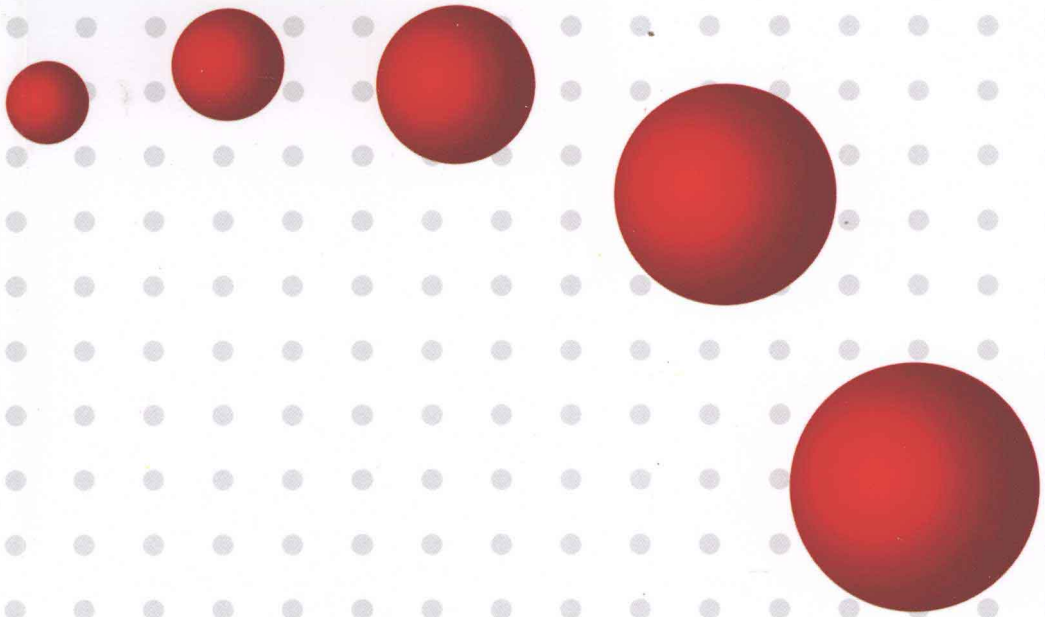
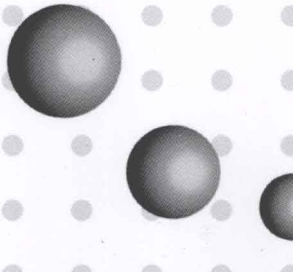


# 科技资源数据库建设的理论与实践

刘明生　刘　辉　李建华　著



科学出版社

# 科技资源数据库建设的理论与实践

刘明生　刘　辉　李建华　著

科学出版社

北京

## 内 容 简 介

科技资源数据的生产、收集、共享、服务与分析是科技资源数据库建设的重要环节。本书围绕这些环节，重点讨论科技资源数据库资源数据规范、科技资源数据库元数据管理、资源数据整合与共享、数据发现与服务、资源数据安全和资源服务的理论、技术和方法，并提供实现这些功能的程序代码片段。本书的特点：依托省级科技资源数据库建设项目，阐明的理论系统，采用的技术先进，提供的案例真实，摘录的代码实用。

本书可供科技资源数据库系统开发人员和信息系统项目管理人员借鉴，也可作为软件技术、计算机科学与技术、管理信息系统和信息管理类专业高年级学生和研究生的教学参考书。

**图书在版编目（CIP）数据**

科技资源数据库建设的理论与实践/刘明生，刘辉，李建华著．—北京：科学出版社，2012

ISBN 978-7-03-034226-3

Ⅰ.①科… Ⅱ.①刘…②刘…③李… Ⅲ.①科学技术-数据库-资源建设-研究 Ⅳ.①G311

中国版本图书馆 CIP 数据核字（2012）第 086675 号

责任编辑：任加林 戴 薇/责任校对：王万红
责任印制：吕春珉/封面设计：耕者设计工作室

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
双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2012 年 6 月第 一 版 开本：787×1092 1/16
2012 年 6 月第一次印刷 印张：17 3/4
字数：373 700

