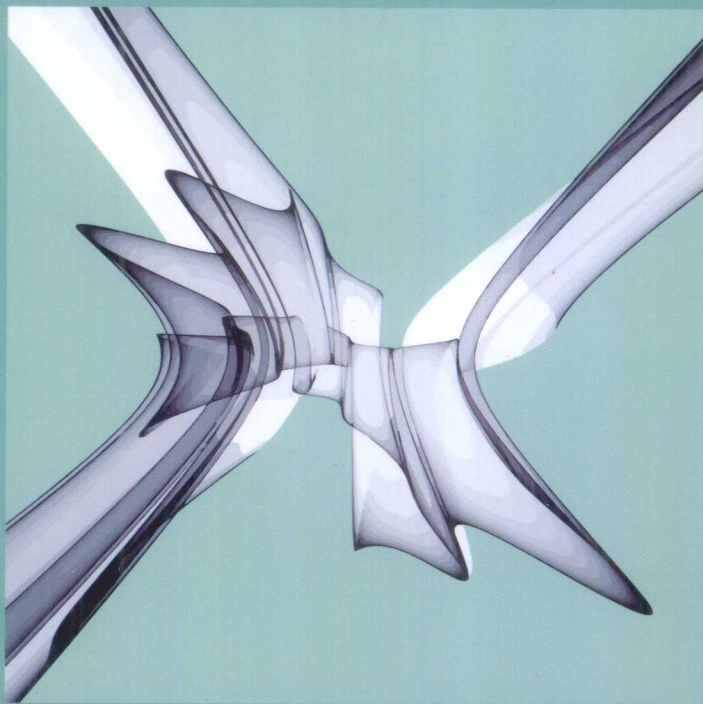


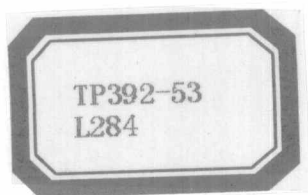
科学数据库与 信息技术论文集

第十一集

中国科学院科学数据库办公室 编



科学出版社



郑州大学*04010745224T*

科学数据库与信息技术 论文集

第十一集

中国科学院科学数据库办公室 编



科学出版社

北京

TP392-53
L284

内 容 简 介

本书收入论文 59 篇, 主要反映中国科学院近年来在科学数据库的建库技术、网络技术、信息服务技术等方面所取得的成果及学术上取得的进展。这些论文也集中体现了近年来国内数据库与信息技术方面的研究和应用水平。

本书可供从事数据库技术、网络技术和信息系统研究的科技人员、工程技术人员参考, 也可供相关学科的研究人员、大专院校师生阅读。

图书在版编目(CIP)数据

科学数据库与信息技术论文集. 第 11 集/中国科学院科学数据库办公室编. —北京: 科学出版社, 2012.1

ISBN 978-7-03-033242-4

I. ①科… II. ①中… III. ① IV. ①数据库-文集②信息技术-文集 IV. ①TP392-53②G202-53

中国版本图书馆 CIP 数据核字(2011)第 280967 号

责任编辑: 陈 静 王 颖 任 静/责任校对: 刘小梅

责任印制: 赵 博/封面设计: 刘可红

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2012 年 1 月第 一 版 开本: 787×1092 1/16

2012 年 1 月第一次印刷 印张: 30 1/2

字数: 741 000

定价: 138.00 元

(如有印装质量问题, 我社负责调换)

编辑委员会

主任委员：孙九林

副主任委员：阎保平 陈维明 马俊才 陈 刚

编 委：（以姓氏笔画为序）

王云鹏 代 亮 庄大方 纪力强

李天宪 李望平 李德铎 杨 锐

肖 云 邹自明 张晓林 张耀南

罗 泽 赵永恒 南 凯 郭 杉

温 浩 黎建辉

第十一集编辑部

主 编：李望平

编 辑：郑晓欢 张丽丽

序

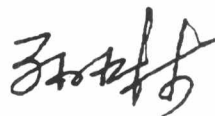
中国科学院科学数据库已经历了二十多年的建设和发展历程。二十多年来，在院、所各级领导关注、指导和全体建库工作者的共同努力下，科学数据库逐步成为国内外有较大影响的重要数据库之一，并在科技创新、学科发展、国家经济建设、国防建设、规划决策、国际合作等诸多方面得到广泛应用，取得了显著的社会效益和一定的经济效益。

