

我国科技信息资源的数字化网络化集成化构建

The Digitized, Web-based and integrated Infrastructure of STI Resource in China

赵新力 罗勇

(中国科学技术信息研究所, 北京 100038)

【摘要】现代信息技术已彻底改变了我国科技信息服务行业的环境,数字化信息资源建设成为我国科技信息行业共建共享的基础。这项工作近年已有相当的发展,初步形成了我国科技信息的共建共享体系。但是,在这一进程中还有许多急待解决的环境建设问题。

【关键词】 数字图书馆 网络 科技信息 数据库 资源共享

Abstract : Information technology has irreversible changed the environment of Chinese scientific and technical information (STI) service obviously. The construction of digitized information resource became the significant foundation of STI joint-building and sharing system. This job has already got considerable development in recent years and formed the China's STI joint-building and sharing system initially. However, there is a lot of environmental improving to be studied in this process.

Key word: Digital library; Internet; Scientific and technical information (STI); Database; Resource sharing;

1. 引言

随着人类社会进入知识经济时代,信息资源已被认为是继能源、物质资源之后最为重要的战略资源,成为国家利益的重要组成部分。世界各国也早已认识到,作为整个社会信息资源一个重要组成部分的科技信息资源的建设对于一个国家的科技进步与创新,乃至整个经济和社会的发展都是一种不可或缺的重要支撑条件。我国的科技信息资源建设从上世纪 50 年代中期开始,经历了波折起伏,总体向上的发展历程。近 10 年来,随着计算机技术的发展,特别是互联网的普及应用,我国科技信息资源的建设进入了一个以数字化、集成化和网络化为特征的蓬勃发展新阶段。本文将对我国科技信息资源的数字化、集成化和网络化构建的历程、成功案例进行介绍和分析,对其发展的趋势和对策进行探讨。

2. 我国科技信息资源体系的建设

2.1 我国科技信息资源及其服务体系的建立

1956年10月，中国科学院科学情报研究所的成立标志着我国科技信息资源和服务体系的建立。自那以后，我国按照专业部门、政府体系逐步建立起了一个从中央到地方、门类齐全的全性科技信息服务体系。到上世纪90年代，我国形成了以中央和地方的综合科技信息、公共图书馆、中科院文献情报、高校图书馆以及独立的专利、标准和国防科技情报等隶属不同行政部门的七大主要科技信息系统。据90年代初的一项统计，当时全国共有各类科技信息机构4000多个，从业人员超过10万人。总体看来，虽然各系统科技信息机构的发展是在国家统一规划和协调下建立的，但在发展过程中却出现了系统间不能有效兼容、缺乏资源共享的机制与条件，经费投入分散、低效，在基础设施和资源建设上低水平重复的现象，无法满足国家建设对科技文献信息资源的需求。

据科技部2002年的一项调查，当时我国每年平均出版科技图书约5万种，同期世界科技图书每年平均出版约70万种，我国引进和收藏约27800种。国内科技期刊出版约4700种，全世界每年出版重要科技期刊约6.6万种，我国引进16433种。全世界科技报告每年出版70余万件，国内只对美国政府出版的四大套报告有系统的收藏。国内学术会议论文每年产生约10万篇以上，但流通性较差；同期全世界每年出版6000多种学术会议论文集，我国国内引进了其中的三分之一左右。国内收藏有国外专利文献4000万件、国内专利150万件，国内外标准、计量文献35万件，占国内外标准计量文献总量（150万件）的23%。

2.2 科技信息资源的数字化构建历程

我国最早在上世纪60年中期就开始探索利用电子计算机进行科技信息的自动化检索，而真正的科技信息数字化资源的构建应该说是从70年代中期起步的。当时，中国科学技术情报研究所（现中国科学技术信息研究所）引进了一台小型电子计算机从事文献的检索研究，主要用于建立馆藏文献的检索数据库。上世纪80年代初期，全国开始了具有一定规模的数据库建设，主要是以科技文献数据库为主。1983年建成的《中国药文学文献数据库》是我国第一个具有实用价值的、在计算机上运行的数字化中文科技信息数据库，而90年代则是我国数据库资源建设和服务业的蓬勃发展时期。到90年代末，全国基于网络平台的科技信息数据库达到7000个左右，但普遍存在着容量少、数据覆盖年限短、更新慢、标引深度浅以及缺少高质量名牌数据库等问题。

