

《国家“十二五”科学和技术发展规划》明确提出“要加强自主创新能力建设、优化科技资源配置、推进科技资源开放共享和高校利用”。创新共享模式、健全服务机制是当今科技资源管理领域关注的重点。2011年12月15日,本刊编辑部在京举行了“中国科技资源管理论坛”,对科技资源共享模式与服务创新的进行了探讨与交流(见本期相关报道),产生了较好的效果。为了引发更加广泛和深入的研究与思考,本期组织了“科技资源共享模式与服务创新”专题,交流实践经验,分享研究成果,共同推进我国科技资源开放共享的进程。

——编者

人地系统主题数据库建设与数据共享服务

廖顺宝² 孙九林² 杨雅萍² 王卷乐² 张金区² 杜佳² 宋佳² 朱华忠² 乐夏芳²

(1. 河南大学环境与规划学院, 河南开封 475004; 2. 中国科学院地理科学与资源研究所, 北京 100101)

摘要: 人地系统主题数据库以人口、资源、环境和社会发展为核心, 服务于人地系统基础研究、社会经济可持续发展和国家战略需求。主题库由自然资源及其开发利用、环境灾害及其治理、人口与劳动力、社会发展、地理背景、遥感及地表参数、典型示范区及全球数据等八大类, 共32个子库构成, 数据资源总量8.8TB左右。主题数据库共享服务平台由一个主节点和3个分节点组成, 除具有一般数据库服务系统所具有的数据检索、查询、下载等基本功能外, 还具有属性数据可视化、数据在线分析、区域可持续发展评价等特色功能。在建设和运行期间, 主题数据库为国家和区域资源开发, 环境保护, 可持续发展研究、分析、评价与决策, 国土资源规划, 生态系统过程监测与调控等领域提供了大量的数据服务。

关键词: 人地系统; 主题数据库; 数据共享; 数据质量; 标准规范; 数据可视化; 在线分析; 科学数据库

中图分类号: G250.74

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2012.01.002

Construction and Data Sharing Service of Thematic Database in Human-Earth System

Liao Shunbao^{1,2}, Sun Jiulin², Yang Yaping², Wang Juanle², Zhang Jinqin², Du Jia², Song Jia², Zhu Huazhong², Yue Xiafang²

(1.College of Environment and Planning, Henan University, Kaifeng 475004; 2.Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101)

Abstract: Thematic Database for Human-Earth System (TDHES) is database system based on population, resources, environment and social development. The mission and objective of TDHES are to provide data service for basic research of human-earth system, national economic construction and development strategy. TDHES is composed of 32 sub-databases, which are divided into eight classes: natural resources and their exploitation and utilization, environmental disasters and harness, population and labor force, social development, geographical background, remotely sensed data and land surface parameters, typical regional data, and global data. Volume of TDHES amounts

第一作者简介: 廖顺宝(1966—), 男, 博士, 教授, 硕士生导师, 主要研究方向: 地球科学数据产品研发及数据质量评价。

基金项目: 中国科学院“十一五”信息化专项项目“人地系统主题数据库建设与服务”(INF0-115-C01-SDB3-02)。

收稿日期: 2010年12月22日。

to 8.8TB. Besides online browsing or downloading of data, the Internet based service platform of TDHES can also provide some functions including visualization of attribute data, online analysis of data and assessment of regional sustainable development. During the period of its construction and running, TDHES has provided already lots of data services for resources development, environmental protection, land planning, and monitoring and control of ecological system process, and study, analysis, evaluation and decision-making on sustainable development,

Keywords: human-earth system, thematic database, data sharing, data quality, standards, data visualization, online analysis, scientific database

