

# 浙江省科技统计特色数据库建设与管理探索

刘君 张乐萍

(浙江省科技信息研究院, 浙江杭州 310006)

**摘要:** 科技数据是反映国家科技创新、经济增长和社会进步的重要战略资源。然而, 目前我国科技统计指标体系不够完善、科技数据缺乏规范性, 难以满足广泛的科技统计需要并有效发挥科技发展决策支持作用。以浙江省通过科技统计特色数据库建设与管理为例, 介绍特色数据库建立的目标、建设思路、数据库指标分类与选择、数据库系统的设计与开发以及数据库系统的维护管理。特色数据库实现了科技统计信息的社会共享, 提高了统计为科技发展决策服务的快速支持能力。

**关键词:** 科技统计; 特色数据库; 信息共享; 统计指标; 浙江省

中图分类号: N37

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2012.06.005

## Development and Management of Characteristic Database for Scientific and Technological Statistics Studying the Case of Zhejiang Province

Liu Jun, Zhang Leping

(Zhejiang Science & Technology Information Institute, Hangzhou 310016)

**Abstract:** Data of science and technology are the important strategic resources that reflect the state of the innovation of science and technology, development of economic and society. However, the science and technology statistics indicator system is not perfect, and the data of science and technology are not normalized in China. Therefore, the extensive needs of science and technology statistics can't be met and the decision support for science and technology development is lacked. Based on the case of the development of characteristic database for Zhejiang province, the purpose of developing the characteristic database, the development strategies, classification and selection of the index of database, design, development and maintenance of databases are discussed.

**Keywords:** science and technology statistics, characteristic database, information sharing, index of statistics, Zhejiang province

### 1 引言

科技统计是现代统计的一个重要方面, 是适应人类社会实践活动和国家对科技工作管理的需要而产生和发展的<sup>[1]</sup>。科技数据是反映国家科技创新、经济增长和社会进步的重要战略资源, 随着社会经济的快速发展, 政府管理部门、企事业单位研究人员对科技数据的要求越来越高, 不仅要求科技统计数据

的准确性、时效性, 而且较强地依赖于系统、

完整、准确、及时的科技数据、资料和相关信息。

目前, 我国科技统计采用的是由华中科技大学管理学院科技统计信息中心提供的离线统计系统来完成机构年报和国家级计划项目执行情况年度调查。其他各省的科技统计也大都采用这套离线统计软件完成统计工作。河北省对科技数据共享问题进行了研究并建立了科技统计数据库<sup>[2]</sup>。广东省在科技统计系统的建设中走在全国的前列, 拥有自己的科技统计软件和数据库, 能实现对高技术产品调

第一作者简介: 刘君(1970- ), 女, 浙江省科技信息研究院副研究员, 主要研究方向: 科技统计分析评估, 科技政策研究。

基金项目: 浙江省软科学研究项目“浙江省软科学研究数据库构建研究”(2007C20016)。

收稿日期: 2012年8月9日。

查、工程技术研究开发中心调查、专业镇调查、发展基金调查、地方财政和开发贷款调查等科技统计功能。科技统计数据库在建设过程中遇到的关键性问题通常有以下几个方面。

一是尚未建立较完善的科技统计制度，现行科技统计运行机制仍是计划经济体制下形成的部门统计管理模式。多数部门的统计是根据部门管理工作的需要开展的，一旦其管理工作思路有所变化，有些统计指标甚至统计项目也随之改变，致使不少统计指标不能稳定、数据不能连续；一些部门没有开展规范化统计，往往根据不同需要提供不同数据。因此，在收集数据时，只能选稳定、连续的数据作为入库的条件。

二是科技统计分散在多个系统中，每个系统中的相关统计数据又可能分散在多个部门和多种统计项目中。将目前多门、多头管理状态下的所有统计数据汇总起来是不现实的，这也是与统计法律法规制度相违背的。目前，科技统计数据主要分布在统计行政管理系统（包括企业相关统计）、科技行政管理系统、教育行政管理系统、国防科工委系统，各部门没有提供数据的责任和义务，获取数据的难度较大，往往要开展有代价的协作。

三是由于多系统、多部门、多种类的统计数据指标缺乏规范性和统一性。统计口径不统一与统计范围不协调的现象比较多，造成数据之间不能兼容，科技统计资料难以综合利用和实现信息共享。进行数据汇总规范数据口径的难度非常大，往往需要做细致的数据分析判别与数据调整工作，这是数据库构建工作中的一个重点与难点。