2.3 科技信息资源的网络化建设和服务

我国科技信息领域对网络化数据库资源的应用始于上世纪80年代初期。1983年，中国科技情报研究所在国内首个开通了以电话专线为基础的欧美科技信息数据库（包括ESA-IRS、DIALOG、ORBIT等）国际联机检索服务。国内各信

息服务机构也在 80 年代末期基本建成了四大联机检索系统，向国内用户提供自建中文数据库和引进国外科技数据库的联机检索服务。以 1994 年中国科技信息研究所接入国际互联网为标志，我国科技信息和图书馆界开始了以互联网为基础的资源建设和服务时代。经过 90 年代中后期的发展，到二十一世纪初，我国已涌现出几个较大规模的科技信息网络体系，包括以 9 个国家级科技信息文献机构共建的“国家科技图书文献中心（NSTL）”、高教系统的“中国高等教育文献保障体系（CALIS）”以及文化部系统的“中国数字图书馆工程（CDLP）”。中国科技信息研究所更是在 2004 年初步建成了基于下一代互联网关键技术-IPv6 的 NSTL 文献信息服务示范平台，开始了向下一代互联网建设的过渡。

2.4 科技信息资源的集成化（共建共享）

世纪之交，随着互联网应用的普及，我国科技信息资源的大规模共建共享条件日臻成熟。据 CNNIC 的统计，截止 2004 年 7 月，我国接入互联网的计算机数量达到 3630 万台，上网人数达到 8700 万人，分别是 1997 年 10 月首次进行互联网调查时的 121.4 倍和 140.3 倍。而按照 CNNIC 的调查，用户利用互联网获取信息的占调查总数的 98.7%。也正是用户对信息的需求进一步推动了我国科技信息资源的数字化和网络化构建，并掀起了共建共享的浪潮。

我国科技信息领域开始进行大规模的、以网络为基础的信息资源数字化建设实际上是与我国数字图书馆发展同步进行的。我国数字图书馆建设比较有标志性的开端是 1996 年立项的国家重点科技项目“中国试验型数字式图书馆项目”（CPDLP）、科技部 1998 年立项的 863/306 项目“知识网络-数字图书馆系统工程”。其中，CPDLP 项目完成了 900GB 的多媒体资源的建设与发布；863/306 项目完成了总量超过 100GB 的五个较大规模的多媒体资源库建设、实现了跨库联合检索并已实用化。

目前，我国在科技信息、教育和文化系统还分别进行着几项由政府主导的数字图书馆工程，大规模地进行综合性信息资源的数字化、网络化和集成化建设。它们分别是 1998 年 5 月立项的中国高等教育文献保障体系（CALIS）、2000 年 4 月开始建设的中国数字图书馆工程（CDLP）及 2000 年 6 月成立的国家科技图书文献中心（NSTL）。这些项目遵循分布式加工、存取和使用的方式，开展信息资源的数字化和数据库建设，通过系统集成和网络手段对信息资源进行共建共享。另外还有一些商业机构也是在这期间开始提供商业化的数字化、网络化科技信息资源服务。

2.5 我国科技信息资源构建的经验

政府强有力的资金支持，保障了科技信息资源的公益性建设和服务。几项由

政府主导的大规模信息资源数字化——数字图书馆建设工作都有政府的强力资金投入。NSTL 近三年来的投入已接近 4 亿元，CALIS 一期财政投入为 5000 万（地方政府配套及参建学校资金投入也接近 5000 万；同期高校在资源建设上共投入 3.43 个亿）。

