

# 科学数据共享工程的总体框架

## 科学数据共享调研组

科学数据是指人类社会科技活动所产生的基本数据,以及按照不同需求而系统加工的数据产品和相关信息。科学数据是信息时代最基本、最活跃、影响面最宽的科技资源,具有明显的潜在价值和可开发价值,并在应用过程中得以增值。实施科学数据共享工程是在国家统一规划、政策调控和相应法规的保障下,应用现代信息技术,整合离散的科学数据资源,构建面向全社会的共享服务体系,实现对科学数据资源的规范化管理及其高效利用,从而为科技创新、政府决策、经济增长、社会发展和国家安全提供科学数据信息资源的保障。它是国家信息化建设的重要组成部分,是国家创新体系中急需发展的现代科学基础设施,是对科技创新能力的建设。

### 一、科学数据管理的走向——战略资源地位及其共享管理的必然选择

#### 1. 科学数据共享管理的逻辑起点是资源

回顾人类对科技活动本质认识的升华和理性日益成熟的 20 世纪,在科学、技术和工程领域的纵深发展,无一例外地依赖于科技工作者的创新活动和现代科学基础设施与科学数据资源的保障。如果我们把创新体系比作知识经济的动力系统,那么科学数据与相关信息则是这个动力系统的燃料资源。因为科学数据能从根本上满足科技创新、经济增长、社会发展和国家安全等多种需求,对建立现代精神文明和物质文明有着重要作用,从而构成信息时代的一种战略资源。

源。它凝聚着人类科技活动、社会活动和经济活动的价值,而其价值又是通过效用来满足人类生存与发展的需求,其面向全社会共享的结果是效益递增。

如前所述,科学数据之所以具有满足国家发展广泛需求的资源属性,在于科学数据具有明显的潜在价值和可开发价值,其价值的实现在于广泛应用。它还有别于一般意义上的资源,即:科学数据资源是人类科技活动的产物及其长期积累,是一种科技资源,它可以按照社会的多种需求提供系统的足够的数据量;其来源可靠、质量保证,又可以不断得到补充或更新;所提供的数量、质量、产品形态及其存储与传输方式可以通过人为控制;其价值的实现与开发者的能力和方法密切相关,而且能在应用过程中增殖与增值。显然,科学数据价值的突显与其可共享性有关:一方面是由于诸多科学技术的突破诞生在学科交叉的前沿领域,它的研究需要多学科领域的知识、信息和数据的支持,一个领域的科学数据已成为不同学科领域的研究者所共同的需求;另一方面是因为数据可以无限制复制,不会因为满足某时某人需求而影响任何时候他人对其需求。这种共享特性的充分发挥,使其效用价值变得更大。由此可见,科学数据共享管理的逻辑起点是资源。

#### 2. 科学数据共享管理是时代发展的必然

科学技术的迅猛发展及其对社会与经济巨大推动作用已成为当今社会的主要时代特征之一。随着全球性经济合作与竞争以及科技日新月异的发展,全球范围的合作与竞争将进入更高层次,竞争的焦点不仅表现在自然资源与市场的争夺,而且对人才与科技的争夺突显,原始性创新成为科技竞争的制高

\* 本文是在科学数据共享调研组提交的“实施科学数据共享工程,增强国家科技创新能力”研究报告的基础上,由黄鼎成根据近期研究进展所撰写。

点。在这一强烈的时代背景下,加速国家科技创新体系建设,加强包括科技人才、现代科学基础设施和科技信息数据资源等在内的创新能力建设,已经提到各国政府的议事日程。基于全球范围内对科技信息数据资源交流的客观需求和现代信息技术的发展,为信息和数据的获取、发布与使用提供了高效率的传播与交流的公共平台。使科学数据资源管理由传统的存档职能,向提供服务与数据产品开发转变;数据应用从单一目标,向面向全社会多目标服务转变;在信息技术驱动下,数据的传播与交流向标准化、数字化、网络化方向转变。互联网及其丰富多样的知识、信息、数据资源正成为科技创新的公共平台。