“十一五”期间，中国科学院科学数据库作为中国科学院信息化建设的重要基础设施，在科学数据环境建设与应用等方面都取得了重大进展与突破。在数据资源建设方面，重点对科学数据库资源进行了整合，形成了8个主题库、4个专题库、2个参考型库以及37个专业数据库，整合可共享的数据量达148TB；在服务环境建设方面，基本形成了由网络运行服务总中心、学科领域网格主节点和数据资源网格节点三层架构的科学数据网格体系。研发部署了数据访问服务等一系列中间件以及科学数据网格门户系统，在化学、生物学、空间科学、人地系统4个学科领域建成了网格应用示范系统。科学数据库已初步形成结构合理的科学数据资源体系，并取得了数据资源整合服务的良好效果。2010年，在线数据访问量年均达707万人次，年均数据下载量达101.8TB；累计为210项科研项目提供了直接的数据服务。

2012年3月在海南省三亚市即将召开的第十一届“科学数据库与信息技术”学术研讨会，主要是进一步研讨自2010年第十届学术研讨会以来，在科学数据库建库技术、网络技术、信息服务技术、共享政策、标准规范、应用服务、总体发展等方面的进展和取得的成果，并对科学数据库今后发展相关的新技术、新思路进行探讨。这次会议收到论文近70篇，经编辑委员会认真评审，选出59篇论文进行会议交流并收入本次会议的论文集。这些论文集中体现了近年来中国科学院科学数据库及有关信息技术的研究和应用的水平，可供从事数据库及信息系统研究的同行和学生参考。

中国科学院科学数据库今后将会更好地面向国家“十二五”发展战略规划和中国科学院“创新2020”的发展需求，更紧密地与社会经济发展计划相结合，与中国科学院知识创新工程和研究所的学科发展相结合，突出科学数据长期有效积累与支持科学研究应用服务，为我国的基础科学数据共享做出更大的贡献。

“十二五”中国科学院信息化专家咨询委员会主任



2011年9月16日

目 录

序.....	孙九林
--------	-----

总 论

中国科学院科技数据资源现状及其发展思考	陈明奇(3)
科学数据的个体识别和跨学科集成	陈维明(10)

数据库系统及建库技术

化合物子结构检索及时响应的流程优化与分布式优化研究	李英勇 陈维明(21)
植物化学成分数据库建设	徐挺军 赵英莉 陈维明(30)
中国植物物种信息数据库关联性设计与查询服务	何延彪 庄会富 王雨华(38)
数字化物种编目的三个重要问题	王利松 杨 永(49)
植物药用知识传承与利用的数据库建设与应用探索	庄会富 何延彪 王雨华(58)
毫米波射电天文数据库研究	逯登荣 孙继先 刘 梁 杨 戟(64)
化合物结构的唯一性编码和化学数据库管理	许 禄 章文军(70)
空间数据存储技术及四大空间数据库发展浅析	
.....	苏 华 罗小青 王云鹏 黎丽莉 杨静学(77)
NoSQL 当前发展及应用状况	黎丽莉 杨静学 苏 华 李 岩 王云鹏(84)
系统生物学中多组学综合数据库的设计与实现	
.....	唐碧霞 王彦青 陈 旭 庞 博 赵文明(91)
Web 数据库安全技术分析及其应用	李宏波 卢振举(98)
Linked Data 在科学数据库中的应用探讨	沈志宏 胡良霖 侯艳飞 黎建辉(105)
数字版权保护与数字水印技术在科学数据库中的应用	刘 峰(115)
基于网络服务的开放矢量地理空间数据管理系统	
.....	冯 敏 尹 芳 诸云强 朱华忠 宋 佳 任正超(126)

数据共享与服务技术

人地系统科学数据网格建设初探	宋 佳 张金区 冯 敏 吕 宁 陈鹏飞(137)
----------------------	--------------------------