1 引言

数据是科学研究的基础,科学数据只有通过广泛的共享,才能最大限度地发挥科学价值、社会价值和经济价值,从而实现科学数据的增值。数据共享可以减少或避免国家为同一目的重复投入,实现投资效益的最大化^[1]。发达国家一直很重视数据共享,通过建立部门和行业数据中心,为社会公众提供数据共享服务,如美国航空航天局(NASA)1990年组建的分布式最活跃数据档案中心群(DAACs)、全球变化主目录(GCMD)、加拿大的地球观测网(CEONet)等^[2]。日本政府从1994年开始投入巨资建设政府部门、大学、科研机构的数据库和全国科研信息网络^[3]。同时,一些推动数据共享的国际组织也纷纷成立,如国际科学联合会(ICSU)领导下的世界数据中心(WDC)和国际科技数据委员会(CODATA)以及世界气象组织(WMO)等^[4]。

自20世纪80年代起,我国在不同层面、通过多种形式推动科学数据共享。中国科学院于1982年开始运筹规划“科学数据库”,1987年正式启动建设;1984年,我国加入国际科技数据委员会(CODATA);1989年,我国加入世界数据中心(WDC)系统;1999年,科技部启动了科技基础性工作专项,提出了国家级数据中心群的概念,并确定世界数据中心所属的9个中国学科中心为首先启动的对象;2001年,气象科学数据共享试点启动,标志着国家科学数据共享工程正式实施;2005年,科技部正式启动国家科技基础条件平台建设专项,科学数据共享是其六大组成内容之一。当前,科学数据共享已进入全面实施阶段,科学数据已经成为国际科技竞争力的战略资源^[5-6]。

科学数据共享是服务于科技创新的基础性工作,它是一项涉及机制、资源、标准、技术和服务的系统工程。在这个系统中,机制是保障,资源是基础,标准、技术是支撑,服务是最终目标,它们

都是科学数据共享工程必不可少的组成要素,对这些要素的分析和研究无疑将推动科学数据共享事业的进一步发展。本文以具有20多年建设历史的“人地系统主题数据库”为例,介绍其在数据资源建设、平台开发、标准规范研制以及共享服务方面的成果和经验,以期为科学数据共享工作的进一步推进提供参考和借鉴。

2 人地系统主题数据库建设

人地系统主题数据库是中国科学院信息化专项“科学数据库”在“十一五”期间支持建设的8个主题数据库之一,其前身为中国自然资源数据库,它是中国科学院科学数据库的专业库之一,从1987年底开始研建,到“十五”末期,已发展成为一个覆盖资源、环境、社会经济、人口等学科的大型综合性数据库,其软件也从最初的数据存储、查询发展到集数据存储管理、数据查询服务、数据定量计算分析、数据空间定位分析等为一体的多功能应用软件系统^[7]。为进一步优化数据应用环境,提升数据应用水平,“十一五”期间,中国科学院设立了“中科院数据应用环境建设与服务”信息化专项。“人地系统主题数据库”是该项目的重点任务之一。

2.1 人地系统内涵

地球上凡是有人类居住的地区,以人类和人类的经济社会活动为一方,以地球表层包括岩石圈、大气圈、水圈和生物圈等人类赖以生存的地理环境为另一方,组成了相互联系、相互作用的综合体,即人地系统^[8]。人地系统是由地理环境和人类活动两个子系统交错构成的复杂、开放的巨型系统,内部具有一定的结构和功能。人地系统是统一的社会-自然综合体,可分解为人口(P)、经济(E)、社会(S)、资源(R)、环境(E),简记为PESRE系统^[9]。人地系统协调发展是一项复杂的系统工程,必须采用定性与定量相结合的方法,充分利用计算机技术、数据库技术、GIS技术、网络技术、专家系统

和决策支持系统等技术，建立一套完善、高效、操作性强的技术支持系统。人地系统主题数据库正是这样的技术支持系统。

2.2 数据资源建设

人地系统主题数据库的建设立足全国、面向全球并关注重点区域，数据资源由自然资源及其开发利用、环境灾害及其治理、人口与劳动力、社会发展、地理背景、遥感及地表参数、典型示范区及全球数据等八大类数据构成。主题库的数据资源架构体系如图 1 所示。

