

# 材料科学数据共享网及其在材料行业创新发展中的应用

尹海清<sup>1,2</sup> 姜雪<sup>1,2</sup> 张瑞杰<sup>1,2</sup> 刘国权<sup>1,2</sup> 曲选辉<sup>1,2,3</sup>

(1. 北京科技大学钢铁共性技术协同创新中心, 北京 100083; 2. 材料基因工程北京市重点实验室, 北京 100083; 3. 北京科技大学新材料技术研究院, 北京 100083)

**摘要:** 材料科学数据共享网是国家科技部科技基础条件平台建设项目, 旨在整合我国材料领域的的数据资源, 建立满足国家不同需求的材料科学数据体系和材料科学数据共享服务平台。本文主要阐述材料科学数据共享网项目取得的成果以及材料数据作为加快新材料的研发、降低材料成本的三大关键技术之一, 在材料数据科学这一新学科的提出以及支撑虚拟材料设计生产线的建设等方面的应用。

**关键词:** 材料数据; 材料基因组; 数据共享; 知识产权; 材料数据科学

中图分类号: TB39

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2016.03.009

## National Materials Scientific Data Sharing Network and Its Application to Innovative Development of Materials Industries

YIN Haiqing<sup>1,2</sup>, JIANG Xue<sup>1,2</sup>, ZHANG Ruijie<sup>1,2</sup>, LIU Guoquan<sup>1,2</sup>, QU Xuanhui<sup>1,2,3</sup>

(1. Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing, Beijing 100083; 2. Beijing Key Lab of Materials Genome Initiative, Beijing 100083; 3. Institute of Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083)

**Abstract:** Materials scientific data sharing network, a scientific data sharing platform in materials science and engineering, is financially supported by the National Science and Technology Infrastructure construction project of Ministry of science and Technology of China, aiming to collect the nationwide materials data resources and construct the materials data system and data sharing and service platform. In this paper, the progresses of materials scientific data sharing network were described, and the application of materials data, one of the three tools of materials innovation, in establishing the materials data science and virtual materials design and production line.

**Keywords:** materials data, materials genome, data sharing, intellectual property, materials data science.

---

**作者简介:** 尹海清\* (1971—), 女, 北京科技大学教授、博士生导师, 研究方向: 材料数据、粉末冶金; 姜雪 (1989—), 女, 北京科技大学讲师, 研究方向: 数据库; 张瑞杰 (1977—), 男, 北京科技大学副教授、硕士生导师, 研究方向: 材料微观组织模拟; 刘国权 (1952—), 男, 北京科技大学教授、博士生导师, 研究方向: 材料设计、材料数据; 曲选辉 (1960—), 男, 北京科技大学新材料技术研究院院长、教授、博士生导师, 研究方向: 粉末冶金。

**基金项目:** 国家科技部科技基条件平台建设项目“材料科学数据共享网”(2005DKA32800); 国家高技术研究发展计划“基于材料基因工程的高性能材料设计、制备与表征技术”(863计划)(2015AA034201); 国家自然科学基金项目“微米尺度氧化铝微型齿轮的微注射成形研究”(51172018); 得到美国肯纳金属有限公司的支持。

**收稿时间:** 2015年12月27日。

## 1 引言

基于数据的科学发现已经被认为是继科学实验、理论推导、计算机模拟仿真三种科研方式以外的新生的第四种科学研究方式<sup>[1]</sup>。材料是人类文明和社会发展的支柱，人类发展经历的旧石器时代、新石器时代、青铜器时代、铁器时代直至现代，材料始终贯穿其中。社会各行业的飞速发展需要材料的基础支撑。随着材料科学与工程的发展，以手册和数据库等非数字化和数字化形式存储的数据越来越多。我国非常重视数据库的建设，对材料数据库技术的研究开始于20世纪80年代初。当时建成了分布于高校、科研院所和企业用户各种各样、大大小小的材料数据库。清华大学、北京科技大学、中国地质大学、福州大学等高校，北京航空材料研究院、钢铁研究总院、有色金属研究总院以及中国科学院的金属研究所、过程工程研究所等科研院所，先后建立了不同材料的数据库。纵观时代发展，人们对数据信息的利用越来越迫切，在社会科学等方面取得了突出的效果，对生物医学数据的研究及成功应用也成为自然科学研究中的成功案例。而材料基因组计划的提出与实施，使材料数据为越来越多的材料工作者所认识并关注。本文针对材料数据的特点和需求，结合材料科学数据共享网科学数据共享平台，从我国材料数据库的历史与发展状况、材料科学数据共享网项目概况、在新形势下材料数据的发展及其应用等方面进行论述。