化学主题数据库：数据网格技术的应用范例	
..... 赵月红 徐俊波 温 浩 陈维明 戴静芳 李英勇 许 禄 章文军(145)	
材料科学主题数据库的数据整合与共享服务	崔丽娜 叶万江 杨 锐(156)
空间科学数据网格模型应用服务框架设计与实现	郑 程 佟继周 邹自明 高文健(161)
中国动物主题数据库服务平台的设计与实现	
..... 林聪田 韩 艳 乔慧捷 原 帅 纪力强(167)	
卫星轨道演示软件的设计与实现	何战科 李 川 杨旭海 李伟超(176)
面向数据密集型计算的科学数据网格作业调度系统	
..... 薛正华 黎建辉 周园春 张 洋 沈 庚(183)	
一个轻量级分布式计算框架的设计与实现	徐衍波 陈维明(191)
化合物理化性质数据的整合和服务	戴静芳 陈维明(198)
基团贡献法物性估算模块的设计开发	冷建飞 徐俊波 赵月红 温 浩(205)
基于专业日志分析的科学数据库平台用户活动研究——以中国数字植物标本馆 (CVH)	
为例	许哲平 覃海宁 马克平 赵莉娜 李 奕 包伯坚(214)
基于 Flex 的基因组可视化工具	吴 军 刘 翟 孙清岚 马俊才(221)
基因组可视化工具 GBrowse 及其应用	唐碧霞 王彦青 陈 旭 庞 博 赵文明(229)
古生物学和地层学专业数据库中地层数据的可视化	侯旭东 樊隽轩 陈 清 张琳娜(237)
基于 MRS 工具的海量数据检索	夏 青 刘 翟 马俊才(244)
LAS 在南海物理海洋数据服务中的应用研究	徐 超 李 莎 米浦春(251)
PHP 常见安全漏洞研究	刘 鹏 张玉清(261)
基于 HITS 的科学数据检索结果排序的研究	滕常延 沈志宏 丁翠萍(269)
RDF 在科学数据集成检索中的应用	李成赞 沈志宏 李晓东(275)
LEQL：一种面向科学数据实体的级联关联查询语言	何 星 郭志斌 巩晓冬(285)
配置管理数据库 (CMDB) 的关键技术研究及实现	董济农 康红勋 牛雄飞 于海姣(292)
大规模遥感图像高性能集群存储处理技术研究及实现	
..... 张海明 黎建辉 周园春 王学志 薛正华 阎保平(299)	
科学数据管理与共享工具及技术研究	苏贤明 沈志宏 李成赞 何 星(306)
通用数据源连接器设计及实现	金日男 沈志宏(315)
973 计划资源环境领域数据汇交服务实践与成效	
..... 杨雅萍 杜 佳 王卷乐 宋 佳 乐夏芳(323)	
移动互联网环境下群组日程协同软件设计与实现	杨宏伟 刘学敏 于建军 马永征(330)
英汉冰冻圈科学词汇收录修订与共享系统	赵雪茹 张耀南(336)

数据共享与标准规范

- 东北亚资源环境综合科学考察数据集成标准规范体系研究与实践 王卷乐 朱立君 杨 懿(349)
- 科学数据学科领域本体的思考与探索——以植物学领域本体为例 朱艳华 胡良霖 刘 宁(358)

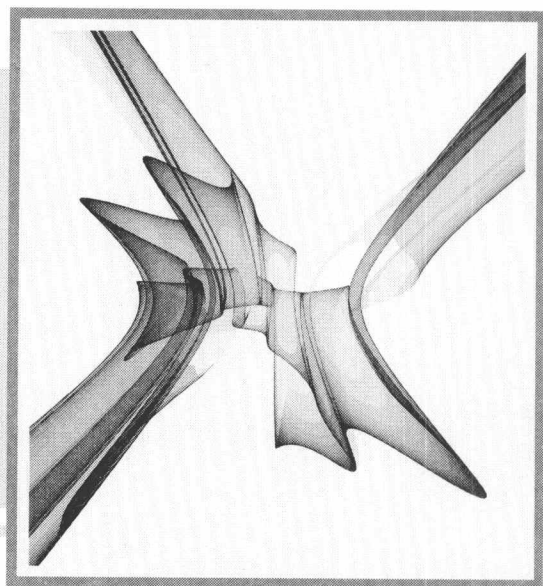
文献数据库及全文检索技术

- 中国化学文献自动标引方法及其应用研究 赵英莉 陈维明(367)
- CSMR 面向生物资源类文献的生物资源信息采集与引用挖掘平台 苏晓林 马俊才 刘 翟 孙清岚(374)
- 基于分布式搜索的科研成果元数据在线收集系统 刘学敏 钱 芳 杨宏伟 李华飏(381)