据估算,人类社会最近 30 年所积累的科学数据总量已经超过了 5000 年发展历史所积累的数据量总和。在科学数据的采集、存储、处理和传输的数量与日俱增的同时,也出现了人们始料未及的新问题:一是科学数据数量猛增,但对其利用不足;二是数据资源共享与数据占有相矛盾;三是数据资源共享亟待规范等等。如何使海量科学数据资源在全社会流动起来,规范数据的管理,最大限度地发挥科学数据作为战略资源的作用,已成为新世纪全球科技发展面临的新挑战。基于科学数据明显的资源属性和在日益激烈的国际竞争中的主动性,使许多发达国家和国际科学组织,更加重视科学数据的采集、管理、分发与服务工作,力图通过科学数据的流通、应用与共享取得巨大的社会、经济效益。

半个多世纪来,我国投入巨资开展了一系列大规模的调查、观测、监测、试验和实验等国家科技事业性工作;通过一系列科技计划支持各专业领域的科学研究,先后建设了一批大型科学工程、重点实验室、工程中心及野外观测研究站网等等。这些科技活动积累了海量科学数据。在科学数据采集和积累方面,初步形成了以部门为主体与科研院所、高等院校互为补充的格局;专门的数据管理机构(组织)相继建成;国际间的数据合作与交换渠道逐步建立。从而支持了大量的科技创新活动,在国家宏观决策、社会发展、经济发展和国家安全诸方面发挥了重要作用。但是,科学数据资源管理滞后,共享问题仍十分突出,已在相当程度上制约了科技进步与创新,乃至国家整体实力的

提高,并造成了重复采集、重复建设和资金的浪费。分析认为,共享问题“久战不下”的原因在于:对科学数据、资料和信息占有的思想严重存在,共建共享观念淡薄;国家层次缺乏宏观管理与协调,政府对科学数据共享的投入不足;没有形成完善的数据交换标准体系和强有力的共享服务技术支撑以及政策、法规体系的保障。

综上所述,科学数据共享管理是时代发展的必然,其实质就是实现科学数据资源的开放与共用,在广泛的应用中最大限度地发挥科学数据资源的潜在价值,使国家的投入获得最大的效益,提高科技创新能力,乃至国家整体科技水平。无疑,实施科学数据共享工程已成为我国信息化建设的重要任务,首先要做好由国家投资获取与积累的公益性、基础性科学数据的共享管理,创建共享新秩序,建立面向全社会的共享服务体系,提高科技创新能力与国际竞争力。

## 二、科学数据共享工程

### 1. 科学数据共享工程的建设思路与总体目标

科学数据共享工程是通过国家的整体规划、政策调控和法规体系保障,应用现代信息技术,对公益性、基础性的科学数据资源进行规范化管理,建立面向全社会的共享服务体系,实现科学数据资源的开放与共用,是增强科技创新能力的一个复杂的系统工程。工程建设的总体思路是:在国家统筹规划下,应用现代信息技术,整合集成各部门、各单位的科学数据资源,充分利用国际科学数据资源,并通过制定共享政策、法规和完善管理体制,把各部门、各单位乃至个人所获取与积累的科学数据资源,纳入国家科学数据共享管理的统一框架;通过国家科学数据中心群和共享服务网的建设以及共享技术的研究开发与应用,形成跨部门、跨地区、跨学科、多层次、分布式的国家科学数据共享服务体系。大幅度提高科学数据的管理与共享服务水平,增强科技创新能力,为国家整体发展和科技整体水平的提高,提供可靠的科学数据资源保障。