以系统集成实现科技信息资源的共建共享。目前较大的资源数字化项目基本上是基于几个传统的图书、科技信息收藏和服务机构以及用户较为集中的高教系统为基础开展的。如：“中国数字图书馆工程”由国家图书馆等文化系统单位发起；NSTL 是由 9 家国家和专业部委科技信息机构共建，而 CALIS 则是以我国教育系统的 100 多家高校图书馆为基础建设的。

以市场机制促进社会化的科技信息内容产业逐步发展。目前，我国民间商业机构主要通过各种期刊数据库来介入科技信息资源的构建，初步形成了我国数字化期刊信息资源群。目前影响比较大的有维普、万方、同方等公司建设的数据库产品。其中，仅维普就宣称能提供 8000 种中文期刊、1000 种中文报纸、4000 种外文期刊的数字化内容；同方中国期刊全文数据库收录有自 1994 年以来国内公开出版的 6100 种核心期刊与专业特色期刊的全文，积累全文文献 800 万篇，题录 1500 余万条，几乎囊括了国内主要的学术性期刊信息资源。目前，国内许多信息服务机构、大专院校或公共图书馆都在利用这些数据库为公众提供服务。

3. 数字化、网络化和集成化的信息资源构建案例

通过国家的集中投入，近年来我国陆续在科技信息资源建设领域开展了几项比较有规模的项目，大大缓解了以前广泛存在的科技信息资源建设投入匮乏的问题，标志着我国科技信息资源的数字化、网络化和集成化迈出了重要的第一步。

3.1 国家科技图书文献中心（NSTL）

NSTL 是 2000 年由中国科学技术信息研究所等国内 9 家科技信息服务机构组建的一个虚拟的科技文献信息服务机构。其宗旨是根据国家科技发展需要，按照“统一采购、规范加工、联合上网、资源共享”的原则，采集、收藏和开发理、工、农、医各学科领域的科技文献资源，面向全国开展科技文献信息服务。

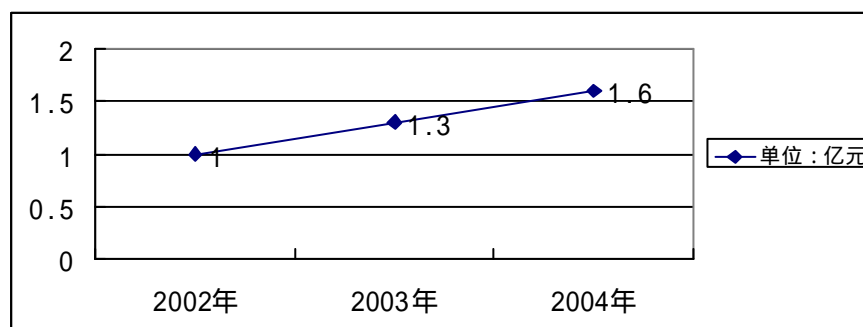


图 1. NSTL 投入增长情况

到 2004 年，NSTL 采集的外文期刊种类达 15000 种，约占国内引进品种的 2/3；外文会议录、工具书等 5000 多种，约占国内引进品种的 60%；另外还收藏有外文学位论文 47000 篇、中文科技期刊 4000 多种、中文会议录 22000 多种、学位论文 470000 多篇。

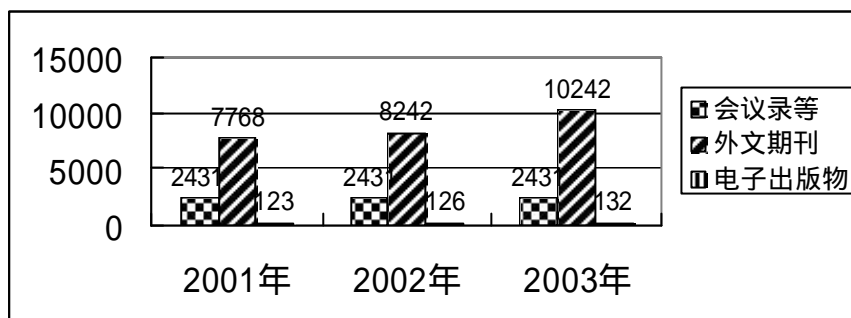


图 2. NSTL 主要信息资源采集情况

为了适应科技文献服务的资源数字化、服务网络化、构建集成化的趋势，NSTL 也加大了对已有馆藏文献的数字化加工。目前数字化二次文献的数量已接近 100GB，上网文献量超过 2700 万条；仅中国科技信息研究所一家所拥有的数字文献资源就达到了 3.4TB（尚未包括网络版期刊、网络版数据库的数据量）。

