

# 迎接科学数据库发展的新阶段

## ——中国科学院科学数据库发展 20 年的回顾与思考

桂文庄\*

(中国科学院高技术研究与发展局 北京 100864)

关键词 中国科学院,科学数据库,20周年

### 一 回顾

上世纪 80 年代初期,计算机和数据库技术的进步为科学数据积累和应用提供了先进的手段,作为科学技术工作的基础设施,我院率先提出了建设科学数据库的思想。1982 年底,院决定把“科学数据库及其信息工程”列入此后 10 年重大基本建设项目,从此开始了我院科学数据库的建设工作。1986 年 6 月,国家计委批准该项目纳入国家重点建设计划。1987 年 5 月,院决定科学数据库作为重大项目进行支持。

1993 年 7 月,院决定把科学数据库纳入基础研究特别支持领域,并成立了“科学数据库工作委员会”进行管理。1996 年重新组建为科学数据库专家委员会。

到 1996 年,13 个单位参加了科学数据库的建设,建成专业数据库 90 个,数据量达到 7.9GB,建成了比较完整的信息系统。1998 年,“科学数据库及其信息系统”获国家科技进步奖二等奖。

在院的持续支持和专家委员会领导下,科学数据库得到了迅速发展。1996—2000 年期间,全院有 21 个单位参加了建库工作,专业库达到 180 个,总数据量迅速增长到 725GB,网上服务站点达到 19 个,有 153 个

库可提供网络检索。8 年的持续支持,使科学数据库得到了稳定、规范的发展。

2001 年,院启动了信息化建设专项,科学数据库作为科研信息化的重要工作之一列入了专项支持。到 2005 年底,我院科学数据库建库单位达 45 个,专业子库达 503 个,总数据量达到 16.6TB。建成以深腾 6800 为主机的信息系统以及科学数据库中心和 45 个研究所的门户网站,提供 7 × 24 小时网上服务。同时,在这期工作中还开展了科学数据库的标准规范与共享政策的研究,研制实施了 21 个标准规范和数据共享政策,423 个数据库发布了共享声明,9.48TB 数据实现了网上共享。2005 年网站年访问量达 174 万人次,平均访问量约为 6 000 人次/天。科学数据库已发展成为国内规模最大、设施先进、管理规范、通过互联网向全国科研和教育界提供科学数据共享服务的平台,在国内外产生了较大影响。

科学数据库涉及物理与化学、材料科学、地理科学、天文与空间科学、生物与生命科学、能源与环境科学等各科学领域。它们与我院科研工作紧密结合,在科研活动中产生、整理并在为科研服务、为经济建设和社会宏观决策服务中发挥了重要作用。例如,“中国自然资源数据库”支持了包括发改委与中科院重大创新课题“中国西部生态-经济区划及典型区可持续发展模式与对策研

\* 中国科学院信息化工作领导小组副组长、办公室主任,原中国科学院高技术研究与发展局局长  
修改稿收到日期:2007 年 1 月 4 日



中国科学院

究”,国家基金委重大研究计划“中国西北近50年城市化与环境互动作用机理研究”等一批重要项目的研究。“病毒资源数据库——病毒保藏数据库”支持了该院多个研究所在禽流感方面的联合重大研究项目。“大气科学与环境数据库”支持了大气物理研究的一系列国家和院重要项目,如“我国重大气候和天气灾害形成机理和预测理论的研究”、“亚洲季风区海-陆-气相互作用对我国气候变化的影响”等,并为院内外多家单位提供了数据服务。“空间环境数据库”密切结合“双星计划”、“子午工程”等重大空间环境探测计划和研究工作,提供实时空间环境数据,支持了“空间环境预报模式研究”、“地球空间暴多时空尺度的物理过程研究”等重要项目,特别是向载人航天工程提供了及时可靠的空間环境预报服务,为保障“神舟”系列飞船发射飞行的空间环境安全做出了贡献。“遥感卫星图像检索数据库”支持了全国国土资源大调查,3年来共计提供了1122万平方公里的卫星数据,受到国土资源部的肯定。

## 二 思考

首先,面临科学数据库发展的新阶段,必须进一步明确我院科学数据库在院和全国科技工作中、在国家科技数据共享平台建设中的地位、特点、价值和作用。

近年来,在我国经济社会持续高速发展的形势下,中央提出了增强我国自主创新能力,建设创新型国家的重大战略任务,科技投入不断增强。科学数据作为科技发展的重要基础之一,得到了国家层面的重视。2002年科技部提出了建设科技共享平台的计划,科学数据库作为科技共享平台的重要内容,开始在全国组织部署。