科学数据共享工程建设的总体目标是:利用 10 年的时间,完成共享政策、法规体系和管理体制的建设,建立健全共享机制;构建 50 个左右的国家科学数

据中心或科学数据共享服务网,基本形成覆盖全面,结构合理,功能先进的科学数据共享服务体系,基本满足科技创新和国家发展对科学数据的需求。

### 2. 科学数据共享工程总体框架

国家科学数据共享工程主要由三大部分构成:数据共享政策、法规与标准体系建设;国家科学数据中心群建设;科学数据共享服务网建设(图1)。

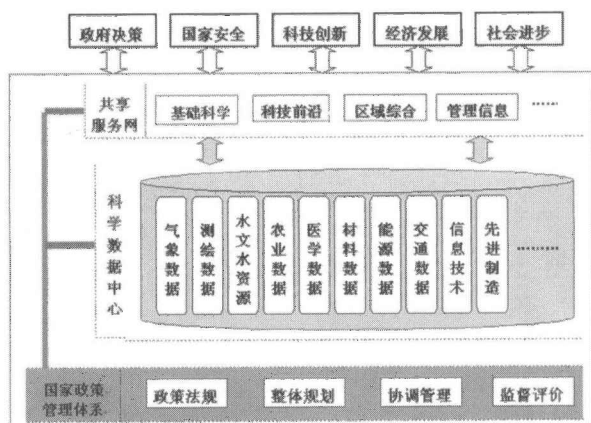


图1 国家科学数据共享工程总体框架

数据共享政策、法规与标准体系建设,是为科学数据共享工作提供政策调控、法规保障和技术前提,并从国家层面上统一规划科学数据共享工程技术框架。科学数据中心群和科学数据共享服务网的建设,是为科学数据共享提供资源保障和技术支撑的共享平台。并在国家科学数据中心和科学数据共享服务网的基础上,构建与上述系统相链接的元数据目录总检索系统(或称科学数据共享工程门户网站)。显然,以数据库为基础的国家科学数据中心和科学数据共享服务网的建设,是在现有网络、数据库、信息发布、身份认证等信息技术的支持下,对分布式数据库(具有物理上的分布性和逻辑上的统一性)和数据集的统一管理,实现目录服务、数据服务、延伸服务等共享服务功能(图2)。

### 3. 科学数据共享工程的主要任务

(1)共享技术平台的顶层设计和政策、法规研究,制定相关条例与标准体系。科学数据共享最大的难点之一,在于缺乏有效的政策调控、法规保障和规范、标准。只有深入开展相关研究,才有可能创建共享新秩序,增强共建共享氛围。为了保证共享系统主流程的科学化、规范化,开展共享技术平台的顶层设计和

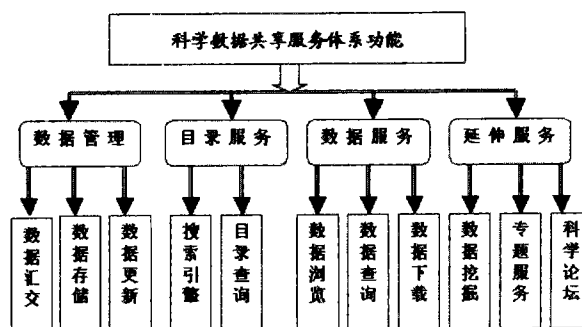


图2 科学数据共享工程服务功能图

政策、法规体系研究,制定相关条例和标准体系的工作是至关重要的。包括:① 科学数据共享的政策、法规体系的研究,立法之前要在国家层次上制定科学数据共享管理条例和相关的规章制度;② 科学数据共享平台的技术体系;③ 科学数据管理与分发服务的标准体系;④ 科学数据分类分级共享及其发布策略等。

(2)构建国家科学数据中心。科学数据中心群由不同领域的国家科学数据中心组成。科学数据中心是由技术平台、数据层、服务层、技术支持层和管理层组成的多层次综合系统,是以技术平台和数据层为基础,以技术支持层和管理层为保障,通过服务层向用户提供科学数据查询、检索和分发服务(图3)。

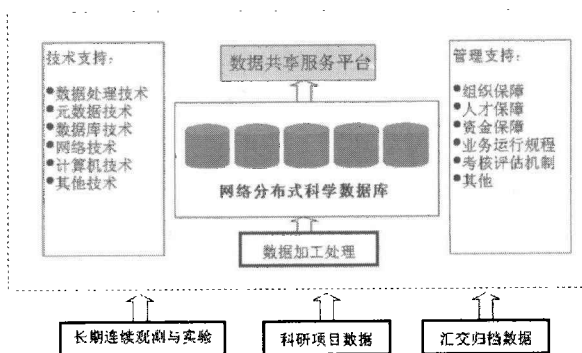


图3 国家科学数据中心结构示意图

国家科学数据中心主要针对国家长期布局的数据采集系统,选择其汇交与管理的数据机构,按科学技术领域分别构建。重点支持:根据科技创新与国家发展的需求,整合已有数据资源以及抢救珍贵历史资料,组织整理加工建库;制定本领域科学数据共享标准与规范;建立数据共享技术平台,面向政府、科技教育界和社会提供数据检索与分发服务。

