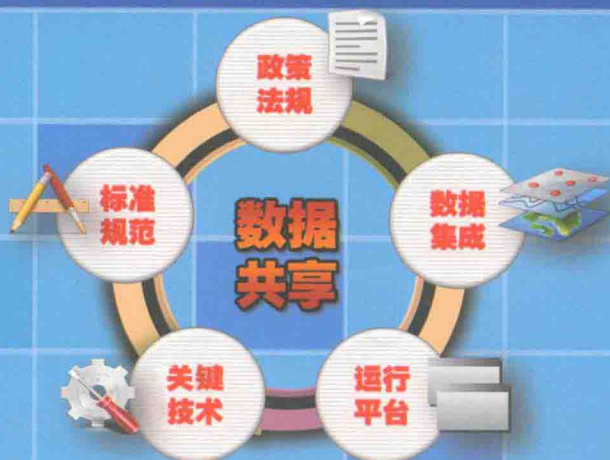


地球系统科学数据集成共享研究：

标准视角

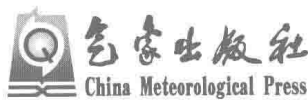
王卷乐 著



国家科技基础条件平台——地球系统科学数据共享平台
科技基础性工作专项项目(2007FY110300、2011FY110400、2013FY114600) 联合资助
中国科学院信息化专项项目(XXH12504-1-01)
江苏省地理信息资源开发与利用协同创新中心

地球系统科学数据集成 共享研究：标准视角

王卷乐 著



内 容 简 介

《地球系统科学数据集成共享研究：标准视角》面向地球系统科学数据集成与共享的难题，立足标准视角分析了数据集成共享的内涵和需求，建立了地球系统科学数据集成共享标准规范体系，研制了地球系统科学数据共享元数据、数据分类等核心规范；形成以元数据为核心的科学数据集成共享技术框架，研究了元数据互操作关键技术，建立数据质量评价模型；开展了国家地球系统科学数据共享平台的数据资源集成共享和典型区域科学数据集成应用，并将这些标准和技术拓展应用到国家科技计划项目数据汇交管理。

本书可供科学数据管理与开放共享相关研究和管理人员以及从事地球科学研究、资源环境调查、自然资源管理与入地关系综合分析等应用研究的科研人员和信息处理的技术人员以及相关学科的教师和研究生参考。

图书在版编目(CIP)数据

地球系统科学数据集成共享研究：标准视角/王卷乐著.

—北京：气象出版社，2015.9

ISBN 978-7-5029-6221-0

I. ①地… II. ①王… III. ①地球系统科学—数据处理—研究
IV. ①P

中国版本图书馆 CIP 数据核字(2015)第 219606 号

Diqiu Xitong Kexue Shuju Jicheng Gongxiang Yanjiu: Biaozhun Shijiao

地球系统科学数据集成共享研究：标准视角

王卷乐 著

出版发行：气象出版社

地 址：北京市海淀区中关村南大街 46 号

总 编 室：010-68407112

网 址：<http://www.qxcbs.com>

责任编辑：李太宇

封面设计：博雅思企划

印 刷：北京中新伟业印刷有限公司

开 本：787 mm×1092 mm 1/16

字 数：320 千字

版 次：2015 年 10 月第 1 版

定 价：60.00 元

邮政编码：100081

发 行 部：010-68409198

E-mail: qxcbs@cma.gov.cn

终 审：章澄昌

责任技编：赵相宁

印 张：12.5

印 次：2015 年 10 月第 1 次印刷

序

科学数据资源通常可以分为两大类型：一类是行业部门按照统一的规范标准长期采集和管理的科学数据；一类是国家各类科技计划项目在研究过程中产生的以及为支持科学研究而通过观测、监测、试验等站点采集的研究型科学数据。这两类数据都是科学数据共享中必须进行整合集成并为科技创新提供支撑服务的数据资源。然而，由于研究型科学数据分散性大、分布面广、标准化程度低，且资源量大，从而给科学数据资源整合和共享带来了更大的难度。

王卷乐博士针对地球系统科学数据整合集成中的难题，结合国家地球系统科学数据共享平台建设等相关实践，从数据共享标准规范入手，以标准视角对研究型数据整合集成的标准体系、技术框架和应用实践等方面开展了系统研究：一，分析了地球系统科学数据集成共享的内涵和需求，建立了地球系统科学数据共享标准规范参考模型，主持制定了地球系统科学数据共享元数据、分类编码等核心标准规范，并开展相关国家标准的制定，为推进地球系统科学数据规范化集成共享提供了标准支持；二，以元数据为核心形成了多学科、多源数据整合集成和发布技术框架，研究了地学元数据互操作的关键技术，制定了地学数据的质量评价模型并开展实践，可为科学数据的集成共享提供技术参考；三，开展了国家地球系统科学数据共享平台的数据整合集成、典型区域科学数据整编等实践，并将科学数据集成共享标准规范成果拓展应用于国家科技计划项目数据汇交管理，可为相关领域的科学数据集成共享实践提供借鉴。