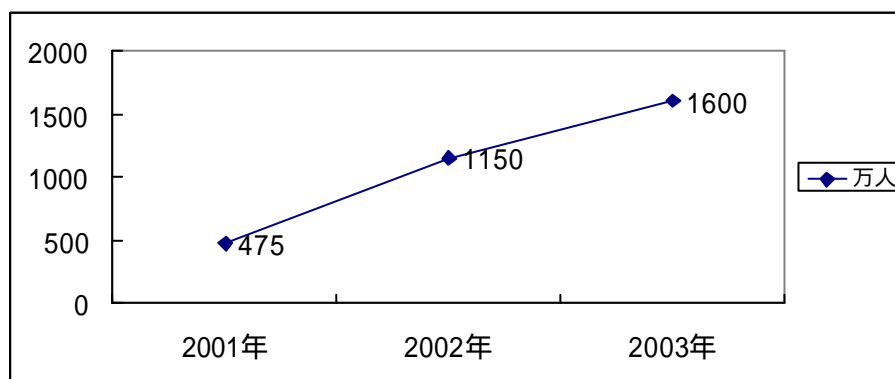


图 3. NSTL 网络检索服务增长情况

这些资源极大地促进了我国科技信息服务的发展。截止到 2003 年底，NSTL 网络系统共向用户提供检索访问 3225 万人次，网上报道的外文期刊 99.58% 被检索利用，向用户提供全文服务 26 万多篇。

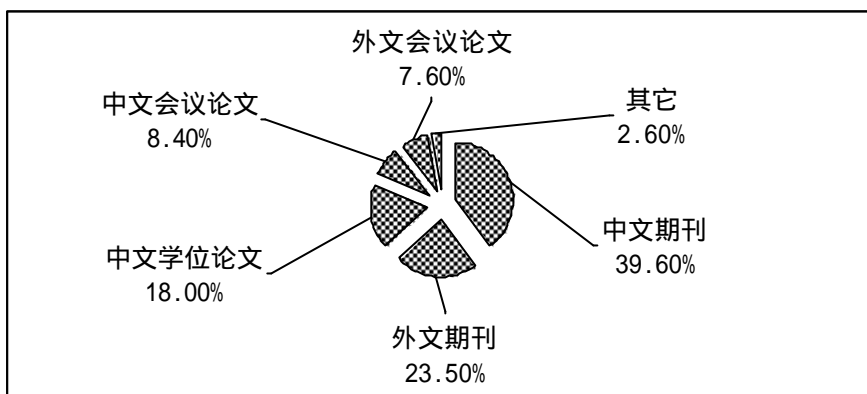


图 4. NSTL 用户检索统计

如今，NSTL 正在以国家中长期发展为导向，以国家科技基础条件平台建设为契机，以“合作、开放、共享”的机制共同构建国家科技文献信息资源战略保障和网络服务体系，推进国家科技文献资源与服务平台的建设。

3.2 中国高等教育文献保障体系（CALIS）

CALIS 是我国高等教育“211 工程”总体规划中建设的两个公共服务体系之一，1996 年立项、1998 年正式启动。其总体目标是在“九五”期间，建成以中国教育和科研计算机网（CERNET）为依托的网上信息资源共享系统，为我国高校的教学、科研人员营造一个与国际接轨的信息网络环境，为高校提供丰富的文献信息资源、先进的技术手段和便利的服务体系。

