

中国材料数据库与公共服务平台建设

尹海清, 刘国权, 姜雪, 张瑞杰, 曲选辉

北京科技大学新材料技术研究院, 北京 100083

摘要 材料数据对于国家安全、工程服役安全、科技创新、智能制造等方面的重要性在数据时代越来越彰显出来,在2011年美国提出的具有变革意义的材料基因组计划中,材料数据与材料计算模拟、材料实验表征一起,为材料发展全流程研究的三大基本工具,使材料研究者与生产管理者进一步充分认识材料数据对加速材料研发进程的推动作用。材料数据具有多样、获取过程复杂、数据间关联关系复杂、知识产权性强等特点,使数据的收集、存储、共享和应用更加复杂。本文就大数据时代下的材料数据的特点、分类、材料数据库的国内外现状对比、中国发展材料数据库与材料数据科学的意义、今后主要发展方向以及存在的问题等方面进行系统分析,提出建设国家材料数据研发与服务公共平台,加大材料数据的收集整合力度,构建国家民用材料数据库与军用材料数据库,开展材料数据及材料数据库相关标准规范建设、定制性专题数据库服务、数据推送服务,同时开展材料信息学、材料数据学等方面的研究,开创并构建材料数据科学这一材料领域新学科。

关键词 材料数据;材料基因组;材料数据库;数据挖掘;智能制造;材料数据科学

中图分类号 N37

文献标志码 A

doi 10.3981/j.issn.1000-7857.2015.10.004

Materials databases and constructing national public service platform of materials data

YIN Haiqing, LIU Guoquan, JIANG Xue, ZHANG Ruijie, QU Xuanhui

Institute of Advanced Materials, University of Science and Technology Beijing, Beijing 100083, China

Abstract Materials data play an increasingly vital role in national security, performance safety, scientific and technological innovation and smart manufacturing in the age of information technology. In 2011, the Materials Genome Initiative (MGI) was launched in the US, of which materials data together with materials computation and materials experimentation and characterization consist of the three tools for accelerating the materials development continuum and reducing the cost. Both the researchers and production managers come to realize the significant role materials data play. The attributes of materials data, such as variety, complex interrelationship, acquiring process as well as the intellectual property issues, drive the process of collection, storage and application ever more complicated. In this paper, the characteristics, classification and status quo of materials data are described. The strategies and obstacles are systematically analyzed for materials data development and database construction. A national platform of materials data for research and public service, a materials data hub in China, is essential and urgently required for MGI implementation. Four aspects are emphasized, that is, materials data repository, infrastructure and cloud service, data mining and international collaboration. The materials data repositories for civil and military uses will be constructed. On the platform, the standards are crucial for materials data and database and big data application, which need to be set up first. The customized database and data push will bring great benefit for materials database users. Materials data science will definitely become a brand new subject in materials science, including materials informatics and materials dataology.

Keywords materials date; materials genome initiative; materials database; data mining; smart manufacturing; materials data science

信息技术和互联网的飞速发展,彻底改变了人们的生活方式,每天都发生各种各样的变化。海量数据(mass data)在

各行业涌现,数据激增速度不断提高,形成大数据的态势,数据从静态变为动态、从简单的多维度变为巨量维度,种类日

收稿日期:2015-04-02;修回日期:2015-04-15

基金项目:国家高技术研究发展计划(863计划)项目(SS2015AA034201);国家自然科学基金项目(51172018)

作者简介:尹海清,教授,研究方向为材料数据与材料设计,电子信箱:hqyin@ustb.edu.cn

引用格式:尹海清,刘国权,姜雪,等.中国材料数据库与公共服务平台建设[J].科技导报,2015,33(10):50-59.

