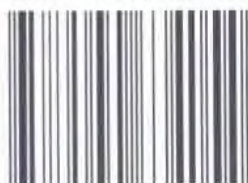


KENETSHUJI



ISBN 7-5046-3405-0



9 787504 634054 >

73.967
217

73.867
217

科学数据管理与共享

孙九林 施慧中 主编



2003年1月18日

中国科学技术出版社

· 北 京 ·

— 00985

图书在版编目 (CIP) 数据

科学数据管理与共享/孙九林, 施慧中主编. —北京:
中国科学技术出版社, 2002.11
ISBN 7-5046-3405-0

I. 科… II. ①孙…②施… III. 自然科学—数据管理—
文集 IV. N37-53

中国版本图书馆 CIP 数据核字 (2002) 第 088511 号

中国科学技术出版社出版

北京海淀区中关村南大街 16 号 邮政编码: 100081

电话: 62179148 62173865

新华书店北京发行所发行 各地新华书店经售

中国科学院印刷厂印刷

开本: 850 毫米×1168 毫米 1/32 印张: 13.75 字数: 306 千字

2002 年 11 月第 1 版 2002 年 11 月第 1 次印刷

印数: 1-1000 册 定价: 42.00 元

(凡购买本社的图书, 如有缺页、倒页、
脱页者, 本社发行部负责调换)

国家科技基础性工作专项 (G99-A-01e) 资助



00985

前 言

我国科学数据资源十分丰富，但科学数据的共享问题一直未能很好解决。科学界长期呼吁，终因共享问题涉及的面广且十分复杂，因而始终进展不大。科学数据不能共享不仅造成了国家投资的巨大浪费，而且在相当程度上制约了我国科研水平的提高。例如，有关青藏高原研究，应该是我国地学能够领先于国际水平的重要领域，我国诸多部门和多个国家计划相继组织了单学科或多学科的综合调查、观测、探测、试验和实验研究，取得了一系列科学数据和具有区域特色的研究成果。但由于共享问题没能很好解决，部门或科学家群体之间的数据、资料不能有效沟通，影响了研究水平的提高。然而，国外学者却利用其资金优势与我国不同单位分别开展合作，在其所致力研究领域中获取了比国内研究群体更为全面系统的科学数据、资料，使其对一些重要科学问题的研究比国内学者更有发言权。

近年来，信息化工作被提到前所未有的战略高度，科学数据的管理与共享工作开始得到了国家的重视。科技部在1999年实施了“国家科技基础性工作专项”，计划陆续支持了一批急需的科技基础数据库建设。2001年又组织专家，就我国科学数据实行共享问题开展了调研，年底专家们提出的《实施科学数据共享工程，增强国家科技创新能力》调研报告，得到科技部领导的高度重视和肯定。科技部徐冠华部长和中国气象局秦大河局长亲自就我国气象数据共享问题进行协商，并于2001年12月14日联合召开了新闻发布会，宣布气象数据共享试点正式启动，从而在国家层面上，把科学数据共享工作向前推进了一大步。

为了推进我国的科学数据共享工作，中国较早地加入了世界数据中心（World Data Center，简称WDC）组织，并组建了9个学科数据中心，在中国国家委员会的领导下，为促进数据共享做了不懈的努力。科技部1999年启动的基础性工作专项不但支持了部分中心的数据资源建设，同时还就数据共享中的若干技术问题委托WDC中国中心开展研究。众所周知，数据共享问题是个十分复杂的系统工程，它既牵涉到法规、政策、机制、管理问题，又涉及理论、方法、技术问题，需要全社会的关注和努力。为了配合科技部实施的科学数据共享工程及中国气象局启动的气象数据共享试点所营造的大好形势，以WDC中国中心系统为主的一批科技人员，于2002年2月5日至7日在京召开了科学数据共享研讨会，就数据共享的原理、原则，法律、法规，理论、方法，标准、规范，技术、条件等问题展开了研究和讨论，试图将我

国的科学数据共享工作稳步推向前进。

在研讨会筹备和进行中，得到了科技部等有关部委领导和科学家的鼎力支持，在此向他们表示诚挚的谢意。本文集所收录的论文就是以这次研讨会上的报告为主，经整理出版，供读者讨论和参考。我们衷心希望有更多的科技工作者参与到这个领域来，共同推进我国的科学数据共享工作，使科学数据共享工程预期目标早日实现：通过 10 年左右时间的努力，“基本建成覆盖全面、结构合理的国家科学数据中心群及其数据共享网络体系，形成科学数据的良性汇交与更新机制，基本满足政府部门和科研教育机构及社会对科学数据的需求”。