本书系统地将该研究中的发现和成果展示给读者。这些成果正随着国家科技基础条件平台——地球系统科学数据共享平台的发展而推广应用，同时也与世界数据系统（WDS）等相关国际数据组织的科学数据共享活动接轨。希望这一研究能够在未来更深入地开展，并在国际和我国科学数据管理与开放共享中发挥更大的作用。

中国工程院院士



2015年7月25日

前 言

地球系统科学是从行星地球角度出发,将地球大气圈、水圈、冰雪圈、岩石圈和生物圈看作一个相互联系的有机整体(地球系统),是综合研究各圈层形成机理、变化规律及其相互作用并为区域可持续发展和全球变化研究提供理论依据和调控方法的学科。地球系统科学研究近年来如火如荼地开展,凸显了基础科学数据支撑不足的问题。随着全球变化、地球圈层关系与机理、人地关系与环境效应等重大科学计划的进展,越来越需要这些多年沉淀和积累的科学数据资源支撑。当前的数据分散、条块分割、难以共享严重阻碍了我国基础科学的研究和发展,并且还无谓地增加了大量的低水平数据资源重复建设,造成严重的浪费。

针对地球系统科学数据的集成和共享问题,本研究从标准视角建立地球系统科学数据集成共享的标准规范体系,分析其集成共享技术框架,并在国家地球系统科学数据共享平台中开展实践。全书共分四部分11章。第一部分是第1章和第2章,介绍科学数据共享的概况,并针对地球系统科学研究数据的特点分析其整合集成的内涵和需求。第二部分包括第3、4、5章,分别是地球系统科学数据共享标准规范体系、元数据标准和数据分类标准,统筹设计了包括指导标准、通用标准和专用标准三个层次,内外兼容的地球系统科学数据共享标准参考模型;参考模型涵盖概念术语、数据分类、描述、集成、建库、分发、服务和质量控制等内容,为地球系统科学数据集成共享标准的科学制定提供了顶层的指导依据;以ISO 19115元数据标准为基础,建立了可扩展的地球系统科学数据共享元数据标准;以用户数据共享服务为导向,建立了扁平化的地球系统科学数据主分类与编码体系。第三部分是技术篇,分别在第6章介绍了以元数据为核心的地球系统科学数据整合集成技术体系,第7章介绍了地学元数据互操作的技术方法,第8章介绍了地学数据质量评价过程与方法。第四部分是实践篇,分别在第9章介绍了国家地球系统科学数据共享平台数据集成概况,第10章介绍了海岸带及近海历史数据资源集成应用,第11章介绍了国家科技计划项目数据汇交管理集成应用。

本研究主要是结合国家科技基础条件平台——地球系统科学数据共享平台

的建设和服务开展的。感谢国家地球系统科学数据共享平台团队对本项工作的支持和指导。感谢苏萍、赵强、柏永青、柏中强等研究生参与排版整理。限于专业领域覆盖面和写作能力，可能会有错误或不足，欢迎批评指正，以便更新时改进。

王卷乐

2015 年 7 月于北京

目 录

序
前言

第 1 章	绪论	1
1.1	科学研究数据共享背景分析	1
1.2	地球系统科学数据共享概况	10
第 2 章	地球系统科学数据整合集成的内涵和需求	14
2.1	研究型科学数据整合集成问题	14
2.2	研究型科学数据共享面临的主要问题	17
2.3	地球系统科学数据整合集成内涵	18
2.4	地球系统科学数据整合集成原则	23
第 3 章	地球系统科学数据共享标准规范体系	25
3.1	数据共享标准规范体系	25
3.2	地球系统科学数据共享标准规范体系设计	30
3.3	地球系统科学数据共享标准规范体系初步构建	35
第 4 章	地球系统科学数据共享元数据标准	41
4.1	元数据的概念和作用	41
4.2	地球系统科学数据元数据标准	43
4.3	地球系统科学数据元数据扩展模型	50
第 5 章	地球系统科学数据共享分类标准	56
5.1	国际地球系统科学数据分类体系	56
5.2	地球系统科学数据共享的用户服务特征	59
5.3	地球系统科学数据分类原则	61
5.4	地球系统科学数据共享目录与关键词表分类	62