“十一五”期间，通过集成整合相关数据资源，特别是国外数据资源和国内重点区域的数据，重点建设了 32 个子库，数据资源总量约计 8.8TB。

2.3 数据质量评价

地球科学数据的质量问题贯穿于数据获取、数据处理、产品加工、数据管理与共享等数据生命周期的各个阶段，其质量评价与控制是一项涉及面广的复杂而艰巨的工作^[10]。在人地系统主题数据库建设过程中，我们在全面关注数据质量的同时，把数据质量评价的重点主要放在数据产品加工上^[11]，并对 3 种常用的地学数据产品加工方法，即属性数据空间化^[12]、矢量数据栅格化^[13]和遥感制图^[14]中的数据质量评价问题进行了深入研究，研究成果在主题数据库的数据质量评价与控制工作中得到实际应

用。

2.4 标准规范研制

人地系统主题数据库是在原专业数据库的基础上重点支持建设的数据库类型，其目标非常明确，即面向人地系统基础研究、社会经济可持续发展和国家战略需求，建立以人口、资源、环境和社会发为核心的数据库体系，满足人地系统研究对数据的需求，为科学研究、经济建设和政府决策提供内容规范、质量可靠的数据服务和在线分析服务。因此，主题库的建设必须符合一定的标准或规范。

在总项目的标准规范框架下，主题库重点开展了《人地系统主题数据库数据分类与编码规范》《数据质量管理规范》《元数据标准》《元数据编写规范》《数据文档规范》《数据资源整合技术规范》《共享服务平台技术规范》《数据共享服务规范》和《数据资源采集与整理工作指南》等 9 项标准、规范的研制。它们不仅与总项目的标准规范兼容，而且内容具体，可操作性强，在主题库建设中都得到了实际应用。

2.5 共享服务平台研发

数据资源整合、集成是数据共享的基础，但只有同时研发一套支持数据共享的软件系统，才能实现数据的有效共享。人地系统主题数据库共享服务平台是一个分布式的数据共享平台，目前由一个主