当前很多行业和领域的数据库服务都开始开展特色数据库建设工作，并在服务的多样化和个性化方面取得了一定的成果<sup>[3-5]</sup>。所谓特色数据库，是指依托馆藏信息资源，针对用户的信息需求，对某一学科或某一专题有利用价值的信息进行收集、分析、评价、处理、存储，并按照一定标准和规范将其数字化，以满足用户个性化需求的信息资源库。特色数据库实现了数据库服务从单一化向多元化拓展的发展模式特点<sup>[6]</sup>。

因此，浙江省为更好地满足政府管理部门、企事业单位研究人员及社会各界对科技统计数据的需求，通过建设全省科技统计特色数据库并对数据进行有效管理，实现了科技统计信息的社会共享，提高了统计为科技发展决策服务的支持能力。

## 2 数据库的建设思路

浙江省科技统计特色数据库指标系统结构及数据收集、选取的基本思路有以下几点。

(1)充分体现建库目的要求。浙江省科技统计数据库建设是为政府管理部门提供科技决策支持、为社会公众提供数据查询服务，是为了全面系统地描述浙江省科技工作的发展，以数据的形式、多角度地反映浙江省科技活动的总体面貌。因此，数据库系统建设既要考虑到现有统计工作情况，也要考虑到统计工作的发展完善；系统指标设置既要包括现成的、可稳定获取的统计数据的指标，也要包括应该开展统计而目前暂无数据的预设指标。

(2)构建一个完整的数据库系统结构。科技作为非独立的生产要素已日益渗透到经济和社会生活的各个领域，成为推动社会生产力发展和现代社会进步的决定性力量。基于国家统一制定的地方科技主要指标数据库，浙江省科技统计特色数据库结合浙江特色，将各类指标组成一个有机联系的系统，以科技活动为主线，按科技活动的阶段性及科技活动的相关要素设定指标体系的层次。构建的基本层次为：科技活动资源与投入（人力、财力、物质等资源）、科技活动进展（项目数量、类型、进展、完成等及相关活动数量）、科技活动产出（成果、奖励、专利、论文、技术转让、新产品等）、科技活动主体（院所、高校、企业等）、科技促进经济社会发展（高新产业及相关内容等）、科技综合水平位次等。其中，科技活动主体按科研院所、高等学校、企业、其他主体等分类展开，反映各类活动主体的发展情况及其科技活动情况。除了各类科技指标，还增设了相关的社会经济发展指标作为附件，内容包括国民经济、工农业生产、人口、物质文化生活、生态环境等方面。

(3)以权威稳定的数据来源作为保障。为使数据库指标系统综合反映浙江省科技发展状况，数据库系统指标分为浙江全省、浙江各市县、全国及各省的数据，并将各宏观与中观指标的数据按年度汇总后组成时间序列数据，形成展示动态变化的时间数列。浙江全省及各市县科技数据以浙江省统计资料数据为主，包括历年积累的各类原始统计数据整理、浙江科技统计年鉴、浙江科技统计数据（小册子）等；全国及各省科技数据以中国科技统计年鉴、中国高技术产业统计年鉴、中国统计年鉴等权

威性的统计资料为主，其他所需数据从“地方科技主要指标数据库”中收集或选择部分重点省市开展协作。国民经济和社会发展的相关数据，主要取自中国统计年鉴、浙江统计年鉴等权威性统计资料。在时间序列上收集2000年开始的相关数据，并将时间序列数据追溯到上世纪90年代，延续到最近期。偶尔一次性开展的统计数据一般不放入数据库，除非特别有意义的数据才考虑以适当方式进入数据库，以提高数据的准确性，使数据能反映事物的动态发展变化。

(4) 制定数据收集与数据库维护、管理工作制度。一是制定数据收集渠道建设方案（包括协作关系、需开展统计内容的建议、协作经费渠道等）；二是制定数据库管理、维护和使用的制度，包括落实数据库管理人员力量、数据库数据修改权限与方式、新数据收集更新与历史数据补充收集要求、数据的长期保存及保存方式、数据保密要求与使用限制、数据深层服务要求等。数据保存采用在线与离线相结合的形式；数据库数据修改权限仅限于数据库管理人员；建立持久稳定的数据库维护经费（包括数据收集经费、管理运行经费和管理人员经费）渠道。