## 2 材料科学数据库研究的背景及意义

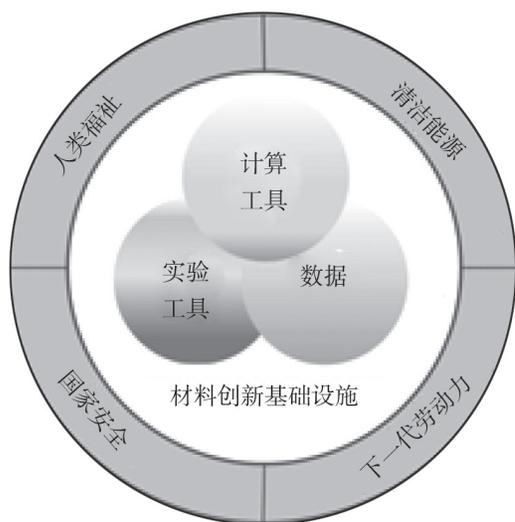
21世纪初，根据《2004—2010年国家科技基础条件平台建设纲要》和《“十一五”国家科技基础条件平台建设实施意见》的部署，由科技部、财政部共同组织实施，开展国家科技基础条件平台建设专项，重点支持6个领域的共享平台建设，形成包括研究实验基地和大型科学仪器设备、自然资源、科学数据、科技文献、网络科技环境等国家科技基础条件资源的整合共享，促进全社会科技资源高效配置和综合利用。

其中，科学数据共享平台的建设，旨在打破条块分割，对相关部门和行业长期持续积累的数据资源以及国家科技计划项目的数据进行整理、汇交和建库，构建集中与分布相结合的国家科学数据中心群，提高与国际科学数据组织的信息交换能力，推动面向各类创新主体的共享服务网建设，形成国家科学数据分级分类共享服务体系。

材料科学数据共享网是国家科技基础条件平台建设项目，以整合、重构现有的、较为成熟的材料科学数据资源为基础，构建材料科学数据共享服务平台，形成材料数据采集提交与存储、质量控制、开放共享的管理体系。20世纪80年代开始建设的一批材料数据库，大多数因学科性强、规模小、数据量少、服务范围小、缺乏更新等问题而导致无法持续运行服务，或在网络上无法找到相应的资源。还有少部分数据库资源被整合进入了材料科学数据共享网中，如中国科学院金属研究所的金属腐蚀数据等，后来汇交入库进入材料科学数据共享网的有色金属及其合金数据共享资源结点。为了避免上述问题再次出现，材料科学数据共享网项目建立异构分布、有序共享的材料数据体系，制定符合材料数据特点的共享机制和数据共享相关的标准规范，并在统一技术标准下，逐步建成面向不同领域服务的共享资源结点和应用主体数据库群。

2009年，Hey Tony基于Grey Jim的科学发现第四范式思想，提出数据驱动的数据密集型科学发现是继科学实验、理论推导、计算机模拟仿真三种科研方式以外的新生的第四种科学研究方式<sup>[1]</sup>，这标志着对数据的研究从此进入一个崭新的阶段。

2011年6月，美国发布了“材料基因组计划”<sup>[2-3]</sup>。该计划是“先进制造伙伴计划”中的核心内容之一，通过计算、实验与表征以及材料数据三大工具深度融合，如图1所示，使新材料从设计到新产品开发应用全链条的进程至少加快1倍，成本至少降低一半。材料基因组计划（我国称之为“材料基因工程”）作为一种新的材料学科方法论，被列为我国“十三五”重点研发专



项。材料基因组计划打破了材料计算的各尺度相对独立、计算与实验相对分离，尤其是数据价值不被重视的局面。其中高通量计算、高通量制备与表征的实施，将产生大量数据，使材料数据由海量向大数据的趋势发展。