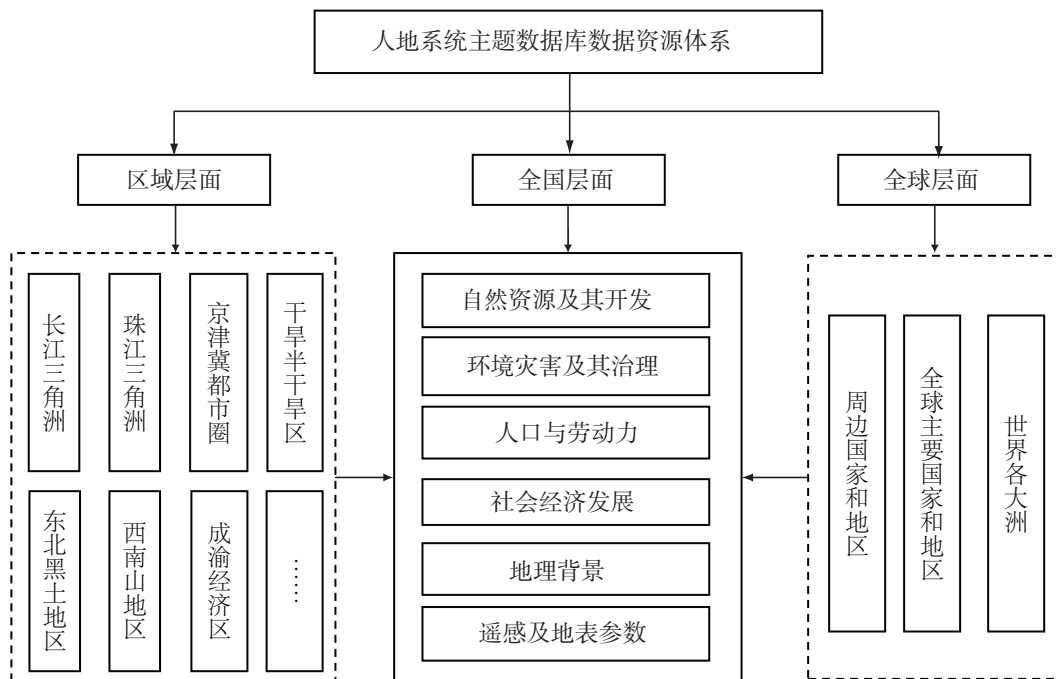


图 1 人地系统主题数据库数据资源架构体系

节点(主持单位)和3个分节点(参加单位)组成。平台除具有一般数据库服务系统所具有的数据检索、数据查询、数据下载等基本功能外,还具有属性数据可视化、数据在线分析、区域可持续发展评价等反映主题数据库特色的功能。

(1)数据检索、查询、下载功能。通过简单的注册,用户即可查询、下载所需数据。主题数据库服务系统提供了多种数据查询方式,具有多种图形浏览功能、统计图和专题图自动生成功能。图2是人地系统主题数据库服务系统的首页(http://www.data.ac.cn)。

(2)属性数据在线可视化功能。人地系统数据具有时空属性,可视化表达是人地系统数据最具特色的展示方式。根据人地系统研究的科学主题和关注重点,以需求为导向,以模型为驱动,主题数据库服务系统实现了数据的在线可视化与分析功能。目前,服务系统集成了我国省级1950年到2007年之间每年100多个数据项,可以对这些数据进行如下处理:①属性数据空间可视化;②基于时间序列的数据空间分布动画模拟;③数据的时间序列趋势分析;④地图和图表的联动显示;⑤不同省份之间的数据对比分析等。其操作及显示界面见图3。

(3)数据在线应用分析功能。主题数据库共享服务平台集成的数据在线应用分析功能包括:①根据坐标信息生成点状图并在线生成点Shape文件;②以行政区域为统计单元的数据,在线生成专

题图;③对带有坐标信息的数据进行空间插值和等值线生成;④各种数据的实时在线图表分析,形成包括柱状图、折线图、散点图、条形图、饼状图、面积图等图表类型;⑤在线计算数据之间的相关系数;⑥在线一元线性回归分析。

(4)区域可持续发展评价功能。人地系统协调与可持续发展是人地系统研究的最终目标,人地系统的可持续性由社会经济的发展水平、资源环境的数量、质量状态及其协调度所决定^[15]。建立可持续发展评价指标体系有助于掌握区域发展水平的动态变化规律与演变趋势,进而适时适度地调整区域发展方向,有针对性地制定达到特定发展水平的区域发展规划^[16]。因此,在人地系统主题数据库应用服务平台中设置可持续发展评价功能模块是人地系统研究的基本需求,它是一种集数据、模型和功能于一体的高级信息服务,是主题数据库区别于一般专业数据库的主要特征之一。区域可持续发展评价功能可以根据数据库内的数据项预先内置一些比较成熟的评价指标模型,用户通过选择这些模型并调用相关的数据即可得到相应的评价结果。同时,用户也可以根据数据库内的数据情况在线自行构建评价模型。不论采用哪种方式,评价结果都可以专题图的形式在线实时显示。