### 3 数据库指标分类与选择

构建的浙江省科技统计特色数据库指标系统分

大类主题层、小类主题层、数据范围层及指标层4个层次，综合反映浙江省科技和经济社会发展状况（图1）。

#### (1) 大类主题

指标大类主题参照科技活动主线、以数据库的基本层次结构进行设置，并反映在数据库查询目录中，其本身无具体数据指标。由于在整个数据库指标体系中，具体的数据表（即小类主题）有近百张，在众多数据表名称中要查找到所需数据表并不容易，因此需要将数据表目录进行分类排列，为此设立大类主题。浙江省科技统计数据库最终确定了12个科技大类和5个社会经济发展大类，详见表1。

#### (2) 小类主题

小类主题反映大类主题的具体要素，是设置数据表的主要依据。以大类“科技人力资源与人力投入”为例，具体的小类主题有科技活动人员、R&D人员、国有企事业单位专业技术人员、科技人力资源培养等小类主题，从小类主题中又可以派生出更具体的主题，如研究机构R&D人员、大中型工业企业R&D人员、科技活动科学家与工程师、R&D科学家与工程师等。一般情况下，浙江全省性数据的数据表可以直接由小类主题作为数据表的主题词（标题）进行检索，浙江省各市县和全国各省数据的数据表由小类主题派生出的具体指标作为数据表的主题词进行检索。浙江省科技统计数据库最终确定