材料数据已经作为材料领域第四种研究方法、材料基因组计划的工具之一，在2014年德国的“工业4.0”计划<sup>[4]</sup>和2015年我国的“中国制造2025”计划中成为实现材料智能设计制造的基础和知识载体。在以材料创新为主体需求的引导下，材料数据进一步显示出新的特点，主要表现为高通量、海量、跨尺度和复杂多维等，对于材料数据的收集、整合、存储、查询检索、管理、共享以及数据价值的深度开发都出了新的挑战。

### 3 材料科学数据共享网的研究现状及成果

材料科学数据共享网项目经过三年的建设期以及后续的稳定服务期，已经从立项时的仅有的一些金属腐蚀数据，发展成为目前拥有70余万条质量可靠的材料数据、拥有近20家高校和科研院所的100余位专兼职建设与服务人员并向全社会免费开放数据库和提供免费服务的材料数据平台。

材料科学数据共享网的数据，按照项目建设初期确立的材料科学数据资源体系，整合入库

涉及资源体系的所有材料学科数据，形成了材料基础、黑色金属、有色金属及其合金、高分子材料、复合材料、无机非金属材料、生物医用材料、能源材料、信息材料、天然材料及其制品等10个材料数据资源结点以及建筑材料、道路交通材料等2个应用主体库。其中，黑色金属和有色金属及其合金的数据量较大，这是由于金属材料发展的历程以及积累了大量数据的优势。材料科学数据共享网的数据资源呈现物理分布、逻辑统一的格局。

材料科学数据共享网的门户网站提供涉及2.4万余种材料的70多万条数据、材料数据处理工具以及29项材料数据及材料数据库建设相关的标准规范草案、5项管理规章制度和4项材料数据分析评价方法。提供了数据的查询、检索、下载，规范草案的下载，部分数据的处理以及专家咨询、材料设计与制备的研发等各项服务，服务的专家涉及参与建设的所有单位的科研技术人员，包括北京科技大学、中国科学院金属研究所、西北工业大学、中国科学院化学研究所、北京大学、北京师范大学、钢铁研究总院、北京工业大学、中南大学、中科院上海硅酸盐研究所、四川大学、北京有色金属研究总院、中国地质大学、建筑材料工业技术情报研究所、长安大学、上海电力学院共16家单位，服务覆盖材料数据资源体系中的所有材料学科，并编制了标准规范（草案）。

材料科学数据共享网服务平台在2010年建成上线，向全社会全天候免费开放并提供数据共享服务，实现了物理分布、逻辑统一的数据库模式，可以通过中心网站进入分布数据库，也可以对分布数据库进行单独访问。对中心网站的访问量呈现逐年递增的态势，如表1所示。截至2015年年底，网站访问量近24万次，其中2015年访问量达到10万次，较上线后的前几年有较明显的增长，其中来自国内的访问量占总量的93.5%。这一比例也是随着访问量的增加而呈现增长的趋势，较2014年的国内访问量增长6个百分点。分析其原因，可能一方面在于大数据在社会多方面

的应用形成的大数据时代这一社会环境的影响，另一方面在于材料基因工程相关热门话题及其讨论在我国受到越来越多人的关注，材料数据也成为一个新的研究热点。对国内不同区域的访问情况分析，如表2可以看到：来自北京的用户访问量占国内总访问量的82%，所占比例远超过其他地区，其原因可能在于北京在高校和科研院所的数量上具有明显优势，对材料领域的研究覆盖面较广，因此，对于由中国工程院和中国科学院发起的材料基因组计划的研讨中参与人员较多，对不同学科的材料数据的关注、了解与需求量增大，导致浏览量的激增。从用户使用的浏览器看，如表3所示，使用最多的三种浏览器分别为google、firefox和IE8，利用上述浏览器的访问量占中心网站访问总量的62%。

除中心网站外，材料科学数据共享网的分布数据源在中心网站的宣传下，也受到了关注。表4、图2显示了部署在沈阳的有色金属及其合金材料数据资源结点的访问量与用户来源情况。2015年上半年访问量统计约为17.5万次，较前一年同期略有降低，其中来自中国和美国的用户占总用户量的98%以上。