3 共享服务与成效

人地主题数据库服务系统对所有用户开放,任



图2 人地系统主题数据库服务系统界面

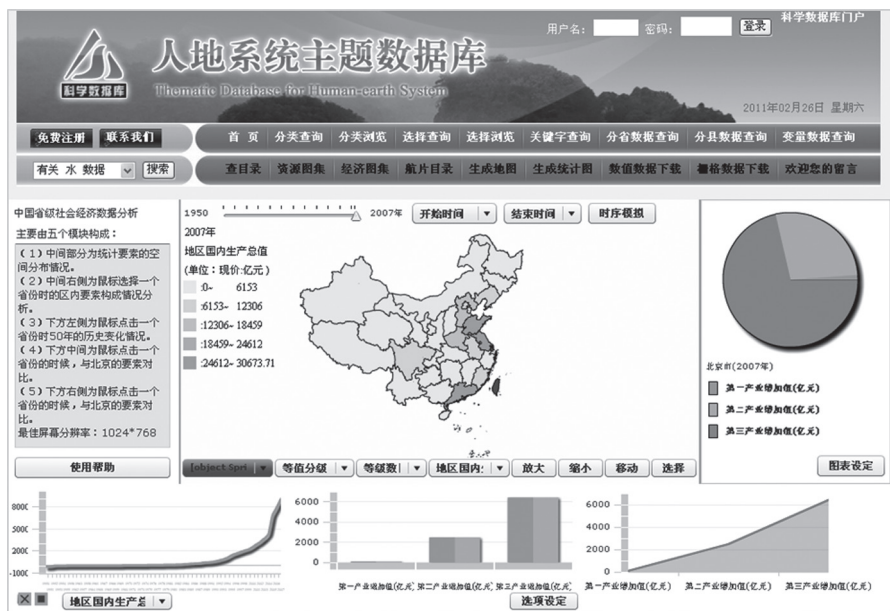


图3 数据在线可视化与时空分析界面

何用户均可通过数据服务网站查询主题库的元数据和数据目录, 用户可以获得在线、离线、定制以及数据在线可视化分析等多种形式的服务。在建设和运行期间, 人地系统主题数据库无偿为国家和区域资源开发, 环境保护, 可持续发展研究、分析、评价与决策, 国土资源规划, 生态系统过程监测与调控等领域提供了大量的数据服务, 取得了明显的社会效益。

自有用户注册记录以来至 2010 年底, 人地系统主题数据库实名注册用户 26934 人。“十一五”期间, 人地系统主题数据库网站总访问量近 112 万人次, 向科技界和社会公众提供了 4277GB 的数据服务量, 其中在线下载量 2954GB, 离线数据服务量 1323GB, 离线服务 1180 人次, 为 70 多项国家重大科研项目(课题)和重大工程项目提供了数据服务。根据中国知网 (<http://www.cnki.net/>) 的检索, 在“十一五”期间, 用户利用主题库提供的数据库发表了有主题数据库标注的学术论文 263 篇。主题数据库有力地支持了科技创新活动。

4 体会与思考

数据既是科技活动的产物, 又是支撑科技创新的基本资源。地学数据尤其如此。人地系统主题数据库以具有 20 多年建设历史的“中国自然资源数据库”为基础, 以人地系统研究的数据需求和功能需求为驱动, 通过“十一五”的建设, 目前已初具

规模, 并赢得了众多的用户。这既是主题数据库建设的成果, 也是进一步建设和完善主题库、更好地为用户服务的动力, 同时对主题数据库建设也有了更加深入的认识、体会与思考。

(1) 对人地系统内涵的认识与理解是主题数据库建设的基础。人地关系是地理学、资源科学甚至整个地球科学研究的基本对象, 其核心是协调人地关系, 人地系统以人口、资源、环境和发展(PRED)为其核心内容。基于上述认识和理解, 在已有工作的基础上, 通过对已有数据的时空序列更新、人地过程典型区域的数据资源整合以及国外相关数据的收集整理, 初步形成了由 8 个大类构成的人地系统主题数据库的数据资源体系及相应的数据内容。然而, 人地系统是一个非常复杂的巨型系统, 主题库的建设也不可能一蹴而就, 因此必须按计划、分步骤长期建设。