数据库应用

- 从数据库的引用标注看科技资源的应用服务成效 廖顺宝(391)
- 基于协同工作环境的黑河流域生态-水文集成研究的 e-Science 建设 张耀南 康建芳 敏玉芳 汪 洋 赵雪茹 何振芳 赵国辉(397)
- 基于科学数据库的黑河流域 e-Science 平台建设 汪 洋 张耀南(405)
- 化合物性质预测方法的比较与性质预测平台的设想 周俊红 陈维明(415)
- 化学反应数据库与量化计算在合成设计中的应用策略 李 佳 解丽娜 胡 静 姚建华(422)
- 基于 MODIS 数据珠江三角洲大气颗粒物反演 杨静学 王云鹏(427)
- 地学台站气温观测数据的时空分析模型构建与应用研究 赵国辉 张耀南 汪 洋(434)
- 基于 3S 技术的广西环江县喀斯特区域景观格局分析 张明阳 王克林(447)
- 黑土生态数据库支持下的农田土壤肥力评价 付 微 谢叶伟 李 勇 赵 军 孟 凯(456)
- 蛋白质组学相关数据库资源分类及其应用 雷小元 卢振举(463)
- 基于 DNA 条形码的物种鉴别系统研究与实现 孟 珍 周园春 黎建辉 王雨华 曾宏辉(470)

总 论



中国科学院科技数据资源现状及其发展思考

陈明奇

(中国科学院办公厅 北京 100864)

摘要 本文首先介绍了中国科学院信息数据资源整体情况和科技数据资源的发展应用现状,然后重点介绍了中国科学院及部分研究所的数据管理和共享实践情况,以及“十二五”期间科技数据资源有关工程的主要任务,最后分析了中国科学院科技数据资源的可持续发展的机制,提出了相应的针对性建议。

关键词 信息资源; 科技数据; 可持续发展机制

2011年2月11日,《科学》杂志推出以“数据处理”^[1]为主的专辑,关注数据洪流带来的挑战与机遇。专辑在导言部分中指出,数据的搜集、维护和使用已成为科学研究的主要方面。对许多学科而言,海量数据意味着更严峻的挑战,更好地组织和使用这些数据会有助于将巨大机遇变为现实。科学由数据所推动。今天的社会因多种不同目的而需要使用这些科学数据,因此,应当让数据能够被更广泛地获取,这成为科学研究的一个基本要素。数据管理是一项需要共同努力的事业。促进科学事业发展最重要的力量必须来自科学界,《科学》呼吁科学界在数据的供应和管理上做出积极贡献。

除了科学界之外,企业界对数据的重视程度也在不断提高,有些企业^[2]和专家提出我们已进入了“大数据”时代的观点。

因此,数据资源不仅对科技,也正在和即将对社会、经济发展产生重大影响。必须对科技数据在科研活动中的基础性地位、战略性地位给予充分认识,不断加强科技数据的积累整合与共享应用,这是中国科技发展必须面临的长期任务之一。中国科学院的科技数据工作,理应在其中发挥重要作用。

1. 中国科学院信息资源与数据资源情况

依据“摸清家底、推动共享、整合集成、服务应用”的原则,中国科学院在2010年开展了中国科学院信息化资源的整理,并出版了《中国科学院信息化资源报告》,目的是尽可能全面地展现中国科学院历年以来通过信息化手段所积累的各类全局性资源,以及具有一定推广和应用价值的专业资源。这为中国科学院信息化资源的推广和共享创造了条件,为整合集成各类信息化应用提供了便利环境。

在此过程中,研究提出了中国科学院信息化资源的定义,即以科研信息化(包括科研活动和科研管理)为核心,围绕数据生命周期全流程而积累起来的基础设施、信息、技术、工

具、政策、人才、应用服务等一切要素的集合。

中国科学院信息化资源体系由基础支撑资源、信息与软件应用系统资源、保障环境资源三个部分构成。基础设施包括网络接入、高性能计算、数据存储、互联网基础资源；信息资源包括科学数据库、教学资源库；业务平台包括科研管理系统、科研系统平台、教育培训系统、网络科普平台、信息发布平台、其他应用软件等。