孙九林 施慧中

2002 年 9 月

目 录

科学数据共享：一项功德无量的大工程

(代序言)	孙 枢 (1)
略论科学数据共享	孙九林 (3)
实施“科学数据共享工程”的若干思考	黄鼎成 (9)
科学数据共享原则与政策措施的初步研究	施慧中 (23)
我国科学数据共享存在的突出问题	
原因及其对策	李晓波 等 (40)
我国国有科学数据公益性共享服务在投资	
环节上的保障条件研究	刘 闯 (47)
建设海洋基础数据管理与共享服务机制	林绍花 (66)
从第谷星表看科学数据和数据共享	汤克云 (78)
WDC—D 地质科学数据库群的网络共享	戴爱德 (86)
WDC—D 冰川(雪冰)冻上学科中心数据	
管理及共享政策	吴立宗 等 (96)

生命科学数据管理和信息共享	纪力强 等 (109)
材料环境腐蚀数据共享服务网构建	王光泰 等 (137)
地震学科数据库系统总体设计与实现	赵仲和 等 (151)
海洋资料基础数据库模式设计	薛惠芬 等 (164)
气象数据标准规范体系探讨	刘小宁 (173)
气象元数据标准与系统建设	王国复 等 (186)
海洋环境基础数据的质量控制理论、方法 和技术	张冬生 (199)
潮汐资料水尺零点考证示范研究	徐炎光 等 (208)
关于同位素地质数据质量的判别与保证	丁梯平 (224)
数据同化理论方法和应用技术研究概述 ...	韩桂军 等 (244)
地震波形数据集	周克昌 等 (262)
虚拟天文台与互联网格技术	崔辰州 等 (272)
为实现数据共享建立地震数据资源信息 服务系统	吴 敏 (291)
基于 Internet 的空间科学数据共享技术	佟继周 等 (303)
WDC—D 地下水资源数据库的 WebGIS 共享	张礼中 等 (314)
海量数据存储的简便技术	赵永恒 (321)
重要历史地震记录资料的抢救与数字化 ...	宋臣田 等 (335)
美国国有数据共享的法律与政策基础	王正兴 等 (355)
国外大气科学领域数据中心发展概况	李集明 等 (367)
加拿大地球观测数据的共享网络	戴爱德 (380)

附录

中华人民共和国国务院令	(401)
地质资料管理条例	(401)
地质资料汇交范围	(407)
气象资料共享管理办法	(409)
中国极地科学数据管理规定(试行)	(415)
国家斥巨资推动科学数据共享	(421)
把珍珠串成项链	(423)
科学数据共享工程启动	(425)

科学数据共享：一项功德 无量的大工程

孙 枢

(代序言)

我国进入信息时代的主要标志之一是大量科学数据产生。科学数据是国家科技进步、社会发展、经济建设和全民教育所需要的重要基础信息，是国家宝贵的资源和财富。这已经成为不争的事实。长期以来，国家投资科研项目产生大量的科学数据不能得到广泛应用已经成为制约我国科技发展最主要的障碍之一。1994年，中国科学院地学部以咨询建议的方式呼吁打破地球科学和资源环境科学数据壁垒实现数据共享，得到了原国务委员宋健院士和原国家科委的重视。近年来，一些人大代表、政协委员和更多的科学家也加入到呼吁科学数据共享的队伍。由于体制上的条块分割所带来的数据滞留和不共享，给国家重复投资造成的巨大浪费和对国家科技创新能力的抑制，已经到了科学数据共享非解决不可的地步了。科学数据共享涉及科学技术的许多领域，涉及国家部门、科研单位和科学家个人科技活动的方方面面，牵涉国家、集体、个人利益的再调整，关系到科学、技术、标准、投资、管理、法律、人才队伍、国际、国内等许多方面，同时又涉及历史数据、现在正在产生的数据和未来越来越多的数据的管理。历史遗留问题和将产生的新问题表明仅有呼吁不能从根本上解决我国科学数据共享问题，我国科学数据共