材料科学数据共享网在数据整合入库与服务用户的各类需求的同时，逐步建立和完善材料数据集材料数据库的基础设施，为材料数据的长期可持续发展提供坚实的专业理论基础。

(1)材料数据资源体系的确立。由李依依院士等10余位院士和顶级专家讨论确立的以师昌绪先生主编的《材料大辞典》<sup>[5]</sup>和材料领域中长期科技发展规划为基础的材料数据资源体系，包括材料基础、金属材料及特种合金、黑色金属、复合材料、有机高分子材料、无机非金属材料、生物医用材料、信息材料、能源材料、天然材料及其制品等10大类材料数据。这一材料数据体系，在国际上尚属首个。多年的建设实践及与国际同行的交流证明了该材料数据体系的科学性、全面性和权威性，并为今后数据科学在材料领域中的建立奠定了知识体系基础。

(2)依据数据来源的材料数据质量控制。材料科学数据共享网的数据来源包括专业书籍、工具书、手册、标准、年鉴、专利数据、期刊论文、学位论文，参加单位的测试数据、热力学计算数据以及行业、生产厂家的生产数据等。通过《材料科学数据质量评估标准》(草案)的制定，

表1 材料科学数据共享网中心网站近两年的访问量统计情况

统计年度	2015年度			2014年度		
	总访问量	其中来自国内	其中来自国外	总访问量	其中来自国内	其中来自国外
访问量/次	237901	222520	15381	131063	113517	17546
占总访问量的比例/%	----	93.5	6.5	----	86.9	13.1

表2 材料科学数据共享网中心网站的国内用户来源统计

地区	北京	广东	河北	河南	江苏	天津	辽宁
用户量/人	181790	4259	1249	1766	4174	1190	1587
用户占比/%	81.7	1.9	0.6	0.8	1.9	0.5	0.7
地区	山东	陕西	上海	四川	浙江	其他	
用户量/人	1826	2511	7619	3578	2223	25598	
用户占比/%	0.8	1.1	3.4	1.6	1.0	11.5	

表3 材料科学数据共享网中心网站的用户浏览器统计

浏览器类别	firefox	theworld	aoyou	google	ie6	ie7	ie8	ie9	360	other
访问次数	43833	345	10436	74227	5421	23594	30762	13298	7285	30169
访问量占比/%	18.31	0.14	4.36	31.01	2.26	9.86	12.85	5.56	3.04	12.6

表4 有色金属及其合金材料数据资源结点的用户量统计

时间段	独立IP数/个	访问人次/次	页面访问数/页	请求数/个	下载量/GB
2014年06月	4369	9040	33000	59299	1.97
2014年07月	3830	7134	40772	60098	2.16
2014年08月	2945	5870	7253	17845	0.52
2014年09月	1077	4110	5860	17288	3.40
2014年10月	844	4950	42943	65177	2.33
2014年11月	1117	5541	39265	71983	2.34
2014年12月	1170	7046	42674	75762	2.43
2015年01月	1239	7678	22805	50985	1.48
2015年02月	1762	6814	24890	47838	1.26
2015年03月	1953	8690	23150	56262	1.35
2015年04月	2144	8553	26783	57957	1.35
2015年05月	2096	8956	28669	58933	1.60
2015年06月	1409	8777	49499	90649	2.59
总计	26226	91259	386553	720077	21.67

数据来源统计



图2 有色金属及其合金材料数据资源用户来源统计

确定材料数据可靠性分为完全可靠、一般可靠和有待验证3个等级。不同来源的数据其可靠性不同，通过每条数据标注包含数据的所有来源信息，同时保证数据的可追溯性。

(3) 材料性能数据的分类整理。通过对各类材料性能数据的系统分类研究，将材料性能分为物理、化学、力学、电学、光学、热学、磁学、电化学等性能、生物学性能以及环境性能、工艺性能、晶体结构性能与其他性能等13类，共整理了不同类型数据580余个，其中包括163个力学性能、89个物理性能、61个化学性能、48个电学性能等。

(4) 材料科学数据提交格式的标准化。通过《材料科学数据提交格式规范》(草稿)的制定，针对不同来源的数据进行了分类，明确规定了材料计算数据、实验数据和生产数据三类数据应包含的全部信息，以满足材料数据的完整性的根本