该项目由国内 152 所高校图书馆组成联盟，“九五”期间中央财政投入 5000 万，加上地方和学校的配套资金，总投资达到 9300 多万；初步搭建起了中国高等教育数字图书馆(CADLIS)的基础框架。

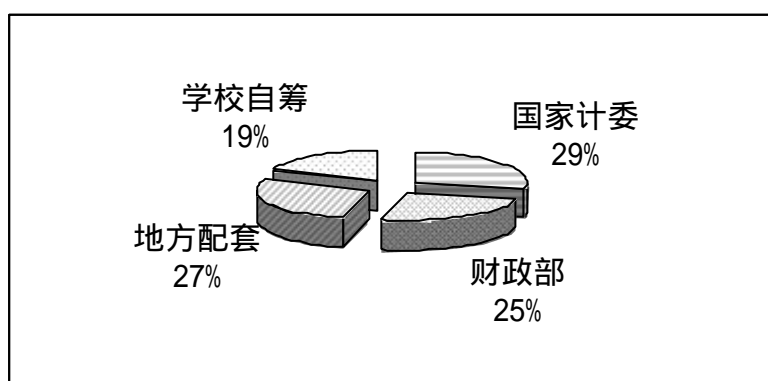


图 5. CALIS “九五” 期间投入情况

CALIS 的服务功能主要有馆际互借、联机合作编目、公共检索、文献传递、电子资源导航、文献采购协调等。同时，它也面向社会提供文献信息服务。其资

源的特点是文理兼备，这也是高等教育的多学科特点所决定的。CALIS 的数字资源建设主要包括两个方面：

引进数据库：国外数据库的引进工作主要在文理和工程两个全国中心进行。数据库引进采取了集团购买的方式，先后引进多种世界著名数据库，提供全文刊 7800 多种，这些数据库大都通过校园网为各高校提供文献服务。

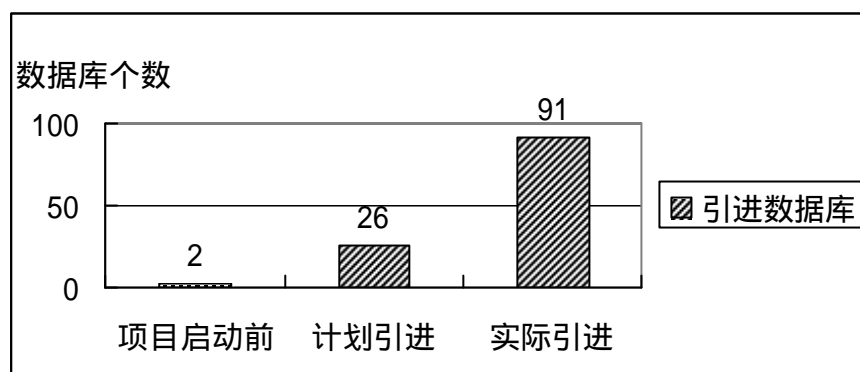


图 6. CALIS “九五”期间引进数据库情况

自建数据库：主要包括中文馆藏联合目录数据库、中文现刊目录库、中文特色数据库、学位论文和会议论文数据库、211 重点学科导航库等。

“九五”期间，CALIS 建设取得了重要进展，增强了国内高校重要外文学术资源收藏与获得水平并促进了高校间文献资源的共建共享能力。通过组织高校图书馆集团引进国外重要学术文献数据库，在以工程中心、文理中心为首的各中心点上建立了较为全面的学术文献资源收藏，重点学科的重要学术资源的国内收藏率达到 90% 左右；使文献保障率从不到 50% 提高到 80%，骨干高校图书馆的设施和服务水平已经接近国际先进水平。

“十五”期间，CALIS 将作为技术的组织者和标准的协调者，组织全国高校图书馆，完善和利用“九五”期间已建立的文献信息保障体系，加大数字资源建设力度，建设一批面向高等教育的特色数字资源，建成若干具有先进应用技术水平数字图书馆基地，于 2005 年初步建成中国高等教育数字图书馆(CADLIS)。