益丰富。大数据(big data)和云计算的发展为社会科学和人文科学带来新的发展方式。2012年世界经济论坛发布的报告中指出,大数据为新财富,价值堪比石油。大数据之父舍恩伯格预测,数据列入企业资产负债表只是时间问题。

2014年初,北京中关村启动了首个大数据交易平台,宣告数据资产正式进入中国市场。《华尔街日报》指出,大数据、智能化生产和无线网络革命将成为引领未来繁荣的三大技术变革。哈佛大学教授Gary King指出,大数据将在学术界、产业界和政界迅速蔓延开来,没有哪个领域不会受到影响。Hey等^[1]在2009年就提出,除科学实验、理论推导、计算机模拟仿真三种科研方式外,基于数据的科学发现成为科学研究的第4种方式。2014年初,英国皇家学院也宣布成立数据科学研究所。

牛津大学教授舍恩伯格在《大数据时代》一书中指出,大数据正在改变我们的生活以及理解世界的方式,成为新发明和新服务的源泉^[2]。近几年,美国先后部署了三大计划:2011年提出先进伙伴制造计划,6月12日发布该计划中重要部分,即材料基因组计划(materials genome initiative, MGI);2012年3月29日提出大数据计划;2013年2月12日提出的脑活动图谱计划,又称脑科学计划。这三大计划都与数据有极强的关系,在全球科学界引发了越发激烈的数据大战和数据科学浪潮。

1 材料领域的的数据

1.1 材料数据对于材料领域发展的应用需求分析

飞速发展的互联网和大数据时代的到来对很多研究领域形成巨大的甚至是颠覆性的影响,材料数据做为材料基因组计划的三大基础工具之一,连同材料的计算、实验与验证一起,支撑着国家安全和经济发展。对材料学科这一从观察和实验开始,最终以有形的材料服务于社会的学科,关于材料数据的思考很有必要。MGI吸引了全球越来越多的人开始用MGI思想,构想各自学科中将计算、数据和实验三者整合在一起,材料科学数据的应用主要表现在以下几个方面。

1.1.1 新材料与新功能的发现与设计离不开材料数据和数据库平台的支撑

社会的高速发展,要求支撑各行业基础的材料领域同样具有高速发展的步伐,才能满足国防新型武器、空天新型飞行器和探测器、各行业的产业升级换代的迫切需求。新材料的发现、现有材料的性能提升和新性能发现,必须依靠创新性的研发模式,必须改变“炒菜式”和“试错法”的研发方式。材料计算与模拟仿真、数据分析发现共性和个性问题与现象,是克服上述问题的最有效手段。

1) 材料计算与模拟仿真的原始数据需求。

大型计算机和工作站满足了材料计算与模拟的硬件要求,第一性原理计算、VASP、Materials Studio、Thermo-Calc & Dictra、Abaqus、Ansys等研究方法,与国外大型软件暂时满足

了国内材料计算与模拟的软件需求,这些软件需要与包含了大量数据的数据库相配套,才能开展计算(如与第一性原理计算相配套的数据库有200余个)。这些国外软件配套的数据库在早期是开放式的,允许自有数据的添加,然而近年来对中国用户全部关闭,只可使用,而无法看到数据,更不能添加新数据,如与Thermocalc软件配套的热力学数据库等。

目前国内已积累了少量的相应数据,但不够系统,也没有商业数据库。无法自行更新的国外数据库及其随时可能发生的停止对中国的销售,都严重制约了材料计算和新材料发现的进程。美国在材料相关关键数据的发布和共享上,更是采取极为严格的控制措施,数据发现后的3年时间内严格保密,不允许对外发布。因此,严重依赖国外数据的材料研发使中国材料从根本上很难全面摆脱“跟从”国外材料发展的局面,只有整合和建设具有自主知识产权的数据库,支撑各尺度层次下的材料计算和设计,才能保证中国材料领域的新材料的创新与发现。

2) 材料计算的有效性和精确性的验证,需要大量实验、生产和产品数据的验证。

材料计算与模拟的结果,如果缺少实验数据的验证,则难以为他人所信服。在材料计算和模拟中做出的一些假设、模型的选择、边界条件、初始条件和计算变量的设定,都将影响计算和模拟结果的精度。因此,以实验和生产数据验证计算和模拟结果,是不可缺少的环节。

