2)主持和参加单位首先将拥有的数据提供出来共享,(3)形成固定的队伍,(4)运用多种方式为用户提供服务,(5)多渠道有效地整合数据,(6)开创并成功实施“边建设、边服务”策略,(7)主动服务,跟踪重大科技项目,(8)提高服务质量,加强平台间互动,(9)集中开发共享软件平台,建立统一管理的分布式服务体系,(10)积极开展国际合作,(11)建立数据共享机制,(12)兼顾各方利益和国家安全,(13)吸引全社会关注,积极投入。

地球系统科学数据共享平台采用的分散数据资源的整合和服务模式得到主管部门和科技界的高度认可。2007年,经过反复论证,科技部正式以此为基础,选定973计划资源环境领域,开展国家科技计划项目数据汇交试点。2008年,科技部正式在全国范围启动了973计划资源环境领域项目数据汇交工作。这是落实国家“科技进步法”的重要举措之一,是平台对分散科学数据资源整合和共享



服务成功经验的重要产出,对于推动科技项目数据汇交机制的建设,完善共享环境;提高国家在数据资源管理与共享上的投入产出比;提升我国科技创新支撑能力和科研项目的管理水平等,都具有重要的意义并将产生深远的影响。

# 开放科学数据 实现共同发展

## ——中国科学数据共享之路

□ 杨兰

导语:科学数据是人类社会在科技活动中产生的各类数据资料,是科技创新的基础,也是经济社会发展决策的重要科学依据。人类的知识,无不是来源于所掌握的数据和信息的分析,尤其是在“大数据”时代的今天,科学研究、政府决策、产业发展越来越依赖于科学数据。因此,科学数据的开放、共享非常重要,特别是在人类社会投资力度不断加大,获取科学数据的能力飞速发展,而数据孤岛却快速增长下,科学数据的共享显得尤为迫切。作为科研大国向科研强国迈进的过程中,十年的科学数据共享努力,中国的科学数据共享到底如何?

## 科学数据共享的价值与难题

半个多世纪以来,通过国家各级、各地政府和部门和公益型事业单位开展的有组织、大规模考察、调查、勘探、监测、试验、以及各类科学工程、科研项目的实施,我国已经积累了海量的科学数据。进入21世纪后,科学数据更是以指数级的速度在增长,我们已经进入“数据洪流”“大数据”的时代。

尽管我国已经拥有丰富的科学数据资源,但数据共享问题仍然十分突出。诸多数据库往往局限于本部门、本单位使用,甚至小群体专用,缺乏部门间的交流和沟通,更没有形成面向社会共享的局面;国家各类科技计划项目所产生的科学数据,也没能得到有效的管理和利用。科学数据共享问题已经造成了重复建设和资源的巨大浪费,在相当程度上制约了科技创新,乃至国家整体科技水平的提高。

孙枢院士专门撰文呼吁:“科学数据共享可以使科研成果得到最广泛的利用和传播,科学家和社会公众可用最低廉的费用获得数据。充分利用他人已有的研究成果,避免不必要的重复,使科研成果得到有效利用”。“国际科技界花了100年的时间认清了科学数据的战略地位。我国应该仔细学习这段历史,一步到位给予科学数据应有的战略地位,这是科技长期可持续发展的重要保障,也是我国参与国际科技竞争的重要支撑条件”,孙鹤烈院士在谈到科学数据共享时表示。秦大河院士长期推动科学数据共享,认为:“科学数据具有非排他性、无消耗性和可增值性。科学数据共享,使科学数据更加开放、交换更加顺畅、应用更加便利”。国际科学数据委员会(CODATA)主席郭华东院士指出:“数据与信息的共享是实现不同学科、领域间合作的核心”。孙九林院士指出:“科学数据的科学、社会、经济价值只有在数据流动和共享中才能充分体现”。

然而,科学数据共享涉及到政策法规、标准规范、数据资源、技术系统、运行服务等诸多因素,是一个开放的、复杂的系统,尤其是涉及到部门利益、理念意识等,实施起来非常困难。“真正实现科学数据共享,要切实改变传统的科学数据占有观念,打破信息壁垒,实现科学数据的合理流通。”科技部原部长徐冠华院士说:“在确保国家安全和相关知识产权的前提下,大力促进信息公开,实施最大限度的科学数据共享,进而更加有效、更大范围地服务于社会和广大民众,是社会发展的必然”。科技部原副部长程津培院士2004年在科学数据保护与共享战略国际研讨会上指出:“科学数据共享要做好五个方面的工作:转变观念、提高认识。统筹规划,建立健全科学数据共享的法律、法规体系和标准规范体系。整合存量,构建国家科学数据共享服务体系”。