**定价：52.00 元**

（如有印装质量问题，我社负责调换〈双青〉）
销售部电话 010-62134988 编辑部电话 010-62137026（BI08）

# 前　言

科技资源数据作为一项重要的基础性战略资源，在科学研究、经济建设、社会事业发展等方面扮演着重要的角色，发挥着不可替代的作用。正因如此，科技资源数据库的建设工作历来受到各国政府的高度重视。美国、加拿大、欧盟各国、日本、韩国等和新兴工业化国家始终把科技基础资源平台建设放在极其重要的地位，将其视为实现创新能力培育的国家目标的重要手段，加大投入力度，在国家战略和相关规划、政策中予以高度关注、重点倾斜。我国科技资源基础平台建设工作起步于20世纪80年代初，以中国科学院组织实施的科学数据库建设项目为标志。2004年7月，国务院办公厅转发科技部、国家发展和改革委员会、财政部、教育部制定的《2004～2010年国家科技基础条件平台建设纲要》把科技资源数据库建设列为科技基础条件平台建设的重要组成部分，从而拉开了全面系统地建设基础条件平台的帷幕。

科技资源数据存放地域是分散的，组织形式是多样的，部署平台是异构的，资源内容是变化的。科技资源数据的上述特点，给资源数据的整合、共享和服务带来了重重的困难。幸而，现代信息技术、信息管理技术的快速发展和全面应用，为科技资源数据深度整合、全面共享和有效服务提供了前所未有的机遇和科学的解决方案。

在2004年国内兴起的科技资源数据库建设大潮中，本书的作者有幸成为河北省科技资源数据库建设工作的倡导者、组织者和建设者。作为倡导者，与河北省科学技术厅等政府部门发起了河北省科技资源数据库建设工作，并得到他们的大力支持，从2005年起立项建设并获得连续五年的资助；作为组织者，带领十个专业科技资源数据生产、建设部门的百余名同行，建成了河北省科技资源数据库中心和二十余个专业科技资源数据库；作为建设者，研究信息管理理论，运用先进信息技术，设计了河北省科技资源数据库总体架构，制定了科技资源数据规范、核心元数据规范、专业元数据规范和数据共享、服务规范和相应的服务应用策略。作者在科技资源数据库建设中，可谓学习理论心得颇丰，开发应用教训多多，于是萌发了用文字记录这些心得和体会的动机，以便使成功之处、正确之处供同行们分享和借鉴，失败之处、不妥之处供业界警示和批评，这就是我们撰写本书的目的所在。

科技资源数据库建设涉及科技资源数据的生产、收集、共享与分析等环节，本书以河北省科技资源数据库建设为例，针对上述环节分七个章讨论相关的理论、技术和方法。第一章介绍了对建设科技资源数据库的认识，列举了支撑科技资源数据库建设的理论、技术和方法；第二章讨论了科技资源数据规范、核心元数据规范、专业元数据规范；第三章研究了元数据管理理论和方法，以及元数据注册系统的开发；第四章探讨了在科技资源数据库建设中，如何以资源核心元数据和全文元数据为基础，利用SOA架构技术实现资源异构数据的深度整合，奠定了资源共享和服务的基础；第五章

针对资源数据几种常见的组织形式，即结构化文本数据、非结构化文本数据（Web页）、图像数据、空间数据，设计了资源数据发现和服务的有效手段，为充分发挥资源数据的作用提供了技术保障；第六章以科技资源数据库中心和各专业资源数据库中资源信息的安全等级为依据，提出了相应的信息安全策略，以保障资源数据的安全；第七章介绍了科技资源数据对外服务的策略、方法和手段，以满足科技资源数据库服务与应用的基本需求。

本书是在建设河北省科技资源数据库基础上完成的，没有项目的资助、没有同仁的支持和鼓励，本书不可能问世。为此，作者要向河北省科学技术厅以及项目各建设单位的同事们表示诚挚的敬意！