科研活动信息化(即 e-Science)已成为现代科研的趋势。通过高速网络联接的高性能计算资源、科学数据资源、数字图书馆、野

外观测台站、大型共用科研装置、计算模拟软件工具等等,形成支撑科研活动的协同工作环境,实现资源集成共享,构成了 e-Science 的基础设施。

科学数据在信息化的科研活动中起到了基础作用。许多科学技术领域都十分依赖数据。随着信息技术的发展,这种依赖大大加强了。新型观察手段,例如遥感、传感器网络、大型科学装置等,使人们在实验观察中得到越来越多的数据,基于超级计算和海量数据存储技术可以对这些极为复杂的海量数据进行分析处理,使得人们可以在更加深入、更加精细和多种尺度水平上发现新现象、认识科学规律、进行工程设计。这极大改变和拓展了传统研究方法,对当代科学技术前沿发展起着不可替代的作用。

e-Science 固然需要传统科学数据(包括数字化的科学文献)的支持,但是传统的科学数据库,还不足以支持信息化的科研活动。我们需要发展适应于 e-Science 需要的科学数据库技术,也需要发展配合 e-Science 具体工作的科学数据库。这些数据在科研活动中产生、应用,又形成了新的积累,为后续研究提供基础,同时也为科学领域数据资源发展提供了不竭的源泉。基于这样的认识,就可明确我院科学数据库在国家层面的定位,突出我们的特色,并形成与国家部署的其它科学数据共享平台互为补充和配合的态势。

因此,我院科学数据库的建设应和 e-Science 紧密结合,与国家及我院科研工作的发展战略相一致,与各领域科研活动的信息化需求相适应。这样才能为我院乃至国家的科研工作有效服务,也才能得到持续发展的动力。

第二,科学数据库是一项需要长期积累的基础性工作,需要长期稳定的支持,也需要多方资源的集成,更需要科学数据库工作

者坚持不懈的努力。

回顾历史,我院对科学数据库给予了长期稳定的支持,使它得以不断发展,形成了今天的规模和队伍。科学数据库作为 e-Science 最重要的基础设施之一,是我院信息化工作的重要方面,应当继续得到长期稳定的支持。

历史经验告诉我们,数据库工作要与我院科研活动紧密结合才能持续发展。1991—2000 年 10 年间,我院对科学数据资源建设的总投资为 1 000 万元,21 个建库单位每年平均得到的资助仅为 4.76 万元,而这 10 年间科学数据库数据量从不到 3GB 发展到 725GB;2001—2005 年,科学数据库资源建设投资 3 000 万元,45 个建库单位每年平均得到的资助只有 13.3 万元,而数据量发展到了 16.6TB。显然仅靠院投入的经费是不可能取得这样成就的。事实上,这里有来自科研活动中各方面的经费的集成,更有数百名科学数据库工作者的巨大努力和奉献。如果说在科学数据库发展初期,主要是靠他们对科学数据库价值的信念和前途的责任感,后来则更重要的是真正体会到了科学数据库对科研活动的重要意义。不少数据库的发展是因为学科领域发展的需要,是建库单位自身发展战略的一部分。有的研究所领导明确表示,无论是否得到支持都要搞科学数据库。凡是这样的情况,研究所领导层面都给了有力的支持,数据库的发展和应用就越好。

数据库资源发展经费的多元化,不仅弥补了院直接经费投入不足,而且还建立了一种机制,它保证了所建设的数据库是科研工作最需要的,保证了研究所关切和持续支持的自觉性,这些是我院科学数据库健康发展最基本的条件。

第三,科学数据库多年来形成了专家委员会决策领导,科学数据库办公室和数据中

心提供运行管理和技术支持服务的行之有效的管理体制,是应当继续坚持和发扬的经验。

科学数据库专家委员会,不仅是一个专家战略研究和咨询评议机构,而且是一个系统工程总体规划、组织实施和运行管理的指挥部。我院科学数据库发展到今天,在规模上、质量上、技术上和应用上都取得了很大成绩,这和参加建设的单位及工作人员分不开,更和历届科学数据库专家委员会所发挥的作用分不开。特别要指出的是,专家委员会并没有掌握大笔经费,在院投入十分有限的情况下,能坚持不懈、努力推进科学数据库事业发展,是一种高尚的事业心和责任心在驱使,是我院科学数据库这面旗帜具有凝聚力的重要原因。

科学数据库办公室和数据中心在日常管理、技术支持、运行服务和制定标准规范、政策研究上,为科学数据库专家委员会的决策领导、组织实施、调度协调、规范管理起到了重要的支撑服务作用,得到了建库单位的认可。正是由于这种体制保证和各方协同努力,科学数据库才得以顺利和成功发展。

科学数据库现在已进入一个新的发展阶段,面临着重要的机遇和挑战。不仅要研究科学数据库如何围绕我院科研需求发展的问题,而且要提升到国家层面研究适应于基础科研活动问题,实行开放、联合的战略,为建设支撑国家基础科研的科学数据库做出贡献。要深入研究我院科学数据库在国家层面的地位、作用、特点、规律以及和其它行业领域科学数据平台的关系,充分发挥它对科研教育信息化的基础支撑作用以及对其它行业领域数据库的先导、源泉作用,制定科学的发展战略,组织好我们的队伍和工作。我院科学数据库一定会有更大发展并将发挥更大作用。

(相关图片请见封三)



中国科学院