第 6 章 地球系统科学数据整合集成技术体系	68
6.1 数据整合和集成的方法	68
6.2 数据整合集成的模式分析	72
6.3 基于元数据的数据发布技术框架	74
第 7 章 地球系统科学数据元数据互操作方法与关键技术	82
7.1 元数据互操作层次框架	82
7.2 地球系统科学数据元数据互操作分析与实践	84
7.3 元数据互操作 Web Service 技术研究	93
7.4 元数据查询服务设计与开发	98
第 8 章 地球系统科学数据质量评价过程与方法	103
8.1 数据质量评价的问题和现状	103
8.2 栅格格网数据质量评价过程	105
8.3 栅格格网数据质量评价方法	110
8.4 栅格格网数据质量评价实例	115
第 9 章 地球系统科学数据集成共享应用	119
9.1 GEODATA 共享平台原型构架	119
9.2 面向服务的地学元数据技术体系研究	124
9.3 地学元数据技术体系开发与实践	127
9.4 基于元数据的应用服务系统开发	135
第 10 章 海岸带及近海历史数据资源集成应用	151
10.1 海岸带研究热点与数据需求	151
10.2 海岸带及近海主体数据库体系设计	160
10.3 海岸带数据资源集成技术	167
10.4 海岸带综合调查历史空间数据集成技术	169
第 11 章 科技计划项目数据汇交管理集成应用	175
11.1 国外科技计划项目数据汇交政策	175
11.2 数据汇交集成与共享技术流程	180
11.3 科技计划项目数据汇交的实践与启示	185
参考文献	189

第 1 章 绪论

1.1 科学研究数据共享背景分析

科学数据是人类社会科技活动积累的或通过其他方式获取的反映客观世界的本质、特征、变化规律等原始性、基础性数据,以及根据不同科技活动需要进行系统加工整理的各类数据的集合(GB/T 31075—2014)。科学数据被誉为科学研究的生命和血液,科学数据资源如同工业社会中的石油。科学数据资源在信息社会的广泛应用,造就了社会财富的巨增,成为重要的国家战略性资源和各国科技实力竞争的重要资本,对科技进步与创新和经济增长、社会发展以及国家安全都发挥着重要的作用。一个好的科学思想、理论假说和应用技术,必须在掌握大量前人资料和科学数据的基础上才能形成,同时也必须在大量相关数据的支撑下才能被证伪。对科学数据进行系统化的综合分析,进而促进新的科学思维的产生,是实现科技创新的重要方式,并推动交叉学科的发展。在竞争激烈的科技创新全球化时代,拥有科学数据就意味着拥有了无穷的创新资源,就有了提升国家科技竞争力的最广泛的基础。

正是由于科学数据的战略性地位,数据信息贫富不均已成为国家和地区发展不均衡的巨大鸿沟,开展科学数据的集成和共享得到世界各国的重视。科学数据的作用只有通过流动和共享才能体现。“填平数据鸿沟,连接数据孤岛”,“把珍珠串成项链”(孙九林等,2002)等数据共享理念,逐渐在科技界的推动下取得了突破和进展。

1.1.1 国外科学数据共享进展

近 30 年来,从国际组织到单个国家特别是发达国家,在信息技术广泛应用与发展的基础上,不断加强科学数据、资料和相关信息的获取、管理与面向社会服务的步伐,积极推进科学数据的流动与低成本使用,并从政策、法律制度、技术规范、组织管理等各个方面保证科学数据信息的管理与应用的正常秩序。

(1) 美国科学数据管理与共享

美国是世界上科学数据拥有量最多的国家,特别是在地球科学和生命科学领域,其数据拥有量占据世界总量的 80% 以上。在 1998 年全球 11339 个数据库中,美国生产的数据库占总数的 63%,无论在数据库数目还是数据量规模,美国的科学数据和数据库在世界上都占有绝对优势(刘闯,2003)同时,美国也是世界上介入科学数据共享管理最早的国家。

为了充分发挥科学数据对于全社会科技进步的支撑作用，早在 20 世纪 60 年代美国政府就介入科学数据共享管理并通过立法规范和促进科学数据共享。

1) 制定数据共享战略部署。在战略思路，美国强调要在国家层面上统筹规划科学数据的管理，实行科学数据完全和开放政策(full and open)。在战略步骤上，美国在国家层面上重点抓三件事，即统筹规划数据共享机制和数据共享体系，数据共享工作预算和投资保障，数据共享政策法规的制定、完善和监督。第一步侧重于数据中心建设。如由美国国家航空航天局(NASA)在全国优选九个数据中心组成国家级分布式数据中心群(Distributed Active Archive Centers)。第二步是侧重于法规和网络共享环境建设。如由白宫设立总统长期专项——美国全球变化研究项目(US Global Change Research Program)，其中包括全球变化数据信息系统(Global Change Data Information System)，历经 10 年建成了世界上最强大的科学数据共享体系。