本书获得河北省教育厅学术著作出版基金资助出版，对河北省教育厅致以衷心的感谢！

# 目　录

# 第1章 绪 论

科技资源数据库作为科技基础条件平台建设的重要组成部分，是创新体系中急需发展的现代科学基础设施。建设科技资源数据库要本着政策调控、法规保障、统一规划、上下协调、平台开放、资源共享的原则，应用现代信息技术，整合离散的科技资源，构建面向全社会的共享服务体系，实现对自然科技资源和科学数据信息的规范化管理和高效利用，为科技创新、政府决策、经济增长提供数据信息资源的保障。

## 1.1 科技资源数据库建设目标和内容

建设科技资源数据库，就是要根据经济建设和社会发展对科技的总体需求，对科技资源进行有效整合、共享、完善和提高，有效激活存量资源，调控增量资源，最大限度发挥现有资源的潜能，逐步实现科技基础条件资源的优化配置，建设以科技资源数据库管理中心和科技资源数据库网络平台为核心的科技资源数据共享服务体系。

### 1.1.1 科技资源数据库建设现状

美国在科学数据资源建设方面一直保持着遥遥领先的优势。20世纪70年代以来，美国科学数据积累迅速增加。据估算，数据库总量一直占据全世界总量的一半以上。到1975年，美国开发177个大型数据库，主要服务目标是政府决策和政府启动的重大科研项目。20世纪80年代末美国利用“完全与开放”的数据共享政策作为美国联邦政府在信息时代的一项基本国策，通过数据的流动和应用激励美国经济的发展，确保美国在21世纪信息时代处于世界领先地位。2006年美国国家科学基金会（NSF）在《21世纪科学研究的信息化基础设施》报告中明确指出，在未来，美国科学和工程上的国际领先地位将越来越取决于在数字化科学数据的优势上，取决于通过成熟的数据挖掘、集成、分析和可视化工具将其转换为信息和知识的能力。因此，他们在国内机构层面、国家层面及国际层面上采取了一系列的行动计划，目的是建立全球性的科学研究数据的公共访问机制，为美国科学研究和经济建设服务，并保持美国在科学研究和工程技术领域的领先地位。

目前，国际上众多国家和国际组织已经认识到开放共享科学和工程数据所带来的广泛社会、经济和科学效益。科学数据的长期积累、开发应用、共享，已经逐步在政府、科学界和社会形成共识。2007年1月，英国科学与创新办公室发布了针对“2004～2014科学与创新发展规划”的需求分析报告，其6个子调研组中有4个与科学数据密切相关。报告建议英国针对基础科研领域的科研项目启动数据资源数字化的长期保存与共享建设，建设国家级永久性信息基础设施，重点建立国家级的超大规模科学数据仓

库，协调现有国家、地方、科研院所、其他相关者等关系，形成强大的数据服务能力。日本在面向21世纪的产业技术开发战略中提出，建设高水平、高效率的先进研究开发设施。文部科学省专门制定了“国立大学等设施紧急整备5年计划”。韩国科技创新计划提出改善科技基础设施，使实验室设备现代化，为天才大学生建立地区科技研究中心。2002年，欧盟发表了名为“迈向信息社会：原则、战略和优先行动”的布加勒斯特宣言，提出对公共科学数据、公共当局持有的信息开放共享的公益性共享原则和指导思想。

在我国，中国科学院于20世纪80年代初启动了“科学数据库及其信息系统”重大项目。20多年来，已建成了内容涵盖化学、生物、天文、材料、腐蚀、光学机械、自然资源、能源、生态环境、湖泊、湿地、冰川、大气、古气候、动物、水生生物、遥感等多个学科，总数据量超过16TB的503个专业数据库系统，并建成了分布式的网络数据服务体系，这是我国目前学科覆盖面最广、设施先进、管理规范、数据积累丰富的综合性科学数据系统。