要求，使材料数据为材料设计和研发创新服务。

(5) 材料数据的知识产权保护。DOI (Digital Objects Identifier)<sup>[6]</sup>，即材料数据数字对象唯一标识符，具有唯一性、永久性、互操作性以及动态更新等特点，成为数据的“身份证”。通过制定《材料数据对象唯一标识符命名与注册规范》(草稿)，研究材料数据DOI的注册机制，制定材料科学数据DOI命名规范一项，按照材料数据分类体系、数据关键信息对材料数据DOI命名进行标准化，保证数据在共享过程中的知识产权保护以及个人利益不被侵害。数字对象唯一标识符成为目前解决数据共享和数据出版的最有效的途径之一。

#### 4 材料数据在材料发展中的应用

自材料科学数据共享网项目实施以来，一直本着“边建设边应用”的原则，不断提高材料

数据的整合利用，同时加强材料数据的应用与服务，在国家大型工程的材料选材与结构设计、监控及无损检测，城市基础设施安全服役，行业创新发展的核心技术开发与工程建设、新材料的设计研发与应用，国际合作与交流以及科研、教学及人才培养等方面发挥了重要的支撑作用，平均每年接到来自用户的直接反馈约8000余件，用户人群覆盖高校、科研院所、市政部门、公司企业等方面。在过去一年多的时间里，材料科学数据共享网为来自不同方面的需求提供了科学而系统的数据服务，典型应用主要有以下几个方面。

#### (1) 对重大工程选材设计与健康监控的支撑

在航天飞行器热结构设计方面，为北京临近空间飞行器系统工程研究所承担的某临近空间超高速飞行器的机翼前缘等热结构的材料筛选、结构与评价等提供了耐高温陶瓷基复合材料的基本力学性能、物理性能、700℃~1300℃的高温力学性能等关键性能参数，为实现防热—结构一体化设计提供了有力的支撑，节省了项目研究的成本和设计周期。

在复合材料飞机结构损伤监测方面，为中国飞机强度研究所开展了飞机结构健康监控等研究，提供了聚合物基复合材料的基本力学性能、物理性能数据以及结构损伤监测相关的国内外技术标准，为飞机结构健康监控系统设计和结构无损检测提供了有力的支撑。

在大型工程的环境服役条件下材料设计、寿命预测评估及服役性能检测方面，基于金属材料在大气及不同水域的腐蚀性能数据及各类技术标准，完成了港—珠—澳跨海桥隧等工程在不同气候条件下的金属材料的选材及其防腐设计、材料寿命预测及评估、工程施工以及材料在使用过程中的性能监测。

#### (2) 对城市基础设施安全服役的支撑

城市燃气管网设施的发展是城市现代化的重要标志，确保城市燃气管网安全运行更是发展城市基础设施安全服役的关键性问题。在材料科学数据共享网数据支持下，由北京科技大学与北京市燃气集团有限责任公司合作研制的城市燃气

管网服役安全信息支撑数据系统，建立了管网地理信息数据库、管道材料数据库、焊缝材料数据库、防腐层材料数据库、土壤环境数据库、腐蚀案例库，为城市燃气管网的安全检测、诊断、信息管理、服役安全和预警提供信息化支撑，至今仍是进行首都燃气管网服役安全评估、确保管网安全运行的重要基础。

#### (3) 对创新驱动需求牵引的新材料的设计开发的支撑

基于材料科学数据共享网的稀土金属与合金基础数据库的数据资源，北京工业大学在硬质合金刀具的研发和生产中取得突破，成功设计并采用原位合成技术和纳米粉末造粒技术，开发了热喷涂用纳米WC-Co硬质合金复合粉末，并成功制备了煤矿用支撑掩护式液压支架活塞杆的表面涂层，替代了传统使用的电镀硬铬层，完全满足了液压支架活塞杆的表面防护的实际需求，与Praxair公司的国际先进同类产品的涂层相比，磨损速率降低了30%~50%。

#### (4) 对《材料基因工程》重大专项立项的支撑

在美国提出材料基因组计划后，我国材料科学专家从2011年开始对该计划进行深入研讨，并逐步形成了我国的思想体系，并分别于2013—2014年由中国工程院和中国科学院组织撰写了材料基因工程重大咨询项目，其中的建设材料数据库平台及数据技术部分就是依托材料科学数据共享网思想和成果完成的，为材料基因组计划在我国成功落地起到支撑作用。