2) 制定公共数据开放和共享的指导政策。美国科学数据共享政策的核心是，除了危及国家安全、影响政府政务和涉及个人隐私的数据和信息以外的国有(公共领域)数据和信息，全部实施“完全与开放”政策。

3) 制定数据共享运行机制。归纳起来，美国科学数据共享的运行机制分为保密性管理机制、完全开放管理机制和市场管理机制三种类型。

4) 构建科学数据共享法律法规体系。其重要法律主要包括《信息自由法》、《政府阳光法》、《电子政务法》、《隐私权法》、《版权法》、政府规章等。

美国的全球数据库规模很大，涉及领域众多，例如大气、海洋、航天、遥感应用、生物医药等。下面仅以医药卫生领域的美国国立卫生研究院(National Institutes of Health, NIH)为例，介绍其数据资源整合情况。NIH 中的一个著名的集成数据库就是基因序列数据库 GenBank(Genetic Sequence Database)。该数据库 1982 年由阿拉莫斯(Los Alamos)国家实验室建立，后由美国生物技术信息国家中心(National Center of Biotechnology Information, NCBI)负责该数据库的建设和维护工作。GenBank 数据最早主要来自于生物学文献，后来则直接来源于测序工作者提交的序列、由测序中心提交的大量 EST(Expressed Sequence Tag)序列和其他测序数据以及与其他数据库协作交换而来的数据。GenBank 每天都与欧洲的核苷酸序列数据库(EMBL)、日本核酸数据库(DDBJ)交换数据，使这三个数据库的数据保持同步。GenBank 的数据呈指数增长，1985 年仅有 5700 条记录，至 2006 年 2 月在传统的 GenBank 分支系统中，已有 597 亿多个碱基，序列记录达到 5458 万余条；在全基因组测序(WGS)信息中，已有 631 亿多个碱基，序列记录达到 1246 万余条。早期，GenBank 主要以磁盘或磁带介质存储和发行，1998 年 GenBank 不再发行光盘版数据库，数据可通过 FTP 直接获取。目前，GenBank 的数据可以从 NCBI(National Center for Biotechnology Information)的 FTP 服务器上免费下载完整的库，或下载积累的新数据。NCBI 没有对 GenBank 数据的使用和分发设置任何限制。

(2) 欧盟的科学数据管理与共享

欧盟通过科技发展规划和项目资助的方式来推进科学数据共享。其在共享政策、数据管理与法律保障方面的做法，可以概括为以下几个方面。

1) 欧盟的科学数据共享政策。欧洲《布加勒斯特宣言》代表了欧盟对公共科学数据、公共当局持有的信息公开共享的公益性共享原则和指导思想。《布加勒斯特宣言》(以下简称《宣言》)是2002年11月9日在为筹备信息社会世界高峰会议而召开的布加勒斯特泛欧大会上发表的宣言。《宣言》认为,“这个信息社会以广泛传播和分享信息、各利益相关方(包括政府、私营部门和民间团体)的真诚参与为基础。在争取让人人都充分享受到信息社会的益处的努力过程中,这些利益相关方的贡献至关重要”。

2) 欧盟的科学数据管理代理机构。当前欧盟有16个依这种单独立法形式而建立起来的欧共体代理机构,如欧洲环境代理处(EEA)、欧洲共同体植物品种办公室(CPVO)、欧洲改进生活和工作条件基金会(EUROFOUND)、欧洲航空安全代理处(EASA)等。

3) 构建数据共享法律体系。其法律法规项目主要包括:《欧盟条约》、《欧共体条约》、信息公开立法、有关科学数据保护立法、有关网络和信息安全立法、知识产权立法、有关基础设施和网络服务的政策法律等。

(3) 世界数据中心的数据管理与共享

世界数据中心(World Data Center, WDC)成立于1957年,在ICSU世界数据中心专门委员会的指导下开展工作,截至2007年全球建有52个WDC学科中心。世界数据中心所涵盖的数据内容广泛,在时间尺度上涵盖了从瞬间(秒)到百万年尺度。这些数据主要用于对地圈与生物圈、外层空间以及天文现象变化的研究,包括对渐变现象和突变现象的研究,对可预见现象和不可预见现象的研究,对自然现象和人文现象的研究。世界数据中心不仅在地球、空间和环境科学领域中极大地推进了数据工作的发展,努力为各类科学团体服务,而且积极参与了许多较大的国际科学计划,为科学发展做出了许多重要和有益的贡献。