2004年7月，国务院办公厅转发了科技部、国家发改委、教育部、财政部联合制定的《2004～2010年国家科技基础条件平台建设纲要》（以下简称《平台建设纲要》）。国务院2006年2月9日发布的《国家中长期科学和技术发展规划纲要（2006～2010年）》，也用相当篇幅来阐述“加强科技基础条件平台建设、建立科技基础条件平台的共享机制”的内容。为贯彻落实《平台建设纲要》精神，加强国家科技基础条件平台建设，科技部联合财政部、国家发展改革委员会、教育部于2005年7月发布了《“十一五”国家科技基础条件平台建设实施意见》（以下简称《实施意见》），进一步凝练出了“六大平台”24个方面的重点建设任务，并在平台建设组织领导、经费投入、监督管理、共享服务等重要环节提出了若干重要措施。筹建国家科技基础条件平台建设领导小组，负责平台建设整体规划和相关政策法规的制定工作，对平台建设重大问题进行协商和协调，联合审定平台重大建设任务，组织跨部门、跨行业、跨地区科技基础条件资源的整合与共享工作。截止到2009年底，全国建成了6个行业科学数据中心、7个自然科技资源共享中心和3000余个科学资源数据库，可供共享的数据资源总量超过140TB，初步建成了遍布全国的分布式数据共享网络体系，注册用户达16万人，访问人次达6171万，数据下载量超过430TB。

在过去的几十年中，我国尽管在科技资源数据的信息化建设上积累了一定的基础，但从总体来看，目前所拥有的科技资源相关数据库不仅远远落后于世界发达国家的水平，而且在基础条件的建设上存在着比较突出的矛盾与问题。

（1）布局分散，重复建设。在科技资源基础条件的开发上，由于缺乏整体规划，一部分科技资源布局分散、重复建设，造成资源的浪费，不同学科或不同研究方向之间科技资源的分布和利用处于不平衡的发展状态。例如，在自然科学方面，重野生资源的开发利用而轻种质资源的保护等现象也比较突出。

由于缺乏集中、有效、系统的管理体制，国家及各省均尚未形成统一的管理和服务体系，相当部分的科技资源分散在有关管理部门和研究单位。据估计，目前我国约有数百万件关于技术方法和工艺的文献资源散落在各个研究单位，还有一些经过多年

积累的、具有相当学术价值的数据、文献等资料散落在个人手中。由于国家的科技报告体系还没有建立起来，已经形成的科技报告要么分散在各研究单位，要么流失掉。以上状况说明已有的资源并未得到充分利用，科技资源利用效率低下及浪费现象普遍存在。

(2) 投入不足，配置不当。绝大多数的资料、信息和数据仍采用传统的档案保存方式，信息检索仍以文本信息为主，以简单语法结构和非结构信息为主，与国际上信息储存和检索技术向多媒体内容储存、高精度检索、文本挖掘、知识发现、信息内容可视化、概念词库、知识搜索引擎、语音识别等领域转化形成了强烈反差。据统计，目前我国科技人员使用互联网流量的95%是访问国外站点。

由于投入不足，一些重要的科技资源因为经费短缺难以得到收集和利用，一些具有战略意义的科技资源基础条件与国际水平之间存在着很大的差距。一些新的开发与服务常常受到经济利益驱动，收费普遍较高，限制了科技资源的广泛利用。科技资源的基础条件愈来愈难以支持社会发展的客观要求。

(3) 管理落后，共享机制缺乏。科技资源的开发利用和管理按照纵向隶属关系运行，因此在科技资源基础条件的建设和使用上普遍存在着部门分割的现象，结果使有限的投入被多个渠道所分流，形成的有限的科技资源又为多个部门所拥有。这必然导致各部门只关注本部门的基础条件建设，对于部门以外的科技资源利用和全社会的科技资源共享并不关心，甚至有些个人和团体为了极端的个人和小团体利益，将种质资源、重要数据资料、标本私有出售，换取本部门经济利益的同时，却使资源遭受到难以挽回的损失。长期以来的分散建设和条块分割的管理模式使有限的科技资源难以得到充分高效的利用和在全社会范围内实现共享。

实现科技资源共享是高效利用科技基础条件的存量资源并为科技创新活动搭建平台的主要途径，不仅需要相应的制度支撑来改善落后的管理方式，而且需要一系列保证科技基础条件实现共享的政策法规和统一规范的标准。目前，我国在这些方面的建设还很欠缺，如国家对资源的发现者和创造者缺少知识产权保护；对优质资源的提供者和利用者利益分享问题也没有发挥出很好的协调作用。在自然科技资源的利用上，少数人保护、多数人破坏的现象也比较突出。我国的数据库由于缺乏标准化、规范化，能够为相关行业提供服务的有效支撑数据还不足数据库总量的1/10。