除政策法规、部门利益外,面向海量、分布、多源、异构的科学数据,如何从机制上保障这些数据资源的持续集成与共享,如何从标准规范、技术方法上有机整合集成这些数据资源,为用户提供“一站式”的数据服务都是迫切需要研究和解决的问题。

## 从试点到平台,中国科学数据共享的发展 启动试点工程,谋划中国科学数据共享

我国科学家一直呼吁科学数据的共享。早在1994年,中国科学院地学部徐冠华、孙枢、周秀骥等院士就提出了加强科学数据共享的建议。1999年,科技部在科技基础性工作和社会公益性专项中,启动了科技基础数据库建设。2001年,科技部成立“科学数据共享调研组”,在黄鼎成研究员的主持下,完成了“基础性工作及国家政策与管理研究”等调研报告,提出了“实施科学数据共享工程,增强国家科技创新能力”的建议。

2001年底,“气象科学数据共享试点”启动。2002年6月,科技部等5部委联合下发《关于进一步增强原始性创新能力的意见》文件,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”。2002年11月,“中国科学数据共享”第196次香山科学会议召开,徐冠华、孙鹤烈、孙枢、程津培、秦大河院士和张先恩研究员担任会议执行主席,与会专家呼吁:“应加强国家对科学数据共享的统筹规划与协调,设立国家科学数据共享工程专项,推动我国科学数据共享问题的彻底解决”。同年,继气象数据共享后,又启动了地球系统、水文、海洋、地震、国土、农业、林业、人口健康数据共享,作为国家科学数据共享工程第一批的9个试点,标志着我国科学数据共享工程的正式



实施,掀开了中国科学数据共享事业新的一页!

在科技部发布的“科学数据共享工程规划”中明确了科学数据的战略资源地位及其共享管理的必然选择,确定了我国科学数据共享的建设目标、总体框架、主要任务与实施方案。通过试点工作探索了分散科学数据共享机制,基本解决了科学数据共享共性的标准规范、技术方法和软件系统等问题。通过试点项目,推动了科学数据共享在资源环境、农业、人口与健康、基础与前沿等领域共24个部门的开展,包括:气象、测绘、地震、水文水资源、农业、林业、海洋、国土资源、地质与矿产、对地观测等行业领域国家科学数据共享中心和地球系统、人口健康、基础科学、先进制造与自动化科学、能源和交通等学科领域的科学数据共享网,为科学数据共享的全面建设奠定了扎实的基础。

## 全面建设科技条件平台,夯实数据共享服务基础

为了进一步推动科学数据等科技资源的共享,2004年7月,国务院办公厅发布了《2004—2010年国家科技基础条件平台建设纲要》。2005年7月,科技部、发改委、财政部、教育部联合发布了《“十一五”国家科技基础条件平台建设实施意见》,同年,科技部、财政部设立平台建设专项,科学数据共享纳入国家科技基础条件平台,标志着科学数据共享等进入了全面建设阶段。

2006年12月,国家科技基础条件平台中心(以下简称“平台中心”)挂牌成立,专门推动和管理各类科技资源的优化配置和开放共享。国家科技基础条件平台重点建设了六类资源的43项共享平台:一是研究实验基础和大型科学仪器设备共享平台,包括全国大型科学仪器设备、研究实验基地、野外科学观测研究站和计量标准体系及检测技术体系;二是自然科技资源共享平台,包括植物种质资源、动物种质资源、微生物种质资源、人类遗传资源、生物标本资源、岩矿化石标本资源、实验材料资源和标准物质资源;三是科学数据共享平台,包括科学数据共享中心和科学数据共享网;四是科技文献共享平台,包括科技图书文献信息保障系统、专利文献和标准文献共享服务系统;五是科技成果转化公共服务平台,包括科技成果转化信息平台、网络科技环境平台,包括国家条件平台应用服务支撑系统、网络计算应用系统、网络协同研究工作环境、全国科普数字博物馆和全国科技信息资源网。