1.1.2 特殊条件、特殊要求下的材料性能、机理等数据的获得,需要数据和数据库平台的支撑

对特殊条件下使用的材料的研究,如:1) 高温、高压、多场耦合、极端及破坏性条件下,材料研究难于甚至无法进行实验,如失重状态;2) 实验过程长,获得结果的时间漫长,如材料腐蚀性能;或实验原料稀缺,实验耗资高,如稀土材料等战略资源;3) 发掘材料机理,如金刚石表面沉积过程中十几个可能发生的反应判定,在几十微米尺度的微注射成形的喂料充型过程和缺陷形成机理等。数据支撑下的模拟仿真已成为解决这类问题的有效手段。

1.1.3 现有材料的性能提升离不开材料数据和数据库平台的支撑

中国关键材料的自给率不足20%,材料领域的大部分核心技术尚未掌握,在关键产品部件的生产上仍依赖进口或过期失效的国外专利技术(与国外至少有20年的差距),如催化剂、粉末高温合金、飞机和高铁刹车片等。国外高温合金的成分有100种之多,元素含量控制范围很窄,其中O、N、S等杂质含量精确在 $1\sim 10^{-5}$,甚至 5×10^{-6} 以下。要做到如此精确地控制几十种合金元素的含量,则需要大量实验数据支持的材料热力学、动力学、相场等多尺度计算和制备加工过程的模拟,以及少量实验的低成本、高效率的开发。

1.1.4 材料选材和安全评价离不开材料数据和数据库平台的支撑

国家重大工程、国家安全武器装备、城市基础设施等的

选材和安全评价均离不开大量的、系统性的材料数据,如天宫一号的一个关键部件在地面实验中出现问题,急需在上天前半个月时间内对该部件的安全性作出评估,正是依靠国家材料环境腐蚀平台积累的大量数据,给出了准确的服役评价,保证了天宫一号顺利完成任务。

1.1.5 通过材料数据的分析与挖掘获取材料知识是材料领域科学发现的新途径

大数据时代中,数据科学是科学发现的新的的重要手段。目前材料数据已呈现爆炸式增长的态势,高通量计算与高通量合成表征是产生大量数据的新的源泉,如基于第一性原理的对元素周期表多元组合的大量并行计算,产生巨量的非结构化的流数据,而组合芯片方法^[3]在一次实验获得10000个不同成分样品,对每个样品的实时检测获得的材料结构等信息,无不成为材料大数据的来源,不仅对计算硬件提出高要求,对数据处理的要求更高,数据分析与挖掘技术成为有效手段。

材料数据是材料设计计算与模拟的基础,支撑高校、科研院所和大型企业的材料创新^[4-6]。材料数据库是材料领域的庞大知识资源最便捷的保存方式,更是未来材料智能制造(smart manufacturing,也称数字制造,digital manufacturing)^[7]的基础。目前材料相关知识的碎片化非常严重,大量研究成果未能很好地保存和利用,国家重复投入、研究进展缓慢。数据库以其存储和调用便捷、及时性强、重用性高等特点,应该也将成为大数据时代的材料知识的高效存储方式。

1.2 材料数据的特点

材料数据在发展过程中形成了其独有的特点,主要表现为:1)材料数据的可靠性要求高。数据的质量不但关系到国防装备、重大工程的服役安全,也关系到工程、产品的质量与成本。而数据的质量只有通过专业手段进行评估;2)材料数据之间的关联性强,影响因素复杂。材料数据涉及从量子到宏观尺度的材料各种属性,而材料的化学组成、制备与加工工艺、环境等众多因素均会导致材料的组织结构与性能、使用效能等数据的变化;3)材料数据获取过程复杂。材料实验数据要通过特定试验设计、专用设备与技术、系列性的重复试验、观测分析才能获得。材料计算需要大量各类数据的支撑,同时又产生大量数据,而不同来源的“同种”材料的数据也可能有很大的差异;4)材料数据分散而且具有很强的知识产权属性。材料数据往往与技术秘密、生产效益和经济利益密切相关,给数据采集、整合与共享带来一定的难度。