近几年来，尽管信息资源的共享和整合越来越多地得到国家有关部门的关注，但我国信息共享规模小，参与单位少，共享的服务能力十分有限。例如，科研设施和数据垄断、科研人员流动性差、缺少交流导致信息滞留的问题就比较突出。

总之，在我国科技基础条件建设中，存在着投入部门隶属关系的体制障碍，缺乏国家层面上的整体规划与统一布局，尚未形成稳定的财政投入渠道，运行经费缺乏和配置不当，指导科技基础条件建设的法律法规尚不健全，没有形成科技基础条件共建共享的运行机制，工作人员队伍不稳定、专业化程度低等诸多问题，目前科技基础条件建设的滞后导致我国科学发展的战略性研究经常受制于人，难以形成重大原始性科技成果，关键技术突破和系统集成能够创新亦难以完成，全社会的科技创新活动得不到科技基础条件的有效支撑。

### 1.1.2 科技资源数据库建设总体目标

科技资源数据库建设总体目标是，坚持资源公开与共用的方针，构建结构合理、面向全社会、网络化、智能化的自然科技资源与科学数据管理与共享服务体系；完善共享政策、法规体系和管理体系的建设，建立健全共享机制；培养一批能适应社会信息化的高素质的科学数据共享管理、技术人才。使科技数据资源的积累与共享达到基本满足科技创新和经济发展的需求，提高科技创新能力和竞争力，最大限度地发挥政府投入的效益。

（1）构建科技资源数据管理与共享服务体系。在政府驱动和宏观指导下，集成政府部门、科研机构、高等院校和相关组织等多方面的公益性、基础性科技数据资源，通过整体布局、资源重组、机制创新，构建资源体系完整、结构合理、技术统一、管理规范、服务能力强的科技资源数据共享服务体系。十年内，在人类遗传、植物种质、动物种质、微生物菌种、生物标本、生态环境、气候、农业土壤、农业经济等资源领域，以及针对重大科技计划和重点领域，构建多个科技资源数据中心。同时，构建科技资源数据库管理中心及其通过元数据技术与上述系统相链接的门户网站，形成面向社会统一、透明的科技资源数据服务体系。

（2）制定和完善科技资源数据共享政策、法规与标准体系。政策调控、法规保障和技术支持，是实现科技资源数据共享的基本保证与前提。要从省级层面上统一规划科技资源数据共享工程的技术框架并与国家资源平台对接，开展政策、法规和标准体系研究。形成统一的技术平台，形成一套行之有效的共享政策和部门规章制度；完成标准框架的制定并全面推动标准化建设。形成完善的政策、法规和标准体系，确保科技资源数据库的正常秩序和高效运转。

（3）增强科技数据资源积累，促进科技数据资源增值。形成科学研究项目科学数据（研究结果及其参照数据或运算数据和一些领域的原始数据）的有效汇交、管理和共享的新局面，不断增强科技数据资源的积累。在保证对现有数据资源的管理和共享服务的同时，组织对濒临损毁的珍贵的历史资料的抢救，进一步提高数据质量控制水平，增强科学数据资源的二次开发能力，及时组织高附加值、多源复合数据产品的生产，挖掘所需要的知识，形成有特色的服务体系。

### 1.1.3 科技资源数据库建设基本内容

科技资源数据库建设内容和工作任务主要包括科技资源数据库中心建设和专业数据库建设两部分内容。

（1）科技资源数据库中心建设。科技资源数据库建设是一个跨学科、跨单位，多部门和单位联合参与的信息化工程项目，为了保证整个数据库建设真正实现“逻辑上的高度统一”，整个建设工程通过建立政策机制、资源体系、标准规范和技术平台等来对各个专业数据库的建设提供支持、服务和约束，并且对整个工程的统一性与一致性提供管理和技术方面的保障。其主要内容包括：