世界数据中心的三个主要活动是数据的收集、交换和服务,同时,又应国际科联的要求,承担了一系列重要的国际科学计划申明要保存的数据的管理。WDC的具体数据共享管理和服务活动包括:与其他WDC中心合作收集、整理数据和信息;有序地保存数据;以最少的复制和发布成本向用户提供数据;与数据集的生产人共同编制数据说明文档;通过数字化的方式保存重要的历史数据;为小尺度、地区性和全球性地球物理学研究整编专题数据;将数据制作成如压缩光碟等媒体,便于用户查找且易于转移为可用的格式;评估数据存储媒体的老化、错误生成率和保存期限的技术性问题;综合多源数据加工生产数据产品,例如求算太阳或地磁活动的指数;编制数字模型以描述随时间和空间变动的地球物理环境;维护与以上活动相关的在线信息服务;建立用户服务机制,使得科学家在获得WDC系统的数据集时能够得到专业工作人员的协助;帮助科学家寻找和获取WDC系统所没有的相关数据。

在WDC发展的50多年中,不仅各个国家都在大力开展本国的科学研究活动,国际间的联合科学研究计划也在蓬勃发展,这些不断加强的科学研究活动需要越来越多、越来越专业的科学数据支撑。尤其是进入21世纪以来,这一需求强烈增加的速度明显超出了WDC自身发展的速度。这使得WDC遇到了前所未有的挑战,大体可以归纳为以下四点(王卷乐和孙九林,2009)。

1) WDC系统可持续性的挑战。WDC是一个松散的科学数据交换与共享组织。ICSU

本身并没有对这些 WDC 数据中心进行任何经费资助，各个数据中心都要依托于本国所在机构的资助而发展。在这种情况下，WDC 各数据中心的发展不得不要受到依托机构自身发展目标的影响，其第一业务还是要首先完成各个机构自身的相关任务，而作为全球数据中心的第二功能是第二位的。如果依托机构的主要任务与起初该数据中心成立时的任务越来越不相符的时候，这些 WDC 数据中心能够继续获得稳定的经费支持，能否还以数据共享为主要使命，能否持续地向前发展？都会是很大的问题。

2) WDC 系统总体布局的挑战。WDC 在总体上的学科体系布局没有顶层框架，各个中心之间缺乏有机的联系。这造成 WDC 系统内的各数据中心在学科和地域布局上不够合理。如当前 WDC 的全球布局只有 9 个中心在中国，1 个中心在印度，其他的中心都在发达国家；WDC 在美国、俄罗斯、日本和中国都设有地球物理数据中心，但是缺少有效的整合和联系。在数据资源内容上也没有顶层设计，没有形成整体数据服务“系统”。

3) 数据支撑服务能力的挑战。WDC 建立之初所确立的数据服务对象就是地球科学和地球系统。这一系统研究所需要的数据具有不可度量的复杂性，需要多学科的数据支持，这正是 WDC 系统所具有的多学科特点。然而，这一机遇的一个对等的威胁就是 WDC 很难采用统一的标准来促进多学科数据的集成与支撑服务，这导致 WDC 体系多学科联合数据服务的局面始终没有形成，严重地影响了其数据支撑服务能力。另外，作为世界数据中心，WDC 有义务瞄准和支持关于地球系统和全球变化研究等全球性科学问题的研究，例如气候变化、海啸、预警、人口压力等等。这就需要综合集成的、多种尺度的数据集支撑服务。然而，目前 WDC 在数据的尺度和地域上都很难满足需求(郑军卫等，2008)。

4) WDC 的网络互操作体系挑战。绝大多数 WDC 数据中心目前只有本中心的数据集目录，各个中心独立对外提供数据服务。并且，各个中心的数据集在元数据格式、数据标准上有着显著的差异，缺少开展联合服务所需要数据互操作接口，从整体上还没有建立 WDC 系统统一的数据检索和获取体系。在这种情况下，如何加强各个数据中心之间元数据标准的统一、开发和建立数据的互操作接口是一个技术与标准上的挑战。