3.3 中国数字图书馆工程

中国数字图书馆工程项目于 2000 年启动，是由国家图书馆等文化系统单位发起开展的。该工程旨在建设超大规模的优质中文信息资源库群，并通过国家高速宽带网向全国及全球提供服务，形成世界上最全面成系统的网上中文信息基地和服务中心。

中国数字图书馆工程的建设坚持“统筹规划、需求牵引、科技创新、滚动发

展”为指导思想，坚持公益性为主、资源建设为核心、统一标准规范、开放建设与利益共享、开发与引进相结合等原则进行建设。其总体建设目标是：通过资源建设工作的组织与实施，建成超大规模的、高质量的分布式中文数字资源库群并提供网上等多种服务；联合引进若干国外专题资源库并实现共享；实现全国大部分地区图书馆文献资源的联合目录系统。以国家数字图书馆国家中心为基础，以行业、地区分中心为辐射，逐步建设具有模块化、开放性、互相联通并且稳定可靠、可扩展的计算机网络与存储体系。集成具有自主知识产权的高新技术成果，努力形成总体技术与国际主流技术接轨的中国数字图书馆总体架构。其建设内容包括：数字资源建设、软/硬件基础设施建设、应用系统开发、标准规范制定、信息服务、人才建设及知识产权的解决。

按照《中国数字图书馆工程建设一期规划》(2000-2005)，这一项目主要以文化信息为主，兼顾包括科技信息在内的其他领域。它以图书文献机构的丰富馆藏为依托，并结合其它文化机构的资源进行数字资源的建设，内容涉及文学、艺术、法律、科技、教育、旅游等各类信息，并依托国家图书馆馆藏进行古籍的数字化加工。文化部还正在此基础上建设全国文化信息资源共享工程。目前，中国数字图书馆工程项目共累积建设了总量近 11TB 的数字资源。仅以国家图书馆为例，该馆已完成 6380 万页图书、近 2000 部影片、22 万首音乐作品、4000 页馆藏西夏文资料、8000 幅金石拓片、180 万拍民国期刊、近 8 万篇博士论文等多种类型资源的数字化工作。预计到 2005 年规划完成时，包括各种数据库在内的数字资源将达到 20TB 以上，并在网上为读者提供便捷服务。

4. 我国科技信息资源构建的发展趋势和对策

随着现代科学技术的发展，尤其是自上世纪八、九十年代以来，计算机和网络技术与信息服务行业的日益融合，计算机和网络已经不可逆转地彻底改变了传统图书馆、情报工作的环境、内容和方法，同时也深刻地改变着人们获取信息的取向和习惯。随着信息技术特别是互联网技术、数据存储与检索、数据挖掘技术的迅速发展，信息资源数字化、信息传输网络化、信息管理自动化、信息分析智能化、信息服务专门化成为科技信息服务行业发展的主要趋势。在数字化、网络化和集成化的资源构建过程中，应该充分重视以下几个方面的问题，才能够真正做到社会化的信息资源共建共享和科技信息服务系统的共赢。

4.1 科技信息资源共建共享的支撑是数字化

我国以网络为平台的科技信息服务环境正在形成，基于网络环境、以分布分层为特征进行资源组织的信息服务体系正在出现，具有互通、互联和互操作能力的网络信息服务已成为潮流。这一切的基础是信息资源的数字化，而且是以网络

为基础、具有分布式特征的数字化资源集群，而这其中的难点是多媒体信息资源的数字化和有序化。因此，我国科技信息资源的共建的重点应放在基础资源的数字化建设方面，这样才有可能通过网络信息联合作业平台、信息指引服务平台和个性化信息服务平台的相互结合来实现信息的共享。