科学数据共享,在国家科技基础条件平台的推动和支持下,得到了进一步的发展和夯实。

## 创新“定、评、补”机制,护航中国科学数据共享

作为一项利国利民的基础性工作,国家科技基础条件平台必须有稳定运行的长效机制。这是平台主管部门、承建单位和科技界的一致共识。为此,2011年科技部、财政部联合下发了《关于开展国家

科技基础条件平台认定和绩效考核工作的通知》,向全社会发布了“国家科技基础条件平台认定指标”和“运行服务绩效考核指标”,形成了国家科技平台“定(平台认定)、评(绩效评价)、补(奖励补助)”的长效机制。

依据“定、评、补”机制,2011年11月,首批23家国家科技平台通过认定,正式进入运行服务阶段,包括科学数据共享领域的6家平台:地球系统科学数据共享平台、气象科学数据共享中心、地震科学数据共享中心、农业科学数据共享中心、林业科学数据平台和人口与健康科学数据共享平台。

2013年11月,财政部、科技部又联合下发了《国家科技计划及专项资金后补助管理规定》,正式将国家科技基础条件平台列入共享服务后补助财政资助方式,规定:国家科技基础条件平台经科技部、财政部绩效考核通过后,给予分类分档定额经费补助,主要用于平台的运行服务,标志着国家科技基础条件平台“定、评、补”制度的完善,为我国科技资源服务的持续发展提供了政策上的保障。

平台中心主任戴国强表示:“科技部、财政部制定并实施科技平台绩效考核和共享服务后补助机制以来,按照‘以用为主,重在服务’的要求,各家科技平台大力开展科技资源共享服务工作,服务数量和质量均有大幅提升,有力支撑了科技创新和经济社会发展”。平台中心运行监督处处长认为:“定、评、补”制度的实施,有效调动了各国家科技平台的积极性,推动了平台自身的建设,极大促进了平台运行服务水平的提升”。

## 地球系统科学数据共享平台,十年见证与实践

作为科学数据共享的首批试点和国家首批认定的23个科技条件平台之一,地球系统科学数据共享平台已经走过了10个年头。十年来,地球系统科学数据共享平台通过自己的实践,伴随、见证着我国科技资源共享的成长,推动、引领着我国研究型科学数据的共享。

“与行业部门科学数据共享平台不同,我们专注于研究型科学数据的共享,即分散在科研机构、高等院校以及科学家手中的各类科学研究活动产生的数据,致力于地球系统科学数据和全球变化创新研究及区域可持续发展决策提供科学数据支撑”,地球系统科学数据共享平台负责人孙九林院士表示,“地球系统科学和全球变化创新研究不仅需要长时序列表单一的自然和人文要素数据的支持,更需要集成性综合数据产品的支持”。

地球系统科学数据共享平台相比于行业部门的科学数据,具有涉及学科多、分散性大、来源广泛、类型多样、标准程度低、时空特征明显等特点,因此,它的共享尤其复杂和困难。为此,地球系统科学数据共享平台组织国内外40多家单位400多名科技工作者,按照试点探索(2003—2005年)、全面建设(2006—2008年)、运行服务(2009年—)三大阶段,系统开展了研究型分散科学数据共享的研究,建成了可业务化运行的地球系统科学数据共享平台,长期坚守、持续提供无偿的数据共享服务。十年发展,地球系统科学数据共享平台取得的主

要成效和启示:

**探索形成分散的研究型科学数据共享机制,推动国家重大科技计划项目数据汇交。**研究型科学数据共享缺乏类似于行业领域数据共享的部门政策法规,因此,首先要解决共享机制问题。十年来,该平台倡导“地球系统科学数据联盟”理念,探索形成了面向数据单位“共建、共享”,面向用户单位“先服务、后集成”的分散科学数据共享的机制,建立了“资金引导、规范整合、服务带动、吸引数据,技术互连、数据互通,联盟合作、分享科技”四种类型十种模式的实施措施。提出了分阶段、分类型的科技计划项目数据汇交体系,支撑“973计划资源环境领域项目数据汇交”的率先开展,为我国重大科技计划项目数据汇交工作的全面推进提供了重要的示范和借鉴作用。截止至2013年底,所有973计划资源环境领域项目都参与了数据汇交工作,汇交数据集3500多个,数据量达到3个TB,并积极对外开放这些数据。