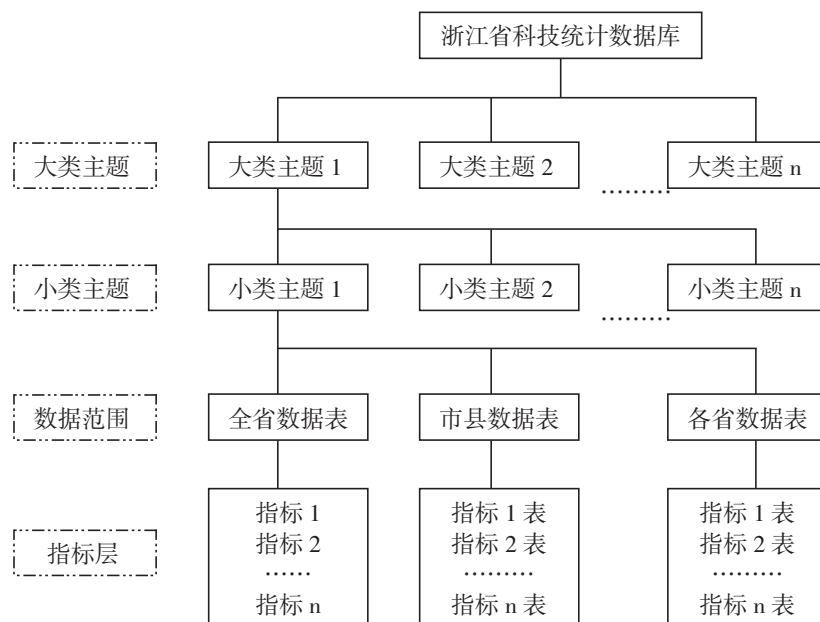


图1 浙江省科技统计特色数据库结构示意图

表1 浙江省科技统计特色数据库指标系统分类主题目录表

一、科技人力资源与人力投入	<ol style="list-style-type: none"> <li>1.科技活动人员</li> <li>2.R&amp;D活动人员</li> <li>3.国有企事业单位专业技术人员</li> <li>4.科技人力资源的培养</li> </ol>
二、科技活动经费投入与科研仪器设备	<ol style="list-style-type: none"> <li>1.科技活动经费筹集</li> <li>2.科技活动经费支出</li> <li>3.R&amp;D经费支出</li> <li>4.地方财政科技拨款</li> <li>5.科研仪器设备</li> </ol>
三、科技活动进展	<ol style="list-style-type: none"> <li>1.国家各类计划项目立项情况</li> <li>2.国家973、863、支撑计划项目执行情况</li> <li>3.国家火炬计划项目执行情况</li> <li>4.国家星火计划项目执行情况</li> <li>5.国家科技型中小企业创新基金项目执行情况</li> <li>6.国家自然科学基金项目执行情况</li> <li>7.国家973、863、支撑计划项目跟踪调查情况</li> <li>8.省级科技计划项目立项情况</li> <li>9.省级重大科技专项（优先主题）项目执行情况</li> <li>10.省级重大计划项目跟踪情况</li> </ol>
四、科技活动产出与技术交易	<ol style="list-style-type: none"> <li>1.科技成果</li> <li>2.科技奖励</li> <li>3.科技论文与著作</li> <li>4.专利申请</li> <li>5.专利授权</li> <li>6.工业新产品</li> <li>7.制定完成技术标准</li> <li>8.技术市场交易合同数</li> <li>9.技术市场交易合同金额</li> <li>10.技术市场技术流向地域（合同数）</li> <li>11.技术市场技术流向地域（合同金额）</li> <li>12.国外技术引进合同</li> </ol>
五、政府部门属研究与开发机构	<ol style="list-style-type: none"> <li>1.县以上政府部门属研究与开发机构</li> <li>2.县以上政府部门属自然科学技术与研究与开发机构</li> <li>3.县以上政府部门属科技信息与文献机构</li> <li>4.县属研究与开发机构</li> <li>5.科研院所转制机构</li> </ol>
六、高等学校科技指标	<ol style="list-style-type: none"> <li>1.高等学校全部科技活动人员与经费</li> <li>2.高等学校自然科学领域科技活动与成果</li> <li>3.高等学校自然科学领域科技活动机构</li> <li>4.高等学校自然科学领域国际科技交流</li> <li>5.高等学校自然科学领域技术转让与专利出售</li> </ol>
七、工业企业科技指标与民营科技企业	<ol style="list-style-type: none"> <li>1.规模以上工业企业科技活动人员</li> <li>2.规模以上工业企业科技活动经费</li> <li>3.规模以上工业企业科技成果</li> <li>4.规模以上工业企业科技项目</li> <li>5.规模以上工业企业其他技术活动经费</li> <li>6.规模以上工业企业科技机构</li> <li>7.大中型工业企业科技活动人员</li> <li>8.大中型工业企业科技活动经费</li> <li>9.大中型工业企业科技成果</li> <li>10.大中型工业企业科技项目</li> </ol>



	11.大中型工业企业其他技术活动经费 12.大中型工业企业科技机构 13.民营科技企业情况
八、其他科技创新主体	1.国家工程技术研究中心 2.重点实验室和中试基地 3.引进共建科技创新载体 4.区域科技创新服务中心（生产力促进中心） 5.科技企业孵化器 6.重大科技创新平台
九、高技术产业（包括高新技术企业）	1.高技术产业规模以上企业基本情况 2.高技术产业规模以上企业科技指标 3.高新技术企业 4.省级高新技术企业研究开发中心 5.高新技术产业主要指标
十、高新技术产业开发区（园区）与特色产业基地	1.国家级高新技术产业开发区建设情况 2.国家级高新技术产业开发区企业经济指标 3.国家级高新技术产业开发区企业科技指标 4.国家级高新区企业技术创新主要指标 5.省级高新技术产业园区主要指标 6.省级高新技术特色产业基地
十一、科协与科普	1.科协系统科技活动情况 2.科协系统省级学会情况 3.全省科普事业与科普活动情况
十二、科技综合评价水平位次	1.全国科技进步统计监测评价指数与位次 2.中国区域创新能力评价综合指标与位次
附一、自然地理基本条件	1.自然地理基本条件
附二、国民经济发展主要指标	1.国民经济主要指标 2.国民经济主要指标对上年增长速度
附三、工农业基本指标	1.规模以上工业基本情况 2.大中型工业企业基本情况 3.农业基本情况
附四、人口与人民物质文化生活	1.人口、就业与专业技术人员 2.居民收入、支出与生活质量 3.文化、教育和卫生
附五、生态与环境保护	1.环境污染状况 2.环境污染治理 3.生态环境保护

了88个小类主题，详见表1。

### （3）数据范围层

在选择小类主题的数据时，既需要反映全省的整体数据，也需要反映各市或各市县的数据，也需要反映全国及各省市的数据。这些不同范围的数据不适宜在一张数据表中显示，需要分别配置全省数据表、市县数据表和各省数据表这3种数据表来反映，这就在小类主题下构成了一个数据范围的层次。这3种范围的数据表配置格式为：全省数据表中综合反映本小类主题的各项指标；市县数据表一般选择本小类主题中的重要指标，按每个指标配置