经征集整理，中国科学院可公开并共享的主要信息资源情况如图 1 所示^[3]。

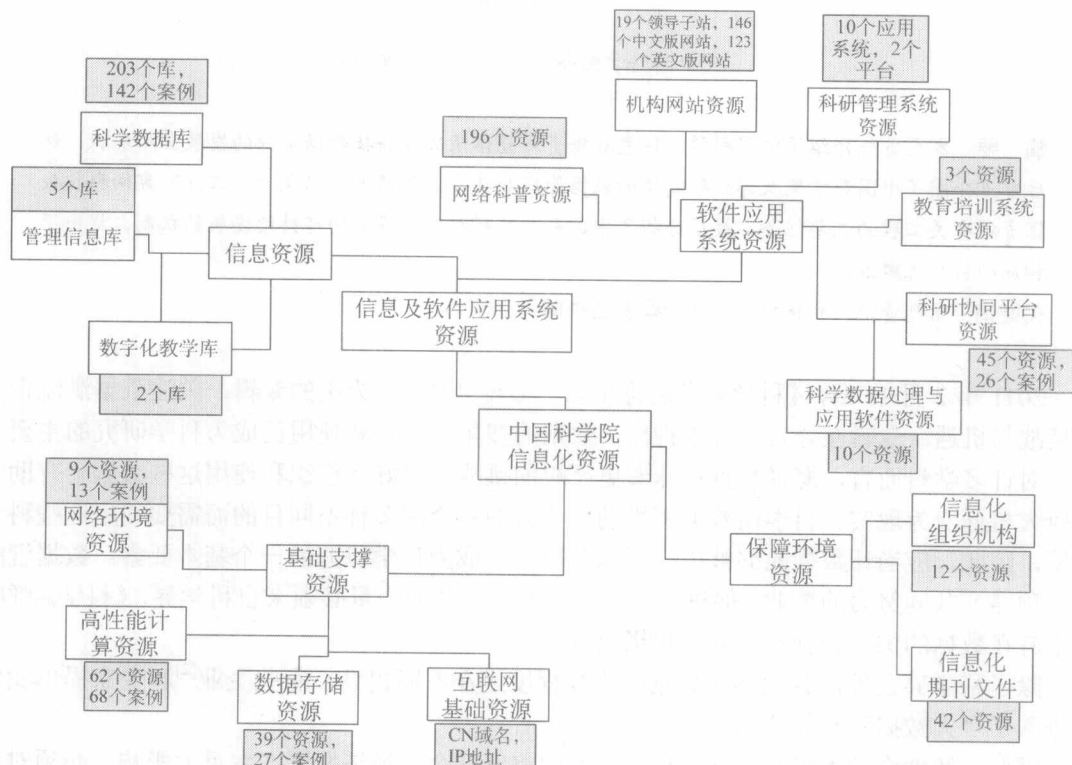


图 1 中国科学院信息资源分布地图

“十一五”期间，中国科学院科技数据基础设施正式列入中国科学院信息化基本环境，并进行重点建设，为科研活动提供综合性的数据应用环境^[4]。全院从信息化战略资源高度，系统地规划科学数据资源体系。根据学科领域的数据整合需求，初步建成了化合物参考型数据库、中国植物物种信息参考型数据库，以及天文科学、空间科学、人地系统、资源环境遥感、化学、材料、中国动物、微生物与病毒等 8 个主题数据库。根据重大项目需求，建设了核聚变、生态系统功能区划、冻土环境、中国科学院青海湖联合科研基地等 4 个专题数据库。同时，根据科研数据整理和共享需求支持建立了 37 个专业数据库。专业数据库由具体研究所承担建设，主题、专题和参考型数据库则通过组织相关研究所进行跨所联合建设。

此外，积极引进了国际上优质的科学数据资源，集中建设了陆地资源卫星遥感影像库（LANDSAT）、遥感影像数据库（MODIS）、数字高程数据库（DEM）、生物信息学数据库、气象资料数据库、斯隆数字巡天数据库（SDSS）、大气科学数据库等数据资源。截至 2010