WDC 于 2008 年正式更新为世界数据系统(WDS)。WDS 变革框架如图 1.1 所示。

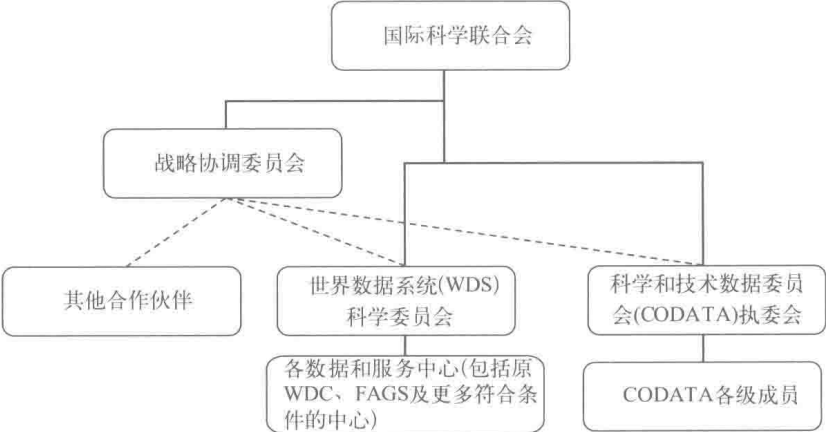


图 1.1 ICSU 的数据和信息体系战略结构图

(Ad hoc Strategic Committee on Information and Data Final Report to the ICSU Committee on Scientific Planning and Review[M]，2008)

WDS 的目标是确保能够广泛和公平地获得有质量保证的科学数据，数据服务、产品和信息；确保长期的数据管理；促进遵守数据标准和公约；提供机制以方便和提高数据和数据产品的访问。

截至 2015 年 6 月，WDS 共有 91 个成员组织，包括 59 个正式成员、10 个网络成员、4 个合作伙伴、18 个协作成员。正式成员见表 1.1 所示。

表 1.1 WDS 成员列表

WDS 正式成员	研究领域
美国地震学研究联合会数据服务中心	地震；大地电磁；气象实测数据；气压；海洋传感器；超导重力仪；次声波
WDC 地理信息与可持续发展中心	空间科学；地球科学；文化和种族研究；经济学；地理学；社会学；计算机科学；数学；统计；系统科学；环境研究与林业
世界土壤信息中心	地球科学；地理学；农业；环境研究与林业；土壤学；可持续土地管理；国土资源信息系统
WDC 德国气候计算中心	地球科学；气候建模
世界气象数据中心，阿什维尔	地球科学；经济；计算机科学；气候科学及相关数据管理；世界土壤信息中心
法国斯特拉斯堡天文数据中心	空间科学；天文学
世界冰川监测服务中心，苏黎世	地球科学；地理；冰川学
WDC 气象学、火箭与卫星、地球自转中心	空间科学；地球科学
澳大利亚南极数据中心	空间科学；地球科学；生命科学；化学；物理学；地理学；环境研究与林业
中国天文数据中心	空间科学；物理学；天文学
WDC 再生资源与环境学科中心	地球科学；地理学；农业；环境研究与林业；区域研究；自然资源；生态；地理信息
法兰德斯海洋研究所	生命科学；计算机科学；环境研究与林业；生物学；生物多样性；生物地理学；分类学；海洋学；信息和通信技术；数据管理
WDC 海洋学中心	地球科学；海洋学
国际地球自转和参考系统	空间科学；地球科学；地理学；计算机科学；数学；统计；系统科学；大地测量学与参考系
台湾鱼类资料库	生命科学(生物多样性)
WDC 海洋学科中心，天津	地球科学；海洋科学
WDC 地球物理学学科中心	空间科学；地球科学；计算机科学
泛大陆—地球和环境科学数据发布中心	地球科学；生命科学
WDC 日地物理学中心，莫斯科	空间科学；地球科学；地磁；电离层；空间射线；太阳活动；行星际环境(临近空间)
热带生态评估和监测网络	地球科学；农业；环境研究与林业；生态
WDC 太阳黑子指数中心	空间科学；统计；历史；太阳物理；太阳活动(中期和长期)；日地关系和气候