① 共享政策与运行机制。完成共享政策与运行机制的制定与实施，建立科技资源信息化领域的数据共享管理机制，保证整个系统建设的整体协调、规范统一。

② 共享资源体系建设。在分析领域资源情况的基础上建立资源规划与管理体系，研究制定相关规范与机制，对共享资源进行合理规划，保证资源建设的规范化、可持续发展。

③ 标准规范体系建设。落实国家、行业和地方科技资源数据共享标准，在建立领域标准规范体系的基础上，进行领域专用标准的研究制定，整合领域现有标准并进行指导、培训和应用。

④ 技术平台体系建设。通过技术开发与应用，建设科技资源数据库共享网络门户网站，保证数据共享及各种服务的进行，并实现对数据中心和共享节点的管理与支持。

(2) 专业资源数据库建设。根据科学研究、经济建设、社会发展、政府决策的需要，建设覆盖面广泛的专业资源数据库，如农业种质、野生动植物种质、微生物菌种、人类遗传、中草药、地理、土壤、气候、环境、生态、交通、地质、矿产、农业经济等科学数据资源专业数据中心。

## 1.2 科技资源数据库建设功能定位

科技资源数据库通过整合和扩充现有的自然科技资源和科学数据，建立和完善以共享为核心的管理制度，稳定和培养技术支撑人员队伍，为经济建设、社会发展、科学决策和科研活动构建便利、充分的资源和信息共享服务平台。

### 1.2.1 总体功能定位

科技资源数据库为了满足应用需求，应提供如下的功能。

(1) 资源共享。科技资源数据库建设为全社会的科技创新活动提供普遍的公共服务，通过平台开发共享，打破科研条件和设备等部门和个人垄断，为所有愿意从事科研活动的人员提供科研活动的条件和场所，提供人才脱颖而出的沃土。

(2) 保护资源。自然科技资源和科技基础数据，具有一旦丢失和灭绝就难以恢复的特点，对科技和社会经济长远发展具有十分重要的意义。这些战略资源对科技研究来说具有经济学意义上的“公共产品”特性，即“共同受益和联合消费”的特点；同时，这类资源积累和维护需要长期、不间断的投入，而科研机构和研究者不愿或无力进行这方面投资，科技资源数据库建设使得这些资源得以积累和维护。

(3) 促进交流。科技资源数据库建设可以促进科技成果的交流、传播和扩散。科技资源数据库网络平台利用先进的网络化、数字化和多媒体信息技术，建立共享服务平台，将最新的知识创新成果和技术创新成果进行积累并交流和传播。从事科学研究、科技推广和科技成果产业化的各类人员，都可以及时有效地通过科技基础条件平台进行科技信息传播和交流。

(4) 提高效率。科技资源数据库建设可以实现国家科技投资的节约，提高科技资

源的利用效率，以合理配置和有效利用科技资源。

### 1.2.2 功能描述

科技资源数据库以科技资源数据中心为主节点，通过互联网与各资源数据中心连接。在信息技术的支持下，借助于公共信息基础设施，实现数据管理、目录服务、数据服务、数据交换服务等服务功能，如图 1-1 所示。

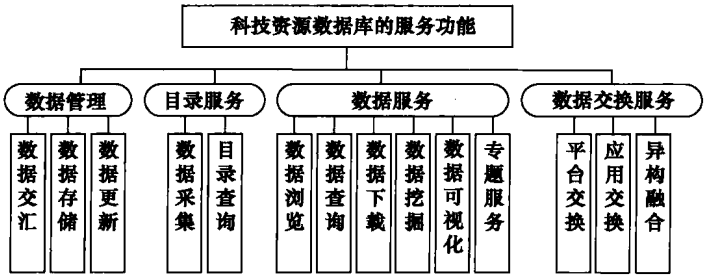



图 1-1 科技资源数据库总体功能

（1）数据管理。数据管理是对海量、多源、异构的数据资源进行有效的综合管理，涵盖了从数据采集到数据管理的多个环节。通过设计合理数据采集和管理过程，可以以最小代价将日常业务数据转化为供科技资源数据共享和决策支持服务的数据，挖掘数据的潜在价值。按照科技条件平台网络和科技资源数据所属部门的数据分布和管理特点，建立综合管理系统模型，明确数据资源科学、有效管理的需求，指导各领域科学数据管理和服务平台的规划和设计，提高各科技资源数据中心内部的工作效率，为不同科学数据资源的整合提供基础。