(2) 模型驱动是主题数据库服务系统的主要特征。一般数据库的服务模式是, 通过有效的数据组织, 把数据分门别类地呈现在用户面前, 用户能方便地获取所需数据。这种服务模式的不足是, 数据库有什么, 用户才能获得什么。显然, 这种服务模式不能满足主题数据库的服务需求。人地系统研究中经常要用到许多分析评价模型, 如区域发展的可持续性评价模型、工业化城镇化水平分析与预测模型等, 人地系统主题数据库服务系统应当具备这些功能, 即: 主题数据库服务系统不仅要提供直接的

数据服务,同时还要提供人地系统研究中经常要用到的模型分析服务,而且模型所用的数据应当来自主题数据库。

(3) 人才队伍是主题数据库建设和持续发展的根本保障。以“人地系统主题数据库”及其前身“中国自然资源数据库”为依托,在20多年的建设历程中,逐步形成了一支知识结构、专业结构、年龄结构合理的稳定、连续并不断壮大的人才队伍。这是主题库不断发展和完善的根本保证。

(4) 在保证日常服务的基础上积极开展主动跟踪服务。数据服务分为主动服务和被动服务(即日常数据服务)。在日常数据服务中,用户需要什么就给什么,数据服务人员难以进一步了解数据应用的成效和应用中出现的问题等反馈信息,不利于主题库的评价和进一步完善。主动服务是由服务人员与被服务项目保持密切联系,随时了解数据服务的效果和新的数据需求,有利于主题库的改进和完善。

参考文献

- [1] 黄鼎成. 科学数据共享的理论基础与共享机制[J]. 中国基础科学, 2003(2):22-27.
- [2] 诸云强, 孙九林, 廖顺宝, 等. 地球系统科学数据共享研究与实践[J]. 地球信息科学学报, 2010, 12(1):1-8.
- [3] 诸云强. 地球系统科学数据共享关键技术研究[D]. 北京: 中国科学院地理科学与资源研究所, 2006.
- [4] 北京: 王卷乐, 孙九林. 世界数据中心(WDC)中国学科中心数据共享进展[J]. 中国基础科学, 2007(2): 30.
- [5] 傅小锋, 李俊, 黎建辉. 国际科学数据的发展与共享[J]. 中国基础科学, 2007(2): 30-35.
- [6] 李娟, 刘德洪, 江洪. 国际科学数据共享现状研究[J]. 图书馆建设, 2009(2): 19-25.
- [7] 孙九林, 李泽辉. 中国自然资源数据库[J]. 现代图书情报技术, 1997(4):9-14.
- [8] 毛汉英. 人地系统的机理与调控[EB/OL].[2010-11-15]. <http://wdcrre.geodata.cn/Portal/wenxian/human4.pdf>.
- [9] 赵明华, 韩荣青. 地理学人地关系与人地系统研究现状评述[J]. 地域研究与开发, 2004, 23(5):6-10.
- [10] 廖顺宝. 地球系统科学数据资源体系研究[M]. 北京: 科学出版社, 2010, 11.
- [11] 廖顺宝, 白燕, 曹俊, 等. 地学数据加工处理中的数据质量评价[C]// 科学数据库与信息技术论文集(十). 北京: 兵器工业出版社, 2010, 6: 385-392.
- [12] 廖顺宝, 张赛. 属性数据空间化误差评价指标体系研究[J]. 地球信息科学学报, 2009, 11(2):176-182.
- [13] Liao Shunbao, Bai Yan. A New Grid-Cell Based Method for Error Evaluation of Vector-to-Raster Conversion[J]. Computational Geosciences, 2010, 14(4):539-549. DOI:10.1007/s10596-009-9169-3.
- [14] 廖顺宝, 曹俊. 数字地图规范性质量评价探讨[J]. 浙江大学学报: 理学版, 2010, 37(5):583-587.
- [15] 温琰茂, 柯雄侃, 王峰. 人地系统可持续发展评价体系与方法研究[J]. 地球科学进展, 1999, 14(1):51-55.
- [16] 方创琳, 毛汉英. 区域发展规划指标体系建立方法探讨[J]. 地理学报, 1999, 54(5):410-419.