续表

WDS 正式成员	研究领域
WDC 海洋学科中心, 奥布宁斯克	地球科学; 物理和化学海洋学
WDC 大气遥感中心	空间科学; 地球科学; 化学; 物理学; 地理学; 计算机科学; 数学
WDC 地磁学科中心, 哥本哈根	空间科学; 地球科学; 地磁学
地磁指数国际服务中心	空间科学; 地球科学; 日地物理学; 空间天气; 地磁学
WDC 地磁学科中心, 爱丁堡	空间科学; 地球科学; 地磁学
WDC 固体地球物理学中心, 莫斯科	地球科学; 地震学; 地磁学(主磁场); 考古学和古地磁; 重力测量; 地热; 近期地壳运动; 海洋地质与地球物理
WDC 气象学科中心, 奥布宁斯克	地球科学; 气象
WDC 太阳活动数据中心	空间科学; 太阳物理学
WDC 地磁学科中心, 京都	空间科学; 地球科学; 物理学; 地理学; 计算机科学; 地磁学
跨学科地球数据联盟	地球科学
澳大利亚气象局信息中心	空间科学; 日地物理学
美国航天局分布式主动档案中心	地球科学; 地理学
橡树岭国家实验室分布式主动档案中心	地球科学; 地理学; 区域研究; 陆地生态; 生物地球化学动力学; 生态数据; 环境演变
世界应力图计划	地球科学; 地球物理学; 地球化学; 地质学; 自然资源
WDC 冰川学中心	地球科学; 地理学; 极地地区和冰冻圈
WDC 电离层和空间气象数据中心	空间科学; 地球科学; 电离层; 日地物理; 空间天气预报
乌克兰地理空间数据中心	地球科学; 计算机科学; 数学
莫斯科地理信息中心	地理学; 结构与环境系统的演化; 环境影响因素; 资源的可持续管理; 俄罗斯及国外人文地理; 大气之间相互作用及岩石圈, 水圈; 制图; 地理信息技术; 遥感; 地理和地质教育
地球资源观测和科学数据中心	地球科学; 系统科学; 环境研究与林业; 遥感; 土地变化科学; 土地变化监测; 评估和预测
语言档案中心	心理学; 语言与语言学; 人类学
WDC 古地磁学中心	地球科学; 气候学; 全球变化
南非开普敦大学数据中心	经济学; 地理学; 政治学; 统计; 环境研究与林业; 健康科学; 区域研究; 信息科学
世界微生物数据中心	生命科学; 微生物学
哥达德地球科学数据和信息服务中心	地球科学; 物理学; 地理学; 计算机科学; 农业; 工程; 大气科学; 降水; 水文; 全球建模; 信息科学; 系统工程
地壳动力学数据信息系统	空间科学; 地球科学; 物理学; 大地测量学; 空间大地测量学
空间科学数据中心	天文学; 空间科学; 计算机科学; 空间物理学; 空间气象学; 行星科学
寒区旱区科学数据中心	地球科学; 地理学
全球水资源中心	地球科学; 计算机科学; 系统科学; 环境研究与林业; 水文循环; 闪电; 极端气象

续表

WDS 正式成员	研究领域
意大利天文档案 IA2 中心	天文学；空间科学
美国校际社会科学数据共享联盟	经济学；性别和性研究；地理学；政治学；心理学；社会学；统计；家庭和消费科学；健康科学；历史；老龄化；刑事司法；人口；教育；法律；药品滥用...
大气科学数据中心	地球科学；大气科学；云；气溶胶；对流层化学
WDC 地磁学科中心，孟买	物理学；地球科学；空间科学；地磁学；固体地球地磁和高层大气科学；高层大气物理学
加拿大天文数据中心/加拿大虚拟天文台	天文学；空间科学
阿拉斯加卫星设施	地球科学；地理学；区域研究；冰冻圈；极地过程；固体地球；磁层；阿拉斯加州地理研究
加拿大海洋信息中心	计算机科学；地球科学；海洋科学；地球物理；海洋物理；海洋生物学；生物化学；海洋工程
瑞典环境气候数据中心	地球科学；环境研究与林业；环境；气候
社会经济数据和应用中心	地球科学；化学；文化和种族研究；经济学；地理学；政治学；社会学；计算机科学；统计；系统科学；农业；建筑与设计；商业；工程；环境研究与林业；健康科学；交通；人类学；区域研究；环境科学；可持续发展的科学；气候学；信息系统科学
美国 NAVSTAR 联合大学	地球科学；大地测量学
陆地过程分布式数据档案中心	地球科学；地理学；农业；环境研究与林业；土地覆盖；土地变化；土地处理

(源自：<http://www.icsu-wds.org>，2015. 6. 30)

(4) 科学与技术数据委员会的科学数据共享与管理

国际科技数据委员会 CODATA(<http://www.codata.org>)是国际科学联合会于 1966 年成立的一个跨学科的科技数据领域的国际权威学术机构，其宗旨是提高所有科技领域内重要数据的质量，广泛推动对重要科技数据的编辑、评价和传播，致力于提高对科技数据的管理、可靠性与可访问性。中国于 1984 年加入 CODATA，并以中国科学院牵头，成立了 CODATA 中国全国委员会，委员来自于国内各研究院所、高校及相关政府部门。近年来，得益于中国科技数据共享及科研信息化等工作的深入推进，中国科学家在国际 CODATA 中的影响和作用日益加大。