1.3 材料数据的分类

从数据的生产来源来讲,材料数据可以分为计算数据、实验数据和生产数据3大类。该分类方法目前已应用于材料科学数据提交格式规范中。

从材料体系看,按《材料大辞典》^[8]的材料分类体系,材料数据分为材料基础、金属材料(黑色和有色)、无机非金属材料、有机高分子材料、复合材料、生物医用材料、信息材料、能源材料、天然材料等10大类。该分类方法目前已应用于材料

科学数据共享网和材料数据对象唯一标识符(DOI)命名与注册规范中。

2 材料数据的发展与应用的国内外对照

中国数字化的数据资源总量远远低于美国和欧洲等国家,每年新增数据量仅为美国的7%,欧洲的12%^[9]。材料领域涉及面宽,应用范围广,材料资源的数字化不仅是技术资源保存的最佳方式,更是智能化制造的基础。

2.1 材料数据库国内外研究现状

材料数据库的发展历程可以追溯到20世纪70年代中期,美国、欧洲、日本先后开展材料数据库的开发工作。目前美国国家标准局拥有数十个数据库,其中材料数据库占有很大比例,如力学性能数据库、金属弹性性能数据中心、晶体结构库、材料腐蚀数据中心、材料摩擦、磨损数据库、免费材料信息资源数据库。由美国的ASM International、NASA格林研究中心的寿命预测部门和Granta Design公司等成立材料数据管理联盟(MDMC),合作开发了Granta MI软件平台,核心任务是维护材料数据的高质量(包括可靠性、权威性等)、可跟踪性、完整性和安全性,保证高质量的数据共享。美国国防部国防技术信息中心建设了10余个国家级数据系统,包括先进材料及制备技术中心(AMPTIAC)等,形成庞大的材料工业数据管理和应用体系,为美国材料工业和制造业保持其世界领先的竞争力和可持续发展发挥重要作用。

欧洲各国开发建成了无机材料晶体结构库(ICSD),成为国际上开展第一性原理计算的支撑数据库,另外还有欧洲热化学数据库,荷兰PETTER欧洲研究中心的高温材料数据库(HT-DB)等数据库,英国有色金属数据中心、国家物理实验室等19个单位建有各自的材料性能数据库。

日本金属研究所、日本金属学会建有金属材料和复合材料力学性能数据库,日本国家材料科学研究院(NIMS)建立了“材料数据平台”(materials database station),现已建成材料基本特性数据库、聚合物数据库、金属与合金数据库、超导体数据库等多个高质量数据库并不断更新。

目前许多大型数据库已经实现了部分全球共享。例如,美国金属学会与英国金属学会合作开发了金属数据文档库,美国、英国、法国、德国、意大利、加拿大、日本等7国联合开发数据库计划(VAMAS)等。在MGI提出之际,美国及世界多国迅速反应,美国尤为强烈。白宫科技政策办公室在2012年的报告中“数据开放与计算工具的开发”是4个关键主题中的首个,在2014年的总结报告中将构建材料数据基础设施和数据的开放做为4大挑战之一单独提出,凸显材料数据的关键支撑作用。数亿美元的政府资助项目以及高校、公司资助的MGI项目陆续开展,材料数据及数据库均为这些项目的基础结构中不可缺少的部分,其中国家标准与技术研究院(NIST)强调数据与标准是材料创新的基础设施,两大主要工作均围绕着材料数据,研发模拟和实验数据的表达与交互作用、计算模型与数据质量评估的标准及工具,提出按照宏观、微观、

原子、电子4个尺度对数据进行了分类,并与ASM International 联合创建金属结构材料开放数据库,包含第一性原理计算、物相、量化表征、过程数据以及材料服役等数据。