#### (5) 对国际合作与交流的支撑

2010年，依托材料科学数据共享网，北京科技大学成功承办和协办了第四届亚洲材料数据会议，并发起成立了亚洲材料数据委员会，确立了我国材料数据在亚洲的三强地位，并多次与国际科学技术数据委员会(CODATA)进行深入的交流与探讨。

#### (6) 虚拟材料设计生产线

基于材料海量数据的整合、材料的本体、数据的多尺度跨层次无障碍传输等形成的虚拟材料

设计生产线,是今后“材料+互联网”的直接体现,也是工业4.0思想在材料领域应用的基础,对于材料基因组计划提出的新材料研发进程加快一倍的目标的实现具有强大的助推作用,对于钢铁材料等关系国家命脉但目前行业形势很不乐观的领域,无疑是解决当前低端产品产能过剩、同质化严重、高端产品性能不稳定等问题的方法。这些工作已经成为钢铁共性技术协同创新中心的主体工作之一,以期实现新技术的工程应用。

## 5 问题与对策

材料科学数据共享网作为科技基础条件平台,是材料科学数据存储、开放、服务的平台,不仅基本实现了项目的整合、重构现有材料科学数据资源,建立满足国家不同需求的跨部门、跨地区、异构分布、有序共享的材料科学数据体系的建设目标,而且不断推动材料数据库成为材料创新的基础设施和有效工具,并在大数据时代和数据驱动的创新活动中发挥重要作用。材料科学数据共享网的建设和发展,有效避免了重复实验,缩短了材料设计周期,降低了设计与服役评价的成本,创造了巨大的经济价值,减少了国家的重复投入,同时,有效避免了信息孤岛问题,有助于政府决策的及时化和科学化,并且有利于对国际科学数据资源的贡献和有效利用,分享全人类的科技成果,产生巨大的社会效益。

然而,材料科学数据共享网在发展过程中,在数据源及数据应用等方面出现的新问题,对于材料数据的发展来讲是极大的挑战,也是材料数据库今后发展所面临的普遍问题,主要表现在以下几个方面。

### (1) 材料数据的知识产权属性及其保护

数据的知识产权属性在一定程度上给数据共享带来困难,因此,在一些领域出现了大数据时代没有数据的状况。大量的材料数据往往与技术秘密、生产效益和经济利益密切相关,赋予材料数据很强的知识产权属性,并隐含着巨大的经济效益并关乎国家安全。同时,我国关于数据知识产权的界定与保护方面的法律法规尚不健全与完

善,给数据共享带来相当的难度。

### (2) 材料数据的整合与工作成果评价

材料领域划分的学科较多,涉及的材料种类众多,而材料数据的来源丰富、分布广泛,难以像气象和地球等领域的原始数据那样通过大量仪器自动采集直接获取。目前,大部分材料数据的获得是依赖人工的内容理解和数据提取,而且要求采集人员具有较高的领域知识水平。然而,大量的人力投入一方面难以换来高影响因子的文章,另一方面往往无法写成文章甚至写成的文章无期刊接收。这对于以论文作为考核指标的高校和科研院所的科研人员来讲基本上是无科研成果。这种投入与成果评价的严重失衡,难以维持数据整合者的工作热情。

### (3) 数据应用服务需求多样化

材料科学数据共享网立项时的建设目标是在不了解其他类型材料时可以通过跨库全文检索来实现材料的跨类型的选材优化。然而,随着经济的飞速发展,对材料的创新需求越来越多。一方面,用户来自材料的不同学科和不同行业,而材料科学数据共享网覆盖材料领域的各个方面;另一方面,对从材料数据分析中获得更多信息量的新的需求也越来越高,尤其在材料基因组计划提出后。但除了材料科学数据共享网的黑色金属数据共享资源节点中对淬透性的预测中运用了人工神经网络(ANN)<sup>[7]</sup>等数据分析技术外,无论国内还是国际上,材料数据库建设者都未曾涉猎材料数据分析或材料信息学<sup>[8]</sup>。