4.2 科技信息资源网络化服务的关键是建立共享机制

要建设起以网络为基础的分布式数据库系统，重视跨系统、跨体制的资源集成。拟建和在建的国家科技数字图书馆、中国高等教育数字图书馆和中国数字图书馆在进行信息资源数字化建设和资源整合的时，应该强调资源的个性、突出特点，要顺应社会数字化内容产业兴起的趋势，找到现代投资体制下合理的资源构建方式。商业投资能解决的资源数字化种类，公共机构就可以通过采购等方式支持而不一定要自己建设，坚持按照市场机制进行资源的配置；公共机构应将信息资源建设的重点放在商业机构没有能力建设的品种上，放在社会资源的整合和搭建基于 web 环境的服务体系上，放在资源共享机制的建立上，培育社会整体的信息资源共享环境。

4.3 科技信息服务共赢的前提是集成化

要加强服务层面的集成、创新。资源整合不仅包括资源建设的整合，也应该在服务层面进行整合；数字资源建设的目的并不是为了扩充馆藏，而更应该考虑如何能快捷方便地为用户所利用，使参与各方都能够从中获益。目前几个大系统的建设一般都是自建一个平台，相互之间数据格式不统一、检索方式各异、服务政策不同。比较典型的是维普、万方、同方三家信息提供商都要求用户安装各自不同的浏览软件，还要多头开户、分别付费才能下载和使用其数字资源。这相当程度上给用户带来了困扰。因此，在数字图书馆建设中加强跨平台跨库的服务整合、形成集成化的科技信息揭示和服务体系、创新服务方式等都应该引起足够重视。

4.4 科技信息资源构建的瓶颈是法规和标准规范

在科技信息资源的共建共享过程中，尚有许多需要解决的技术保障和法律保障问题，包括数字资源建设的规范标准。统一的标准体系是科技信息资源共建共享的重要基础条件。同时，对网络条件下的知识产权的认定和保护、信息公开与保密的相关法律法规、行业认证等问题，都需要加大政策研究的力度，制定具体可行的办法和措施，为信息资源的共享完善社会环境。

5. 结语

随着科技信息活动的重点从文献资源的获取转向为知识的发现和应用，个性

化的信息需求越发明显,科技信息服务将逐渐走向知识服务。在数字化信息资源的支撑下,科技信息服务机构已不再囿于传统的文献收集、加工和服务领域,内容深度开发和知识开发将成为信息服务的重点,数字化的科技信息资源建设已成为我国科技信息实现共建共享的必由之路。我国科技信息行业也将遵循“7R”的理念——专业的人员(Regular Person)把正确的内容(Right Content)以用户要求的方式(Requested Way)、合适的成本(Reasonable Cost)、在确定的时间和地点(Right Time and Right Place)提供给需要的用户(Right User)。可以相信,在信息内容产业蓬勃发展的大前提下,我国科技信息行业一定能够通过构建数字化、网络化、集成化的资源体系,共建共享的良好氛围,以不断满足社会日益增长的科技信息需求,为国家创新体系的形成和国家信息化建设做出自己的贡献。

参考文献

- [1] 李立群. 信息产业的界定与科技信息机构的发展对策[J]. 科学与管理, 1994, 14(3): 49-50
- [2] 中国科学技术信息研究所等. 我国科技电子信息资源的开发和利用研究[R]. 北京: 北京图书馆出版社, 1999年5月
- [3] 王义道. Calis项目“九五”期间建设情况汇报[R].
(<http://www.calis.edu.cn/rm/calis211.pdf> 2004.11.15.)
- [4] 肖珑. CALIS“十五”期间数字资源的整体化建设思路[R].
(<http://www.calis.edu.cn/shul/calis2.pdf> 2004.11.15.)
- [5] 杨毅,“CALIS‘集团采购’引进电子资源回顾”[R](CALIS数字化服务工作研讨会暨第二届国外引进数据库培训周·大连·2004.5.)
- [6] 袁海波,“基于网络环境的文献资源共建共享的实践与启示”[R](中欧数字资源长期保存国际研讨会·北京·2004.6.)
- [7] 周和平,“加强文化信息资源建设,推进经济与社会全面、协调、可持续发展”(信息资源开发利用高层论坛·北京·2004.9.)