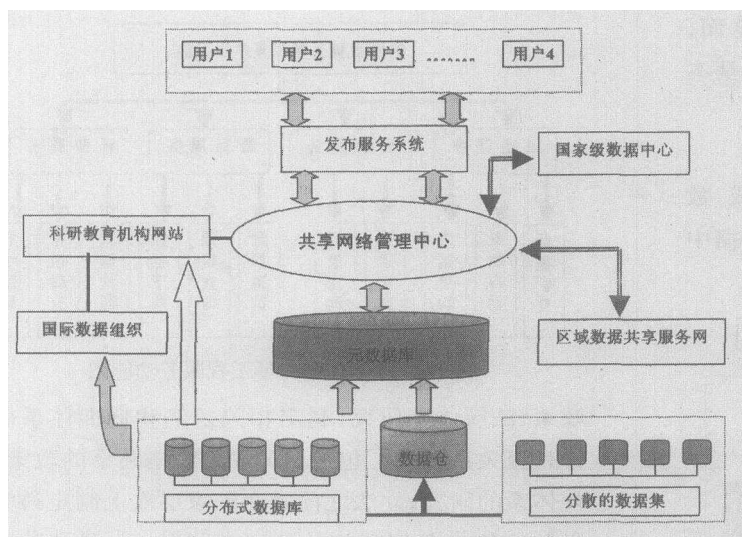


图4 国家科学数据共享服务网结构图

(3)构建科学数据共享服务网。科学数据共享服务网建设主要针对基础科学与科学技术前沿领域以及重点区域,利用互联网构建物理上分布的、逻辑上统一的科学数据共享服务网。即:通过网络连接各科研院所、高等院校和国际数据组织的相关专业数据库而构成,具有网络化、虚拟化、集成化的特点(图4)。科学数据共享服务网包括3种类型:①基础科学与科学技术前沿研究领域的科学数据共享服务网;②重点区域综合性科学数据共享服务网;③科技管理信息共享服务网。

重点支持:集成离散在各研究机构、高等院校乃至科学家个人手中的数据资源上网服务;国际科学数据资源的交换、收集与开发;针对科技前沿研究组织有关数据管理机构和专家共同加工数据产品;组织或参与相关领域数据共享标准与规范的制定;开发与推广适用于本领域的网上数据服务系统;促进各有关单位的数据共享服务网站建设。

### 三、科学数据共享工程的实施

#### 1. 科学数据共享工程的实施要点

如前所述,科学数据共享管理的核心在于实现科学数据资源的开放与共用。为此,我们必须从科学数据对人类生存与发展的意义出发,摸清影响科学数据

共享管理系统的主控因素及其相互作用过程,才能掌握这一复杂系统运行的原理,寻求正确的方法和相应技术的支持。分析认为,科学数据共享管理这一系统的稳定性,取决于彼此关联与相互作用的作用网络的平衡与协调,如图5所示。通过对国家“发展需求”和相关信息的分析,所产生的设想(方案)最具活力,它促进了“组织管理”者以发展需求为导向,制定科学数据共享管理的整体规划与组织实施;体现国家意志的整体规划推动了离散的“数据资源”的整合;“数据资源”的存量、分布状况、产品形态及其价值特性和可共享性决定了共享政策和共享方式的多样性,必然导致维持共享秩序的“共享规则”的产生;“共享规则”(即办法,包括数据的分类分级共享及其

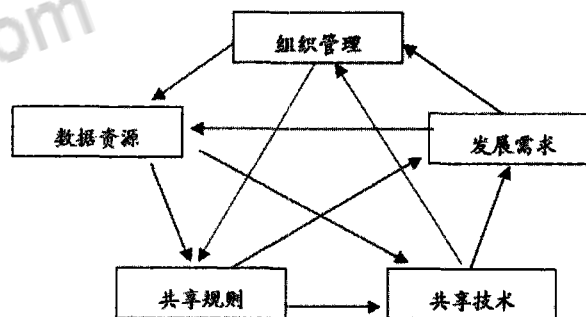


图5 科学数据共享管理系统主控因素相互作用图

发布方式和政策、法规体系)确定后,就需要寻求适合的技术手段,便引导了“共享技术”的选择;在适应社会信息化的“共享技术”的支持下,才能分别满足不同的需求。这是5个要素相互促进的一面,它们之间又存在相互制约的一面,即:“发展需求”对“数据资源”的要求;“共享技术”实现的可能性与创新又对“组织管理”产生影响,包括对整体规划的前瞻性与可行性的检验,以及对高素质数据管理人才的要求;“共享规则”对不同需求用户的共享行为产生约束;“数据资源”的结构、产权分布状况及其更新与维护将对“共享技术”不断提出要求;体现了国家意志的整体规划,又直接影响“共享规则”的形成。上述5个要素在科学数据共享管理系统的运行中起控制作用,按一定秩序相互促进,相互制约,以维持共享管理系统的动态平