中国从2001年底国家启动了科学数据共享工程,逐步建立不同领域的国家级数据共享网。材料领域的数据共享网站有两个,均依托北京科技大学建设,其中包括多种材料的国家材料环境腐蚀平台已积累了材料腐蚀数据40多万个,建立材料环境腐蚀数据库子库20个,数据已在天宫一号、天宫二号、大飞机工程等国家重大工程中获得应用。材料科学数据共享网的数据覆盖材料学科的全面,数据系统、成熟,牌号较全,适用于选材,已建成材料基础、金属材料、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、天然材料等材料的数据子库。建成了跨部门、跨地区、异构分布、有序共享的材料科学数据体系。国家材料服役安全科学中心已基本建成,重点建设重大工程材料服役安全评价的8套试验装置,建成后可生产大量材料服役行为数据,为重大工程设计和安全服役提供保障。

中国科学院在“十一五”信息化建设专项“数据应用环境建设与服务”项目的支持下,建立了一批数据库,其中材料相关的数据库包括化合物活性数据库、纳米数据库、物性及热化学数据库、纯化合物相变数据库、非电解质体系汽液相平衡数据库、共混聚合物相容性数据库、聚变材料数据库、储氢材料数据库、光学材料库、镀膜材料及膜系数据库、光学晶体数据库、古陶瓷数据库等。中国在航空航天、武器装备等材料上也建设了材料数据库,但使用范围较小。

2.2 对比国际发展中国材料数据的发展情况

1) 顶层设计超过发达国家。国家科技基础条件平台的共享工程,针对中国科技资源的碎片化,开展了大规模的数字化和整合工作,材料科学数据共享网的建设使中国成为世界上第一个拥有国家级材料科学数据库的国家,而日本在2014年4月刚刚批准建设国家材料数据中心计划,成为继中国后世界上第二个建设国家材料数据中心的国家。目前,材料科学数据共享网整理了561个性能指标,包括89个物理性能、163个力学性能等。已面对社会开放,数据量达61余万条,形成物理上分布、逻辑上统一的大型分布式数据库。

2) 部分专题数据库达到世界领先水平。国家材料环境腐蚀平台建设的材料腐蚀数据库已成为该学科世界领先的数据库,覆盖材料在大气、土壤和自然条件下腐蚀的实验数据。材料大科学工程项目建成后也将形成极端条件、大尺寸材料试验数据,对材料应用具有很高价值。

3) 中国数据库的数据量还不够,不够精细,在材料计算、模拟用材料数据库的建设上空白。尽管对国家973计划、863计划等项目成果的数据汇交的要求已实施多年,但成果数据化和标准化的环境仍未在材料科学界全面形成。碎片化资源的整合工作还有相当长的路要走。

4) 中国对材料数据的深度利用关注和开发不够。材料数据的深度开发与数据化息息相关,开发与利用的不足直接

影响了数据化的进程,中国20世纪80年代建设的诸多材料数据库,除部分后来纳入材料科学数据共享网外,其余大部分逐渐消失。

3 发展材料数据库与材料数据科学的意义

就目前情况而言,发展材料数据库与材料数据科学的意义主要分为两个方面。

1) 重要性。美国在经历了金融危机和经济震荡后,针对制造业低迷的态势提出材料基因组计划,作为先进制造伙伴计划的一部分,以解决制造业、占领制造业的世界制高点和社会需求的根本问题。而中国在经济放缓、钢铁卖出白菜价的时期,更应该顺应国际发展潮流,应该说,材料基因组计划思想更加适用于中国从“山寨”到自主创新过程的发展。材料数据是材料基因组计划的三大工具之一,具有不可替代的作用,而在数据时代背景下更是具有尤为突出的意义,预示着材料科学与时代大潮融为一体,并借助时代环境优势为自身发展的驱动力升级。