**构建共享管理与标准体系,支撑不同学科和类型的地球系统科学数据的规范化集成和服务。**系统研究和设计了地球系统科学数据共享管理与标准体系。运行管理体系包括:联盟章程、共享条例、管理办法、绩效考核细则等,技术标准体系包括:数据分类、描述、集成、建库、分发、服务和质量控制等。编制了“地球系统科学数据分类编码”“地球系统科学数据核心元数据标准”等23项管理与标准规范,形成了多项国家标准。

**发展异构数据集成方法,开展数据整合与交换,建设全国和典型区域地球系统科学数据库。**针对多源、多尺度、异构地球系统科学数据集成问题,发展了高精度的定位观测和统计数据空间化及误差评价方法、数据同化方法和模型动态集成方法。通过“抢救性整理、规范化整理、挖掘再加工”等方式,建成了以陆地表层系统数据为核心,涉及地球系统5大圈层18个学科的地球系统科学数据库和极地、青藏高原、黄土高原、西南山地、东北平原、黄河中下游、长江三角洲等典型区域专题数据库。成为我国规模最大、学科面最广的研究型地球系统科学数据库。

**攻克分布式科学数据共享关键技术,研制自主知识产权的分布式科学数据共享软件系统。**研究解决了多标元数据统一存储管理、优化检索、多源地球系统数据统一访问、数据共享质量评价与控制等系列产品、地球系统科学数据共享系统。自主研发了全服务化的跨平台部署、二次开发和个性化定制分布式科学数据共享软件,为数据共享系统的快速构建奠定了软件基础。软件系统除在自身平台中应用外,还推广应用到国家人口与健康科学数据共享平台临床医学中心、国家环境档案信息资源共享系统、陕西省科学数据共享平台、宁波市生态环境数据共享平台、973项目“中国主要水蚀区土壤侵蚀过程与调控研究”数据共享系统等。

**构建一站式数据共享服务网络体系,完善和创新服务体系,提供持续的数据服务。**建成了由1个总中心、6个学科分中心(全球变化模拟、湖泊一流域、冰川冻土、地球物理、空间科学和天文科学)和9个区域分中心(极地、青藏高原、新疆与中亚、黄土高原、西南山地、东北平原、黄河中下游、长江三角洲和南海及邻近海区)构成的物理上分布、逻辑上统一的地球系统科学数据共享网络体系。建立了基于网络在线数据共享服务和离线专业队伍服务共同构成的,以数据内容服务为主,技术服务、软件服务和咨询服务相结合的数据共享服务体系。领导创建了东北亚合作研究网络,与国际联盟世界数据系统、兴都库什—喜马拉雅山空间信息共享网、美国全球变化主目录、973计划资源环境领域项目数据汇交管理中心等建立了“国际—国家—部门”3个层次的国际数据交换网络。为国内外各类用户提供了持续、无偿的数据共享服务。

到2013年底,地球系统科学数据共享平台为8万多名专业用户,1700多项重大科研项目,30多项重大建设工程,30多项重大民生工程等提供了有效的数据支撑服务,支撑发表论文和硕博学位论文2000多篇,产生了显著的社会经济效益。

根据中国科学技术信息研究所2012年的第三方评价:地球系统科学数据共享平台在全国84个可评测的科学数据共享网站中综合排名第1。地球系统科学数据共享平台工作也得到主管部门和科技界的高度肯定:2011年获得“十一五”国家科技计划执行优秀团队奖,入选参加“十一五”国家重大科技成就展,并获2013年“河南省科技进步一等奖”。

**结语:**党的十八届三中全会提出要全面深化科技体制改革,明确国家重大科研基础设施依照规定应该开放的一律对社会开放。作为公共服务重要载体和科技资源开放共享重要抓手的国家科技基础条件平台,将面临更高的要求 and 更紧迫的形势。只有全社会都理解、支持科技资源共享,每个产生科学数据、使用科学数据的人都行动起来,才能真正实现完全开放的数据共享,让科学数据的价值得到最大体现,实现科技界多年科技资源共享的梦想,更好地服务于我国的科技创新和经济社会发展!