独立的数据表；各省市数据表的配置方式与市县数据表相同。浙江省科技统计数据库共创建数据表438个。

### （4）指标层

指标层为描述小类主题及其派生主题的各项指标，即各张数据表中的指标，具体指标详见浙江科技统计网。

## 4 数据库系统软件的基本功能

浙江省科技统计特色数据库研究开发了基于Excel的数据库软件，已实现数据处理和管理的一体

化,能够满足统计网上数据处理、管理和发布的需要,目前已在浙江科技统计网上上线使用。数据库系统软件的基本功能有以下4个方面。

(1)网上在线发布功能:数据库以浙江省科技统计网作为建设平台,正式入库的数据均已在统计网上显示,并将及时更新数据,实现了科技统计数据的外发布和社会共享。用户登录科技统计网后选择“科技统计数据”模块即可查询数据,也可以复制数据库的数据,或将整个数据表导出。数据表查询方法是根据前述数据库指标体系,采用逐层展开方式查询,查到所需的数据表名后即可显示数据表数据。

(2)数据查询与分析功能:系统根据数据库的内容编制关键词,设置了根据关键词查询数据的功能,以便用户快速查询到数据;设置了数据简单分析功能,选择数据后可方便地绘制出简单的统计图表,如柱形图、条形图、曲线图、散点图、饼图等,同时也可以根据需要截取表中的不同字段制作新的报表,满足用户的简单分析需求。

(3)数据库后台管理功能:数据库后台是数据库数据整理加工的平台,其功能主要有数据表目录调整与修改、数据表增删、数据表名及指标修改、数据年份增删、数据录入与修改、数据更新与修改后的上网发布、数据表复制备份等各种功能,满足了数据库管理维护工作的需要。

(4)个性化服务功能:针对不同类别的用户,在系统预配置功能中进行用户属性特征的登记和配置。系统在各相关功能点,根据预先配置的用户属性和需求实现相应的功能并推送相应的个性化信息。

## 5 数据库系统维护及管理

(1)建设数据库数据来源渠道。科技统计特色数据库包括400余个数据表,每个数据表都有若干个指标,因此涉及指标数据庞大,只有明确每个指标的数据来源才能有效地更新数据。根据每个指标的数据来源,对于来源于统计年鉴和现有统计报表的数据,清楚了解各统计年鉴的出版时间和统计报表的报送时间,不定时更新数据库数据;对于目前尚未开展统计而数据库中需要的指标,通过和相关单位的协商努力开展长期、稳定的协作关系,确保数据的获取。

(2)设置数据库维护权限。科技统计特色数据

系统是基于Excel的数据库软件,数据更新方式是先将数据更新到线下的Excel表中,再批量或单个上传至网站,而数据的更新又是不定时的,因此,必须落实数据库管理人员,负责数据库的日常维护。为保证数据库后台管理的安全性,系统设定数据库数据修改权限仅限于管理员,负责历史数据补充收集、新数据收集更新、数据表线上更新等工作。

(3)及时做好数据备份。数据库的破坏是难以预测的,如软件或硬件系统瘫痪、人为误操作、存储数据的磁盘被破坏等,为此必须采取能够还原数据库的措施,而数据备份是在数据库修复中最为重要的手段。浙江科技统计特色数据库数据保存采用在线和离线相结合的方式,除了采用服务器数据备份及按需Excel表导出导入的备份和还原方式外,还采用计算机存储、移动工具存储、网盘存储等物理存储方式,多方式长期保存数据。

## 6 结语

浙江省科技统计特色数据库及管理平台于2008年建成,建立在科学的统计指标体系之上,充分保证数据来源的权威性和稳定性。完善的数据库建设运行和管理保障了数据库的良好维护性,实现了科技统计信息的共享,并为科技发展决策提供了良好的服务,用户涵盖了浙江省科技厅、全省地市科技部门及相关企事业单位和个人。运行几年以来,特色数据库为浙江省科技统计工作发挥了很好的支撑作用。

## 参考文献

- [1] 孙学范.科技统计学[M].北京:中国人民大学出版社,1994.
- [2] 胡宝民,李子彪,徐大海,等.河北省科技统计数据库构建及共享平台建设[J].科技管理研究,2005(12):70-73.
- [3] 贺晋霞.青藏高原文化特色数据库建设的思考[J].青海师范大学学报:哲学社会科学版,2011,33(6):153-154.
- [4] 王金秀,张洁.湖北省高校图书馆特色数据库建设现状及发展策略研究[J].情报探索,2011(12):74-77.
- [5] 陈全平.我国特色数据库建设问题与对策[J].图书馆理论与实践,2010(12):14-17.
- [6] 方旭,梁洁,文庭斌.特色数据库个性化服务模式的推广探析[J].科技情报开发与经济,2010,12(6):35-37.