衡与协调而健康发展。可以认为,国家整体规划与管理的积极主导作用,既是实现共享规则和共享技术的保证;又是科学数据资源整合与共享的权力体现,它可以通过国家层次的协调和投入而拥有整合资源和创建共享新秩序的能力。而整合数据资源、规范共享行为、强化技术支持,则成为实现科学数据共享的关键环节。

(1)整合存量,规划增量,搞好资源建设。实施科学数据共享工程应盘活已有的数据资源,并规划国家科技投入新产生的数据资源,包括:①应用统一标准与技术实现对离散的国内外科学数据资源的整合,在激活已有资源的基础上,针对国家发展和重大科学问题开展数据产品加工,提升资源的价值;②对国家科技计划项目所产生数据实施高效管理;③对具有高效用价值的珍贵历史资料进行整理与数字化。

(2)政策调控,法规保障,创建共享新秩序。科学数据共享政策、法规体系是建立健全共享机制,提高全社会共建共享氛围,创建共享新秩序的根本保障。国家对科学数据共享工作的原则性规定,既体现在国家层次的宏观管理与协调上,又体现在法规体系建设的指导思想和所必须遵循的原则之中。共享法规体系是以科学数据共享以及由此产生的社会关系为主要调整对象,并对共享行为进行全面规范,重点要解决共享的原则、共享的措施、共享的权利与义务以及共享的评价、监督和法律责任等问题。制定政策、法规的基本要素应包括:①科学数据共享管理旨在推进科技进步与创新;②在我国目前情况下科学数据资源的获取与积累,国家投入是主体;③科学数据资源只有在广泛应用中才能实现其潜在价值,开放与共用是共享管理的核心;④科学数据分类分级共享及其相应的发布策略。同时还必须遵循:无歧视(平等)原则、国家安全和知识产权保护原则、引入竞争机制和反信息垄断原则以及发挥国家宏观调控作用原则。

(3)统一标准,统一技术,实现有效共享。没有适应的社会信息化的技术支持,就很难实现科学数据资源的有效共享,它是健全共享机制,创建共享新秩序的技术保证。应该说标准的统一是数据资源共享的前提,首先要从科学数据共享的标准框架、共享标准体系的指导标准和通用标准入手,继而制定逐个领域

的元数据标准、数据分类和编码标准。通过统一管理、统一技术和统一用户界面,建立在互联网上的共享技术平台是最基本的工作,是实现对多种类型数据的整合,并为不同层次用户提供服务的综合性服务平台。也就是说,应用现有开发软件和理论基础,通过技术集成和进一步开发,抓好科学数据共享技术平台和标准体系的研究与建立,是实施科学数据共享工程最基本的技术保证。

## 2. 试点先行,分步实施

科学数据共享工程是一项长期、复杂的系统工程,要在国家统一规划的框架下,充分发挥相关部门、机构(或组织)的积极作用。其实施原则是:统筹规划,统一标准,联合建设,分步实施,试点先行,集中力量,重点突破,边建设边服务。考虑当前我国的实际,采用分步实施,即:第一期(2005年前)为规划与试点阶段;第二期(2006年—2010年)为全面推进阶段。

(1)以基地的思路,开展中心、服务网的建设与共享服务试点。无论是科学数据中心,还是科学数据共享服务网的建设与共享服务试点,都存在对试点机构(或网管机构)的遴选问题。遴选原则是:①主管部门的推荐与支持,并在组织协调和人力、财力、物力配置等方面予以保障;②具有稳定的数据来源和足够的数据存量及其规模化的主体数据库,并在数据汇集、归档和服务诸方面具有与开展科学数据共享管理相适应的能力;③愿意并承诺对公益性、基础性科学数据的无偿(和非赢利性有偿)共享和数据维护提供长期服务;④具有服务意识的高水平数据管理队伍,具备对相关领域的数据处理、分析、产品加工和管理的技术能力。试点的主要任务是:

——摸清本领域和相关领域科学数据资源的分布状况,形成数据生产的稳定供给与更新、补充机制;

——深入开展科学数据需求分析,以国家发展需求为导向,按照相关标准对数据资源进行整理、分析、加工和质量控制,构建或完善与维护主体数据库;

——对数据资源进行分类分级共享,建立数据共享技术平台,并通过网站、光盘等多种方式向用户提供数据检索与分发服务;