围绕着国际科学联合会 2006—2011 战略计划，CODATA 致力于发展高质量的科学数据信息所提供的科学与社会可持续发展机制。工作重点包括数据科学领域的数据存取问题(Data access)、数据质量(Data quality in the Internet era)、数据归档(Data archiving)、网络数据资源交换(Interoperability of Web data resources)、重点科技数据集(New key data sets)、数据科学(Data science)、发展中国家的科技数据共享与应用(Developing countries)等。

至今，CODATA 在数据科学与技术上已有近 50 年的领导科学数据前沿研究、知识开

发和数据资源建设的大量经验，不断提供科技数据处理的最新思想和技术。CODATA 关注科学技术各个领域的来自实验、观察和计算的各类数据，这些领域包括物理科学、生物学、地质学、天文学、工程、环境科学、生态学及其他学科。从基础科学数据到前沿科学数据的评价、传播与应用，CODATA 都取得了众所瞩目的成绩。CODATA 为数据科学的发展进步提供了一个国际舞台，在信息时代里成为国际数据资源共享的畅通渠道，其引领的数据科学将日益促进国际化的科技交流与创新发

展。全球性的综合研究是当前科学研究的一个重要趋势，在这一方面国际组织起到了很大的作用。在区域性、综合性的科学研究过程中，产生和积累的大量数据对科学研究具有重要的支撑作用。除了 WDS、CODATA 以外，许多国际组织都在全球科学数据共享领域做出了重要贡献，例如世界气象组织(WMO)、原地球系统科学联盟(ESSP)的四大国际科学计划，即世界气候研究计划(WCRP)、国际地圈生物圈计划(IGBP)、国际全球环境变化人文因素计划(IHDP)、国际生物多样性计划(DIVERSITAS)等等，都在开展的科学研究的过程中，积累和传播科学数据及其产品。类似这样的科学计划还有很多，尤其是在跨学科集成研究领域，在此不再赘述。

1.1.2 国内科学数据共享历程

随着科学技术的迅猛发展和信息化的推进，科学数据的共享与服务逐步为人们所共识。“填平数据鸿沟，连接数据孤岛”是信息社会科技发展的自身需求与时代的必然，同时也是增强国家科技创新能力和国际竞争力的必由之路。正是在这种背景下，我国自 20 世纪 80 年代末就在多个层面上逐步推动科学数据的共享，最终建立了科学数据共享工程和国家科技基础条件平台。科学数据共享的历程可以概括为以下几个大事件：

20 世纪 80 年代在原国家科委和科协支持下，中国科学院联合有关政府部门和科研教育机构，组建了世界数据中心中国中心(World Data Center—D, WDCD)和国际科技数据委员会中国委员会(Committee on Data for Science and Technology, CODATA)；在基础科学和若干前沿领域的基础性数据，开展了有效的国内外交流与服务。

1994 年由原国家科委、自然科学基金委员会、中国科学院的有关司局联合组织的《走向 21 世纪的中国地球科学》的调研，提出了加强包括科学数据、资料和信息；重大仪器装备；标本馆与地质遗迹；文献情报等等在内的科学基础设施建设和建立、健全共享机制的建议。中国科学院地学部向国家提出关于加强科学数据共享的建议，随后又多次提出国家科技规划中要切实解决科学数据共享问题。

1995 年原国家科委相继建立了虚拟的国家科技图书文献中心和中国可持续发展网络信息系统等。

1997 年原国家科委还就青藏高原综合考察科学数据共享进行了部际协调。多年来，国家自然科学基金委员会坚持项目数据资料的汇交管理。

1999 年，科技部在科技基础性工作和社会公益性研究专项中，启动了科技基础数据库建设。

2001 年，完成《实施科学数据共享工程，增强国家科技创新能力》的调研报告。同年

年底,中国气象局在科技部的支持下率先启动了气象数据共享试点,开展气象数据共享服务。

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、孙鸿烈、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。

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、野外科学观测研究台站和计量基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、微生物菌种资源、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,包括科技成果信息服务体系、公益与行业共性技术转化平台和技术标准支撑体系;六是网络科技环境平台,包括国家条件平台应用服务支撑系统、网络计算应用系统、网络协同研究与工作环境、全国科普数字博物馆和全国科技信息服务网。

2008年12月科技部、财政部联合发布《关于进一步推动国家科技基础条件平台开放共享工作的通知》,并于2009年组织开展在建平台评议(科技部,2008)。2011年末,科技部、财政部通过组织专家评审,完成了首批国家科技基础条件平台(以下简称平台)认定评