利用分布式数据库技术、元数据技术和网络技术，建立以分布式为主、集成式为辅的标准化数据管理系统，包括数据交汇、存储和更新等功能，实现对科技资源数据的有效管理。

（2）目录服务。目录服务是以元数据为核心的目录查询，是元数据系统利用元数据技术提供信息服务的一种标准模式，它通过元数据标准的核心元素将信息以动态分类的形式展现给用户。用户通过浏览门户网站提供的元数据摘要（核心元数据）快速确定自己所需的信息范围，然后要求门户网站在该范围内进一步搜索。

科技资源数据库目录服务最基础、最核心的技术是元数据技术。元数据与一般的数据没有本质的区别，在计算机中元数据的组织和编码根据需求由软件实现，并用可扩展标记语言 XML（Extensible Markup Language）实现基于互联网的数据共享与交换。科技资源数据共享的元数据信息在存入元数据库的同时按照万维网联盟 W3C（World Wide Web Consortium）的 XML/RDF（Resource Description Framework，资源描述框架）标准保存为 XML 格式文件，并将该文件同时保存到数据库中。该实施方法既考虑不同开发用户需求，又同时考虑 XML 在元数据交换和共享中的优势，使系统具有良好的可扩展性。

用户首先通过目录服务查询远程的元数据库，然后利用返回元数据所提供的数据

集获取方法取得相应的数据集。元数据专门用于说明专业数据集的各项特征，是一种伴随性数据。由于专业数据集的结构一般都比较复杂且专业性很强，所以相应的元数据通常只能由数据集生产者亲自建立，这就使元数据在采集和管理工作时需要以分布式方式进行。

目录服务基于开放的网络信息搜索和提取协议，使得元数据的网络发布与元数据的采集和存储系统相互独立，可以比较好地适应元数据在存储和管理方面分布式的特点。目录服务提供 Client/Server 和 Browser/WebServer 两种模式对元数据库进行远程访问。在 Browser/WebServer 下，可以建立元数据的分布管理、集中发布体系，从而为数据集的规范化共享活动提供直接、可靠的技术支撑。

（3）数据服务。数据服务是在目录服务基础上的数据内容服务，提供多种多样的数据类型，能够实现各种结构化、非结构化数据的浏览、查询、下载、可视化和专题服务等功能。数据服务为用户提供一系列工具，以便在众多来源的海量数据中进行数据搜索、多源数据整合，及时发现所需要的知识，提高科学数据的利用率。各数据资源建设组织可以针对自身的优势，构建其有特色的服务体系，如科技论坛、统计分析、专题计算等。

数据服务涉及各类数据的多种服务功能，而且有大量的数据还存储在分布式系统中，需要在充分利用已有信息产品的基础上进行全方位的开发。根据目前的信息发展趋势，可以充分利用分布式异构数据库集成共享中间软件开发多种跨平台的信息发布服务。中间件技术将用户提交的基于公共数据模型的查询分解翻译成一个或者多个对数据源的查询，然后将数据源的查询结果综合处理成公共数据模型的数据，并将结果返回给用户。这种方法向用户屏蔽了底层数据源的差异，使得用户的查询表面上是针对单一数据源的，而实际上查询是由各个数据源的子查询结果综合而成的，因此也称为“虚拟”方法。

数据服务提供针对重大科研项目的专题数据服务，进行集中数据挖掘，提供虚拟现实环境，其内容和形式随着应用的深入而不断丰富和发展。科技资源数据库功能性开发的一项重要任务就是能够便于用户在众多来源的海量数据中，进行各种预测和决策的辅助支持，提高科学数据的利用率。数据挖掘是近年来数据库技术应用的一个新突破，在对数据资源的重组和辅助决策支持方面有着广阔的应用前景。科学数据共享服务系统在进行数据挖掘方面有独特的优势。因为共享服务系统所涉及的数据大都是一些行业性的数据信息，有着较强的规律性，可以分别建立不同的科技资源数据模型，对分布式的数据库与信息系统进行集成挖掘。