针对上述存在问题,总结材料科学数据共享网以及国外有影响力的材料数据库的建设经验,今后在材料数据及材料数据库的建设发展中,可以在以下几个方面开展深入研究。

### (1) 推行材料数据知识产权保护的切实可行的措施

只有明确了成果的知识产权,才能使数据拥有者愿意将自己的成果拿出来共享。积极学习期刊、论文等电子出版物的知识产权保护的相关法律、政策、方法、技术等,明确科学数据知识产权的法律界定,不断探索数据的知识产权保护,

学习并推广材料数据DOI体系在材料科学数据共享网建设单位中的普及，并对有数据资源的单位和个人加大宣传，保护材料数据的知识产权。同时，加大数据共享的机理机制建设，为数据共享奠定基础。

#### (2) 推广数据出版和数据引用的理念

只有当数据整合这项基础工作得到全社会，尤其是高校和科研院所的评价体系认可时，才能有更多的人愿意参与到材料数据共享这项工作中来。由国际科学技术数据委员会(CODATA)中方委员会2015年创办的《中国科学数据》杂志，深入探讨了科学数据出版的问题。同时，中国科学技术信息研究所及CODATA中委会等单位，在科学数据引用上也进行了积极的探索与尝试，希望通过建立数据引用影响因子(Data Citation Index, DCI)，最终实现被科研机构评价体系接受的、与SCI同地位的DCI评价体系。

#### (3) 各类材料数据整合与国家急需材料或热点话题相关数据整合两手都要抓

材料科学数据共享网在前期的建设中，已经构建和完善了材料数据知识体系。一方面作为国家基础设施和资源建设内容，材料数据知识体系覆盖的内容都要建设；另一方面不断拓展和深化材料数据，在国家安全和经济建设中或在热点专题上做好材料数据整合和推广应用。在夯实材料数据库基础的前提下，不断以需求为牵引带动材料数据的快速应用型发展。

#### (4) 材料数据价值的挖掘与跨学科协同

科学发现第四范式的提出，为当今的数据时代指出了新的研究方式，为材料数据的发展和应用提出了新的思路。材料数据挖掘与知识的深度提取是数据驱动的材料研究，是材料信息学的研究核心，通过不同的技术，研究数据在不同学科、不同研究命题中的深入应用，来发现新材料、发掘新规律等。材料数据的本体<sup>[9-10]</sup>以及统计分析方法、机器学习方法<sup>[11]</sup>、神经网络方法等不同挖掘方法，均是今后需要深入研究并突破的关键问题，而这一工作具有典型的交叉学科的特点，对材料、计算机、数学、物理、化学的研

究充分融合。依托材料科学数据共享网项目的成果，以及国家十三五专项支持的材料基因工程实施后可能产生的高通量计算和高通量表征数据<sup>[12]</sup>的分析与挖掘，可以获得新材料的发现以及新工艺的突破<sup>[13-15]</sup>，这就要求材料数据库建设者在数据价值的深度挖掘上需要率先迈出第一步，材料信息学和材料数据科学可能将成为今后材料科学与工程人才培养的重要方向之一。

## 6 结论

材料科学数据共享网是国家科技基础条件平台建设项目。通过整合大量材料数据，形成了国家级材料数据库，并通过材料科学数据共享网门户网站对外发布，支撑材料选材，加快材料设计与优化，对国民经济建设、国家安全、国家重大工程等方面具有重要意义，避免了材料信息孤岛，其系统性和科学性在促进材料基因工程及其材料数据工具发展方面发挥指导作用。材料数据在大数据时代正扮演着越来越重要的角色，推动着材料科学创新的纵深发展，与信息等领域的交叉融合对材料数据工作者提出了更大的挑战。基于材料科学数据共享网整合的材料数据的分析与挖掘，将满足知识的数字化存储、计算、实验与表征的交互，同时形成数据驱动的科学发现研究模式。今后，材料科学数据共享网将继续加大数据整合与数据共享的力度，在发展材料数据科学的同时，加大材料数据的应用潜力的开发，以在材料基因工程等方面实现更深入的应用。

## 参考文献

- [1] HEY Tony, TANSLEY Stewart, TOLLE Kristin. The fourth paradigm: data-intensive scientific discovery[M]. Washington: Microsoft Corporation, 2009: 109-130.
- [2] CHEN Liquan. The materials genome initiative and advanced materials[J]. Engineering, 2015, 413(2): 169-169.
- [3] JUAN Pablo J de, BARBARA Jones, CORA Kovacs Lind, et al. The materials genome initiative, the interplay