——按照统一的标准化框架,组织制定本领域科学数据共享标准体系(包括元数据标准、共享服务标

准、数据产品与生产标准、数据质量的评价方法与检测规范、应用服务技术规范和管理规范),研究开发共享技术。

(2)以项目形式开展共享技术平台的顶层设计与共享规则的研究制定。以项目形式组织专家结合共享服务试点,进行科学数据共享技术平台的顶层设计和政策、法规体系研究,制定相关条例、规章制度和标准体系,是建立健全共享机制,创建共享新秩序的根本保证。主要包括:

——科学数据共享技术平台的顶层设计(总体技术框架、共享服务平台的解决方案);

——有关元数据技术、异构数据库技术、网络化数据服务技术;

——科学数据共享标准化,特别是标准化框架和共享标准体系的指导标准、跨领域的通用标准以及各领域科学数据共享标准体系的研究与制定;

——科学数据共享的分类分级与发布策略;

——科学数据共享服务绩效评估体系;

——科学数据共享政策、法规体系的研究,包括:元政策、法规体系框架的研究。目前情况下急需研究制定“科学数据共享管理条例”,上报国务院审批后执行,以保证立法之前共享行为有章可循;研究制定“国家科技计划项目科学数据汇交管理办法”和“科学数据共享工程管理办法”等。

#### 四、实施科学数据共享工程的保障措施

认真分析我国科学数据共享问题“久战不下”的原因,不难看出实施“科学数据共享工程”的艰巨性。人们长期对数据、资料和相关信息的占有观念能否彻底打破?长期存在的部门、单位间资料壁垒能否有效消除?共享法规体系的建设能否顺利推进与有效实施?技术和资金能否及时落实?等等,都需要我们认真对待,逐步解决。当前情况下,建议采取以下措施:

##### 1. 强化数据共享工作的政府职能

在国家层次上,加强政府对科学数据共享工作的统一规划与管理,强化政府在数据共享工程中的政策引导与协调作用;并将数据共享列为相关政府部门的工作职责。

##### 2. 健全科学数据共享工程管理体系

建立科学数据共享工程部门协调联席会议制度,形成工程建设实施与运行的最高决策层;成立专家咨询委员会提供决策咨询;设立两会办事机构,对国家科学数据共享工程建设与共享服务实施协调管理。

##### 3. 加快规章制度的建设与立法进程

在开展科学数据共享政策、法规和标准体系研究的同时,完善相应的规章制度,并着手起草适用的科学数据共享管理条例。制定与实施国家科技计划项目科学数据汇交与管理办法,从计划制定和项目立项、中期评估与验收各个环节把握数据、资料的质量与有效管理。

##### 4. 建立科学数据共享工程管理规范

包括工程建设与运行的管理办法、数据管理规范、共享服务监督评价办法等。

##### 5. 保证建设与运行的经费

以基地的思路实施对科学数据共享工程的建设与运行管理,国家设立“科学数据共享工程专项计划”,对科学数据共享工程的建设与运行给予稳定、足额的财政支持。

实施科学数据共享工程,使作为信息资源重要组成部分的科学数据面向全社会共享,有助于提高全民的科学素质;有利于对国家资助获取的数据资源进行有效管理,避免资源的流失,减少国家为同一目标重复投入,维护和提高国家投入的整体效益;有利于对国际科学数据资源的有效使用,分享全人类的科技成果;有利于推动社会信息化,培育科学数据服务业,带动信息产业化。随着全社会对科学数据信息资源共享观念的增强与共享机制的健全,有可能减少重复获取数据的投入,从而提高科研经费的使用效率。随着数据共享工程的实施与健康运行,将对中国 GDP 年增长率的拉动产生重要作用。

#### 参考文献

- [1]黄鼎成、郭增艳主编.科学数据共享管理研究.北京:中国科学技术出版社,2002
- [2]孙九林等.我国科技数据管理和共享服务的新进展.世界科技研究与发展,2002,(5):15-19



论文写作，论文降重，  
论文格式排版，论文发表，  
专业硕博团队，十年论文服务经验



SCI期刊发表，论文润色，  
英文翻译，提供全流程发表支持  
全程美籍资深编辑顾问贴心服务

免费论文查重：<http://free.paperyy.com>

3亿免费文献下载：<http://www.ixueshu.com>

超值论文自动降重：[http://www.paperyy.com/reduce\\_repetition](http://www.paperyy.com/reduce_repetition)

PPT免费模版下载：<http://ppt.ixueshu.com>

---