对数据挖掘的研究应该是科技资源数据库服务系统的一个深层次的应用，虽然在短期内可能成效不大，但只要坚持发展下去，做到从数据中产生知识，最终会实现在数据高度共享基础上的深入挖掘应用。

（4）数据交换服务。在科技资源数据库信息集成中，数据整合成为系统建设中的一项主要基础性工作。数据整合的内容包括数据同步、消息交换、数据聚合、实时、定时、按需服务等。数据交换服务为有效实现数据整合和交换，完成异构平台、应用间的数据整合和数据共享服务提供支撑。数据交换服务重点解决的问题包括以下几方面：

① 异构性。数据交换中面临的首要问题是应用系统的异构性，主要表现在系统异

构和模式异构两个方面。系统异构是指与数据存储相关的应用系统、数据库管理系统以及操作系统之间的一种或几种不相同而产生的异构；模式异构是指数据在存储模式上不同产生的异构，如相同含义的数据在不同的应用系统中存储类型不是完全一致。

② 完整性。异构数据库间数据交换的目的是为应用系统提供统一的访问支持。为了满足不同应用系统处理数据的条件，经过数据交换服务后的数据应保证原有数据的完整性，包括数据完整性和约束完整性两方面。数据完整性是指完整提取数据本身；约束完整性是指完整提取或转换数据与数据之间的关联关系，保证数据间逻辑特征的完整性。

③ 灵活性。数据交换服务应能为不同的应用提供不同的个性化服务，即用户可以根据自己的需求来选择、编排或定制服务。从而保证随着用户需求发生改变，用户可以快速、便捷地对平台的服务进行修改和扩充。

④ 安全性和权限。由于数据库资源可能归属于不同的部门或单位，所以如何在访问数据库的基础上保障原有数据库用户的权限不被侵犯，如何实现对原有数据库访问权限的隔离和控制，就成为数据交换平台必须解决的关键问题；另外，数据交换的过程中还可能涉及企业需要保密的信息，所以在数据传输的过程中必须建立基于数据加密和身份认证等安全机制。

⑤ 跨越防火墙通信。通常基于安全方面的考虑，人们使用防火墙将应用系统与外界网络隔离开来。因此，为了能使应用系统之间能进行正常的数据交换，数据交换平台必须能够跨越防火墙进行通信。

## 1.3 科技资源数据库建设架构

科技资源数据库建设架构由四层组成，自下而上分别为网络层、数据库层、应用服务层和用户层，如图 1-2 所示。

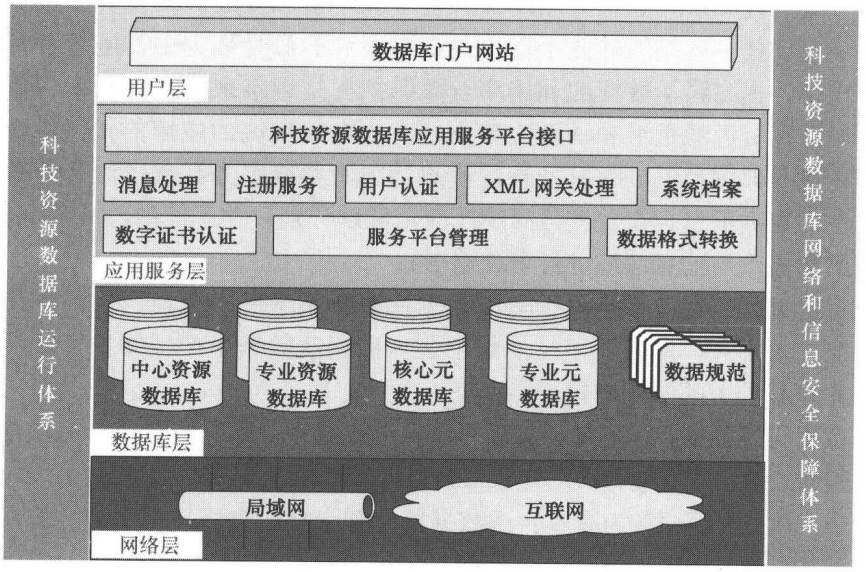



图 1-2 科技资源数据库建设架构