(下转第71页)

大、复杂,人始终是这一网络的核心,无论任何时代、任何潮流,人类社会都以人为最主要的核心。技术越是进步,越应该为人类带来方便。所以科技信息资源共享工作应该充分考虑用户的主体性和参与性,使用户充分融入到科研工作的各个环节,整合集体智慧,创造更大的价值。

第三,国家应加强政策引导。国家应该合理引导科技信息资源的宏观布局,积极制定科技信息资源共享的方针和政策,充分激励不同科研主体进行深度合作。另外,在基础平台建设方面,国家应该加大资金和人才投入,为全国性的科技信息资源广泛、深度的共享打下坚实基础。

### 参考文献

[1] 欧阳日辉. 从“+互联网”到“互联网+”: 技术革命

如何孕育新型经济社会形态[J]. 人民论坛·学术前沿, 2015(10): 25-38.

- [2] 于扬. 创业者需要找到自己的互联网+[EB/OL]. [2015-03-20]. <http://it.sohu.com/20121114/n357591531.shtml>.
- [3] 师榕. 以“互联网+”为驱动, 推进我国经济社会创新发展[EB/OL]. [2015-03-10]. [http://www.qsttheory.cn/subject/2015-03/10/c\\_1114592028.html](http://www.qsttheory.cn/subject/2015-03/10/c_1114592028.html).
- [4] 马化腾. “互联网+”是种能力激活更多信息能源[EB/OL]. (2015-04-29) [2015-05-06]. <http://tencent-research.com/Article/lists/id/3877.html>.
- [5] 阿里研究院. “互联网+”研究报告[EB/OL], [2015-03-10]. <http://www.askci.com/news/chanye/2015/03/13/85637vpqv.shtml>.
- [6] 赵伟, 彭洁, 张新民, 等. 科技信息资源共享的约束与调控分析[J]. 中国基础科学, 2012(5): 34-38.
- [7] 王晓丽, 刘洁, 孙洁. 科技查新信息资源共享探讨[J]. 中国科技资源导刊, 2016, 48(1): 100-104.

(上接第65页)

- of experiment, theory and computation[J]. Current Opinion in Solid State and Materials Science, 2014(18): 99-117.
- [4] NATIONAL ACADEMY OF SCIENCE AND TECHNOLOGY. Recommendations for implementing the strategic initiative industrie 4.0[R]. 2013, 4.
- [5] 师昌绪. 材料大辞典[M]. 北京: 化学工业出版社, 1994.
- [6] GORRAIZ Juan, MELERO-FUENTES David, GUMPENBERGER Christian, et al. Availability of digital object identifiers (DOIs) in Web of science and scopus[J]. Journal of Informatics, 2016, 10(1): 98-109.
- [7] 陈蕴博, 左秀荣, 王森辉, 等. 人工神经网络在钢铁材料研究中的应用[J]. 材料导报, 2009, 23(4): 1-3.
- [8] RAJAN Krishna. Informatics for materials science and engineering[M]. Netherland: Elsevier, 2013.
- [9] ASHINO Toshihiro. Materials ontology: an infrastructure for exchanging materials information and knowledge[J]. Data Sci. J., 2010, 8: 54-61.

- [10] ZHANG Xiaoming, ZHAO Chongchong, WANG Xiang. A survey on knowledge representation in materials science and engineering: An ontological perspective[J]. Computers in Industry, 2015, 73: 8-22.
- [11] CERIOTTI Michele. Machines learn to recognize glasses[J]. Nature, 2016, 12: 377-378.
- [12] NOSENGO Nicola. The materials code[J]. Nature, 2016, 533: 22-25.
- [13] OLSON B G. Genomic materials design: the ferrous frontier[J]. Acta Materialia, 2013, 61: 771-781.
- [14] ALAN A Luo. Material design and development: From classical thermodynamics to CALPHAD and ICME approaches[J]. CALPHAD: Computer Coupling Of Phase Diagram And Thermo Chemistry, 2015, 50: 6-22.
- [15] AGRAWAL Ankit, DESHPANDE D Parijat, CECEN Ahmet, et al. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters[J]. Integrating Materials and Manufacturing Innovation, 2014 (3): 3-19.