享问题必须从国家层面上统筹规划,协调各个方面关系,必须从机制上建立起我国科学数据共享的保障体系。

科技部和原国家科委多年来为此作了许多工作。3 年以前成立以黄鼎成研究员为首的“科技部科学数据共享调研组”,先后与国家 8 个部委领导、数十位科学家召开了 20 余次专题讨论会,对我国目前科学数据共享存在的主要问题和可能解决的办法等一系列问题进行了详细的调查研究,认真听取了各个方面的意见,客观分析了我国实施科学数据共享的难度和可能性,终于完成了“实施科学数据共享工程,增强国家科技创新能力”的国家科学数据共享总体框架。他们从调查研究入手,从国家整体布局着眼锲而不舍的工作精神值得赞赏。国家科技部徐冠华部长下决心解决我国科学数据共享问题,并批准了实施科学数据共享工程计划,这是对我国科技发展具有重大意义的战略决策。国家实施科学数据共享工程标志着我国科学研究开始全面进入信息时代,对我国现代科技发展也具有划时代意义,对促进我国科技发展的作用将是全面的、深刻的和长远的。科学数据共享工程将是一项功德无量的大工程,它对我国科技进步、社会发展和科教兴国战略实施的促进作用必将逐步地、全方位地显现出来。

略论科学数据共享

孙九林

(中国科学院地理科学与资源研究所)

1999年,我参加了以黄鼎成教授为首的科技部基础司组织的科技基础性工作调研组,以及后来开展的科学数据共享工程的研究与讨论,深感科学数据共享工作在我国将有望向前推进。我作为一名长期从事科学数据工作的科技工作者,对此十分激动,相信广大科技工作者和我会有同感,因而将一些有关想法写在下面。

任何一项有价值的科研成果,都是科技工作者站在前人的肩膀上所开展研究的结果,他们除了个人聪明才智和辛勤的劳动之外,还必须吸收前人的劳动成果,或失败的经验教训。前人的经验、成果,甚至失败的教训,都是通过大量的数据资料、文献或具体实物而沉淀下来的,这些宝贵的财富能否被后来的人所利用,发挥它应有的作用,就成为后人能否在此基础上有所发现、有所发明、有所创新的关键性因素之一。所以每个科技工作者都把能否顺利地获取前人或同行的这些成果、资料等视为自身研究工作的重要内容,所以他们是多么期盼国家能够实施有效的办法,在不损害国家、单位和个人基本利益的前提下,实现科学数据资料共享,使他们所从事的科技工作真正能在前人的基础上有所突破和创

新，为国家科学事业多做贡献。

科学数据共享问题是长期困扰我国科技界的重大问题之一，人人期盼共享，然而实施起来却十分困难。究其原因甚多，长期以来存在人们思想上的小农经济思想意识、私有观念十分严重，把本应属于国家所有的数据资料视为个人或单位的小集团所有，使宝贵的人类长期积累起来的财富无法发挥更大的效益；在市场经济利益的驱动下，一些单位把由国家出资获取的数据变为本单位的财富，成为一种资金收入，填补单位经费的不足；国家缺乏长期稳定的资金投入，使得宝贵的数据资源任意流失，队伍不稳，无法形成一支训练有素的科学数据管理和服务队伍；国家没有科学数据产生、管理、共享服务的政策法规体制和运行机制；缺乏大型科学研究计划的数据管理和健全的汇交制度；科学数据共享的基础设施不健全，缺乏数据共享的规范、标准。凡此种种，使大量科学数据资料，任意散失，重复获取，严重浪费。归结为一点，说明我国法制还不够健全。科学数据走向有效管理和共享机制的形成，必须在经济发展，法制不断健全，人们的思想认识不断提高的前提下才能逐步完善，也就是说必须有一个过程。所以过去所存在的问题是发展中的问题，或者说是难以跨越的一个发展阶段，不是任何个人所为。

科学数据是人类活动的产物，它代表了人类的文明和社会的进步，对它的开发应用又可以进一步推动社会向前发展。可见，科学数据具有重要的科学价值、经济价值和社会价值，是人类认识自然、改造世界的重要基础和知识的源泉，是一种重要的非物质和能量的资源。和其他资源一样，对数据资源的开发利用同样存在一个合理开发和有效管理与保护的问题。如果掌握数据资源的公益性、基础性、共享

性、传递性及增值性的特点，去充分开发利用它，对社会将产生巨大的效益。很多人把数据资源当成一般的物质资源，认为采用交易的手段，去产生经济效益，是利用数据资源的惟一方式，这是一种狭隘的认识。这就是人们常说的，有人把数据资源当成过去地主老财的钱罐，个体农民用于防灾荒的粮垛，个体手工艺者的工具箱等。是因为他们没有看到数据资源产生效益的过程，即数据→信息→知识→政策（方案）或是理论→推动社会进步和科技发展。这是科学数据在充分流动和开发应用中所产生的效益。数据资源的价值发挥是采用一种特殊方式而产生的，不是简单的金钱交易，而且需要全社会的各行业人员，在不同的岗位，按照不同的需要，去开发应用数据资源在不同阶段所表现的形式，从而产生不同的效益。例如，有科学家需要原始数据为他的原始创新发挥作用，有人采用数据变成的成果去指导某种行动，而有人将信息转变成知识并上升为理论或者是可实施的方案，去推动科技进步或社会发展。可见，科学数据的价值需要千千万万的人在不同领域和阶段去利用和挖掘它的实际价值。这就充分说明，科学数据应该在充分的传播和流通中，让全社会的人去利用，在各种环节发挥不同层次人员的作用，即只有在共享状态下，科学数据才可以发挥应有的功能和效益。