年底,科学数据库建成可共享数据 148TB,通过资源整合形成了 528 个数据库。中国科学院科学数据库数据资源领域分布情况如图 2 所示。

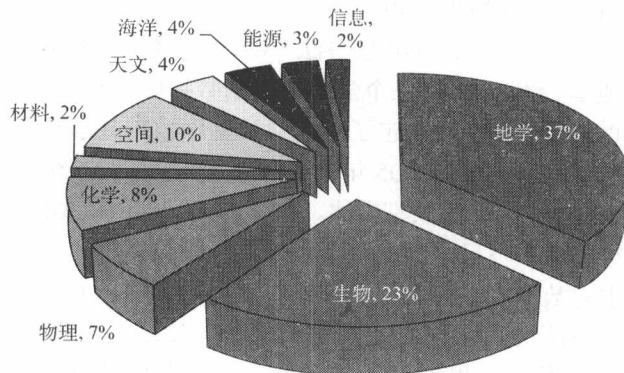


图 2 中国科学院科学数据库数据资源领域分布（专业库个数）状况

通过科学数据库门户推动科学数据共享和提供集成服务,由建库单位的数据网站以及部署在科学数据中心的镜像保障服务稳定运行,同时还开展必要的离线数据支持服务。截至 2010 年底,“十一五”期间科学数据库访问量累计达 2606 万人次,数据下载量累计达 243TB,提供数据服务的各类重大科研项目 210 项,典型应用案例 101 项。中国科学院科学数据库网络年度访问人次增长趋势图如图 3 所示。

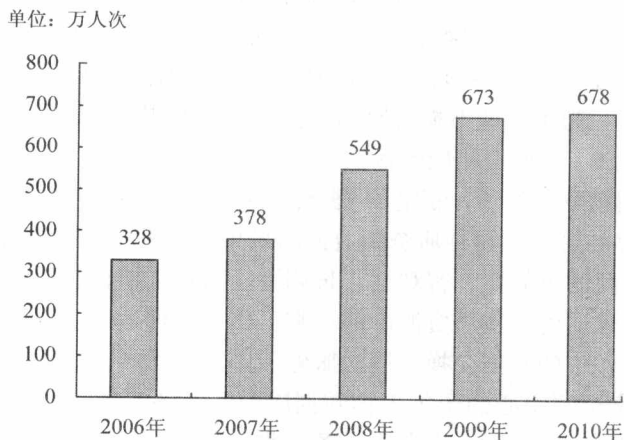


图 3 中国科学院科学数据库网络年度访问人次增长趋势图

2. 中国科学院科技数据资源共享实践

中国科学院科学数据共享的提出、研究和实践由来已久。1982 年,科学数据库工作被列为中国科学院“七五”和后十年十项重大基本建设项目之一,这就是早期科学数据共建共用的具体举措。经过二十多年的发展,特别是伴随着信息技术的革新及其在科学活动中的广泛且日渐深入的应用,众多学科领域的科学数据已呈海量趋势并仍在快速增长中。同样,在此

过程中,科学数据共享的探索也在逐步深入,部分科研项目、学科领域和研究所都分别基于自身的需求、特点,在科学数据共享方面进行了积极的研究和实践,并取得了很好的效果。

2004年前后,中国科学院数据库项目相继完成并发布了《国内外科学数据共享政策研究报告》、《中国科学院科学数据库设计共享办法(试行)》、《科学数据库建设指南》、《科学数据库服务指南》,以及地学、化学、生物三个学科领域的数据库分类分级共享服务指南等成果。院计算机网络信息中心科学数据中心制定了“科学数据库数据共享声明”,由此各数据库明确自己的数据共享和发布策略。截至2005年10月,参加项目的45个单位均发布了其数据共享的声明,明确了423个专业子库的共享方式,其中77%以上的数据库为公开共享。2008年,“数据应用环境建设与服务”项目组面向“十一五”科学数据库建设的新特点,特别是数据库类型的多元化设置主题数据库、参考型数据库、专题数据库、专业数据库,对《中国科学院科学数据库设计共享办法(试行)》进行了修订完善,使之更加适合项目工作的需求。

从《中国科学院信息化评估报告2010》来看,我院科学数据库建设整体规划和统筹管理方面有所加强;大多数研究所比较重视数据库规范性建设,69.3%的研究所能提供科学数据库的相关质量管理措施和方法。此外,也有一些研究所对数据库的运行管理和共享制定了具有自己特色的制度规范与管理体系,或者积累了较好的实践经验。