当人们议论到在科研活动中的数据资料无法寻找，甚至很多数据资料在国内无法得到，而要到国外去获取国内的数据时，感慨万千。有些国外回来的学者工作一段时间，实在无法忍受国内的这种状况，只好再次返回国外。美国充分认识到科学数据共享的意义和效益。所以他们实行了国有科学数据的完全开放共享的政策，使得美国在近十年中，科学数

据从混乱状态走向有序的管理和全社会共享,得到科学界和全社会的赞赏,并得到很多国际组织的认可与效仿。科学工作者都有深刻的体会,即完成一个科研项目要花费大部分时间去收集已有的科学文献和数据资料,有时为了收集数据不惜工本和精力。如果把这个时间用在深入分析和综合研究工作上,不但会提高研究成果的质量和水平,而且可以去完成更多的国家任务,为国家的科学事业和社会进步多做贡献。

长期以来,为了科学数据共享,一批科学家和科技管理人员为此付出了巨大的努力,甚至做出了不少牺牲。他们站在不同层次上到处呼吁和努力,国家相关部门也做出了很多响应。经国家原科委的批准中国在 20 世纪 80 年代末申请加入了世界数据中心组织。目的是将中国的科学事业与国际接轨。中国成立了 9 个学科中心,一批科技工作者投入科学数据的共享行列。经过十多年的努力,虽然与国际的同行们有交往和接触,但因多方面的制约,真正的科学数据共享工作在国内收效甚微。近十年来,科技部采取了多种方式,试图推进科学数据共享,如设立国家可持续发展数据共享试点项目,在一些重大项目开始时,就注意研究数据资料的共享问题等。实践证明仅靠科学家呼吁,而政府没有措施,或者政府部门积极,而没有全社会响应,都是不行的。只有科学家和政府部门相互配合才能使科学数据共享真正得以实施,而且要在若干政策法规、标准的指引下,才能使其健康的向前发展。

科学家希望有一个科学数据共享的环境,但有少部分科学家又把共享理解为让别人把所产生的科学数据和资料贡献出来,而没有意识到自己的科研成果资料和数据也应拿出来共享。也就是说应该建立一个共建、共享、互惠、互利的科

学数据的共享运行机制。我们每一位科学家有权利享用国有科学数据资源，更有义务利用国家投入，产生科学数据和成果供社会共享。真正把共享机制建设好，要求我们每一位科学数据的生产者具有宽阔的胸怀，在利用自己生产出来的科学数据研究出成果以后，将数据毫无保留地贡献给社会，这批数据对你来说已经产生了价值，让别人利用以后又可能产生新的价值。要摒弃我不用的也不让你用的狭隘意识，现在特别需要我们大批科学家积极投身到科学数据共建共享的行列中来，只有这样，政府有行动，科学家有要求（对科学数据的需求，希望贡献自己科学数据的愿望和行动），使我国科学数据共享取得真正的成功。

在有关部门领导和专家们的指导下，我从事了 20 多年的科学数据工作，历经艰难和困惑，有时顺利，有时是走投无路，但始终没有丧失信心，科学数据管理和共享工作实际是一件具有实际应用价值的科学管理工程，我从来就认为这是科学研究和技术工作。多年来，很多人看不起数据工作，这是科技界的一件憾事。仔细想一下就会知道，哪一项科技创新不和数据有关，哪一项高质量的研究成果或重大工程的实施没有大量的、系统的、科学的数据在支撑！随着信息技术迅速发展和应用，以科学数据为核心的新的科研环境的构建，正成为人们的共识。1998 年，当我和我的同事们提出虚拟科研环境的时候，部分人还认为这是很遥远的事情，现在已经被很多人所理解甚至向这方向努力。虚拟科研环境需要大量的科学数据做基础。中国科学院花了近 20 年时间在努力建设科学数据库系统，已将 40 万册图书全部数据化在网上提供阅览；中国台湾中央研究院开展了典藏工程。所有这些均是利用现代信息技术去挖掘和整理中华民族长期以