中国科学院水利部成都山地所制定了《山地科学数据共享服务条例》,主要规范了山地科学数据库管理,包括总则、管理机构的职责、日常运行、安全管理和保密守则等;数据采购、更新与汇交,包括数据采购流程、数据更新周期和内容、数据汇交流程等;数据共享发布与安全保密,包括不同涉密等级数据共享原则,数据共享办法和实施细则等。

中国科学院地理科学与资源研究所对CERN数据和用户实行分类、分级管理,同级用户拥有同等的权利和义务。数据用户被分为5级,分别为:0级——国家及院有关部门;1级——CERN各成员单位;2级——中国科学院内研究单位;3级——国内研究单位及其他非营利机构;4级——其他。为了保证不损害原始数据生产方和提供方的利益,保护其合法产权,对各类数据分不同用户级别设定相应的保护期限。

中国科学院昆明植物研究所,制定了《野生植物种质资源野外采集数据整理整合标准规范》,明确了长期保存的野生植物种质资源数据的采集、整理及上报注意事项,确保了种质资源库植物资源数据来源的可靠性、客观性、长期性,为野生植物种质资源的采集和保存提供数据保障。该规范不仅适用于种质资源库的日常工作,同样适用于同国内外不同种质保存机构之间的种质交换工作中的数据交换,同时服务于国家自然科技资源平台项目。

中国科学院国家空间中心的子午工程数据库中心与各设备、各节点签订了数据接口规范,明确了数据汇交要求;同时,在各个设备验收过程中专门建立了针对数据质量的专家评议体系,数据通过专家评议后,才能汇交到数据库系统中,数据入库过程遵循统一的数据管理规范,以保证数据在存储和发布过程中的规范性,切实保证数据资源的可靠性和可用性。

3. 科技数据资源整合与共享工程的主要内容

目前,全院面向科研信息化整体运维服务的数据应用环境已初步形成,为科技数据的存储与积累、分析与处理、共享与应用奠定了良好的发展基础。为此,中国科学院在“十二五”期间,将面向数据驱动科学发现的新趋势和当代科技数据海量化的新特征,计划实施科技数

据资源整合与共享工程^[5],该工程在“十二五”期间涉及的科技数据信息资源的主要任务是:

(1) 加强科技数据整合,提供科技数据与信息资源服务。面向中国科学院“科技海”中积累和产生的海量数据,以中国科学院大科学装置产生的实验数据,各类地面、海洋、大气和空间监测体系获取的观测数据,各类社会经济发展数据,仿真模拟与计算数据等为重点,实现各类科研数据的有效保存、管理与共享。

(2) 完善科技数据网络,实现数据资源共建共享与协同服务。基本建成中国科学院“科技海”中各类科技数据资源自动汇聚、发布、集成、共享的动态科技数据网格环境,形成由网格运行服务总中心、网格主节点和网格节点组成的层次架构。基于科技数据网格环境,面向“科技海”中各类个性化应用的需求,提供各种开放接口,实现科技数据与科技文献、专利资源的关联与集成,为用户提供多源异构的数据服务(data as a service, DaaS),直接调度、按需使用分布式、海量的科技数据与信息资源,形成从数据到知识的智能服务环境。并作为科技数据运行服务的基础平台,着力面向各类项目的应用需求,支撑领域数据深度集成,实现开放、稳定和安全的科学数据网格服务。

(3) 加强科技数据政策和标准规范研究制定,确保科技数据海—云服务可持续发展。深入分析科技数据资源保存和共享的发展趋势,结合科技活动实际情况,制定全院科技数据政策,推动基础科研数据管理规划、科研数据归档保存和科技数据资源整合服务,与科研项目管理相结合,逐步建立科技数据的归档、追溯、发布与引用等管理和评价机制。通过实施全院一致的基本数据政策,指导科研数据的全生命周期管理和推动有竞争力的科技数据资源整合建设与共享服务。针对基础科研数据的管理和科技数据整合服务,制定数据归档管理和资源整合的相关技术标准和指导规范,重点形成科研项目活动的数据管理规范,基础科研数据的管理、发布、引用等技术标准和规范,数据基础设施的运行管理规范、质量与评价的服务质量(SLA)标准规范,面向科技数据云服务的资源定位、数据接口、数据分发等规范,以及数据密集型处理的大规模数据集模型规范、数据流框架等。通过相关标准规范的确立,保障科技数据的长期运行服务,以及资源整合和深层挖掘应用。

通过实施该工程,将推动全院科技数据资源、数据存储与处理设施、科技数据与信息资源共享、科学数据应用环境等建设与服务,形成以海量科技数据为核心的系列“海—云”服务,成为科技云的重要支柱,进一步提升科技数据战略资源管理和支撑服务能力,逐步建成面向科技界开放共享的国家级科技数据中心,为中国科学院及我国的科技创新提供强大的数据基础。

4. 推动科技数据资源的可持续发展

科技数据是最宝贵的资源,这已成为各方的共识。在中国科学院参与单位及专家学者的共同努力下,中国科学院的科学数据库发展了二十多年,在数据资源积累、标准规范、数据共享技术与数据应用、服务,以及人才培养方面都取得了长足的进步。尤其经过“十一五”期间的整合,形成了由专业数据库、主题数据库、专题数据库和参考型数据库组成的科学数据资源体系。孙九林院士等专家的有关思考和建议对我院科技数据库的发展具有重要的参考作用^[6]。下一步发展面临的主要问题是持续加强信息资源的整合共享与有效利用水平。此外,缺乏可持续发展机制也是面临的一个主要问题,需要有所突破。

“十二五”期间,为了推动中国科学院科技数据资源的可持续发展,需要从不同层面上继续建立相应的运行机制。

从研究所层面,应改变由课题组或项目组管理数据资源的情况,应基于研究所层面来管理数据资源并提供共享服务,确保数据资源的长期可持续发展。研究所应基于本所涉及的主要学科领域创新发展,高度重视科技数据资源的持续积累和整合,将之视为学科发展的重要基础设施,为研究所的“创新 2020”以及“十二五”期间的“一三五”发展战略提供支撑服务,通过建立专业机构,建设专门基础设施或依托院基础设施,出台专门政策等措施,改变科技数据资源实际上掌握在课题组或者 PI 手中的局面,确保数据资源为所里的战略发展服务。

从院层面,一方面是改变立项机制,应建立科技数据资源质量和服务效果的科学评估系统,制定一套类似 SCI 论文引用指数的科技资源引用指数,对科技数据资源的共享及使用情况进行评价;即基于后评价制度,对科技数据资源进行不同额度支持,从而推动其运行、可服务,并推动质量和效果优秀的科技数据资源获取国家和社会力量支持。另一方面,应建立促进科学资源积累、共享与利用政策。制定落实相关政策制度,推动我院科学数据的积累、共享、共建、共用。例如,制订并发布我院科学数据共享办法(试点),借鉴国内外已经制订的科技计划项目数据汇交办法,开展针对我院某一领域或某一先导专项的科技计划项目数据共享方案的研究工作,重点研究解决包括共享数据的内容和方式、各方职责的划分与绩效考评等在内的若干关键问题。

相信通过上述不同层面的努力,逐步建立包括统一数据管理及共享数据政策体系、考核机制、运维保障机制等在内的可持续发展机制,中国科学院科技数据资源事业必将在相应的可持续发展机制的保障下不断取得进步。

参 考 文 献

- [1] <http://www.sciencemag.org/site/special/data/>.
- [2] http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf.
- [3] 中国科学院信息办. 2010 年中国科学院信息化发展报告. 北京, 2010.
- [4] 中国科学院信息办. 2010 年中国科学院信息化资源报告. 北京, 2010.
- [5] 中国科学院“十二五”信息化发展规划编写组. 中国科学院“十二五”信息化发展规划, 北京, 2011.
- [6] 孙九林. 科学数据库发展战略思考//科学数据库与信息技术论文集. 北京: 兵器工业出版社, 2010.

The Status and the Sustainable Development of Scientific Data Resource in CAS

Chen Mingqi

(General Office of Chinese Academy of Sciences, Beijing 100864, China)

Abstract Firstly, the overall situation of information resources of Chinese Academy of Sciences(CAS), and the

current development and application's status of scientific data resources of CAS are investigated. Then the practices of CAS and some institutes of CAS are mainly introduced, and the objective and main tasks of the scientific data resource project of the Twelfth Five-Years plan are introduced. Lastly, in order to achieve the sustainable development of scientific data resources, the related mechanisms are analyzed, and suggestions are proposed.

Key words information resource; scientific data; sustainable development mechanisms