2) 紧迫性。中国在材料计算与模拟用材料数据库的建设上空白,在数据积累上与有几十年甚至百年历史的北美和欧洲相比,存在相当大的差距。然而,最近几年,在材料领域的国际合作与交流上出现了两种情况:一是在材料相关国际会议对大陆学者全部拒签的现象,二是美国等国在与日本等亚洲邻国合作时,对参与人员的身份限制很严格,未入籍的中国学者被直接排除在项目之外。而在新材料的市场销售上,出现的现象为,当中国无法自主生产时进口材料价格居高不下,但当听说中国可以生产时价格就直降为原价的30%,而当国产材料上市时,进口材料的价格进一步降低。可见,无论技术封锁,抑或是虚高价格,都使得决定材料数据库是否自主研发的关键时刻。

中国在材料领域的发展积累了大量成果,但目前可共享的数量还不足以建设材料计算用数据库,计算结果难于找到实验数据验证,导致计算成果在材料制备和生产中的应用步伐缓慢。资源碎片化导致管理上的信息缺乏和沟通不畅,科研投入重复严重,使得本应用于更多更深入创新工作的科研经费大量浪费,也导致在成果评价上出现不恰当的结论。

4 重点发展和建设方向

4.1 国家材料数据研发与服务公共平台建设的基本原则

1) 民用与军用相结合。依托国家材料科学数据共享网、全国材料环境腐蚀与老化站网、国家材料服役安全科学中心以及其他中国积累的材料数据库的基础上,建设民用材料数据库,提供各类服务并全部向军用材料需求开放;军用材料数据库服务于国家安全相关的材料研发对材料数据的需求,整合国内现有航空材料、军用武器装备材料数据以及其它军用材料数据库资源,建立并形成中国全面系统的军用材料数据仓库,形成国家材料知识库储备。

2) 集中与分布相结合。分布式异构数据库是处理数据

量和并发访问量激增、数据类型复杂等情况的数据库系统,具有物理分布性、逻辑整体性、站点自治性、集中与自治相结合的控制机制、数据分布透明性等特点,数据存储在多个站点,但逻辑上多个站点是一个整体,能够被所有用户共享,并由一个分布式异构数据库系统统一管理。分布式数据库是目前国家材料科学数据共享网的模式,也将是 MGI 计划的民用材料数据库的建设类型。而军用材料数据库由于安全性问题,采用集中式的数据库建设和管理模式。

3) 基础与专业相结合。化学元素周期表显示了所有元素,这些元素构成材料的最基本组成,而这 112 种纯元素的物理、化学性质,以及二元、三元化合物/合金的性质,构成材料基础数据库的核心内容,也构成了金属、陶瓷、高分子等各类材料的设计基础。专业材料数据库则是按照材料体系分类建设的数据库,如金属材料数据库、无机非金属材料数据库、有机高分子材料数据库、复合材料数据库等。

4) 通用与专题相结合。通用材料数据库即为数字化的材料领域的知识库,这是数据时代的要求和未来发展的必然

趋势,因此构建通用材料数据库的意义就在于构建材料知识库,成为材料知识积累的宝库。只有积累了足够的数据库,才能实现材料数据各行业用户的各类需求,以及智能制造等新方向的新需求。在《华尔街日报》提出新世纪 3 大技术变革中,智能制造的突出亮点即为材料科学,指出,兴起的材料科学正改变事物的构建,从分子结构开始进行设计、建造,优化功能,甚至创造新材料,大幅提高质量、减少浪费,指出 3D 打印技术^[10,11](又称直接数字化制造)是目前智能制造最具发展前景的方向之一,已出现 3D 打印的汽车、五层建筑、飞机座椅等产品及部件。该技术涉及的数据包括从量子尺度计算到打印过程模拟、生产过程控制整个生产流程,构建 3D 打印数据库将成为专题数据库的典型示例。专题数据库就是针对国家需求、研究热点和用户需求构建的某一材料(如高温合金、石墨烯)、技术(3D 打印)的专门的数据库。

4.2 建设思路

国家材料数据研发与服务公共平台的构建思路如图 1 所示。

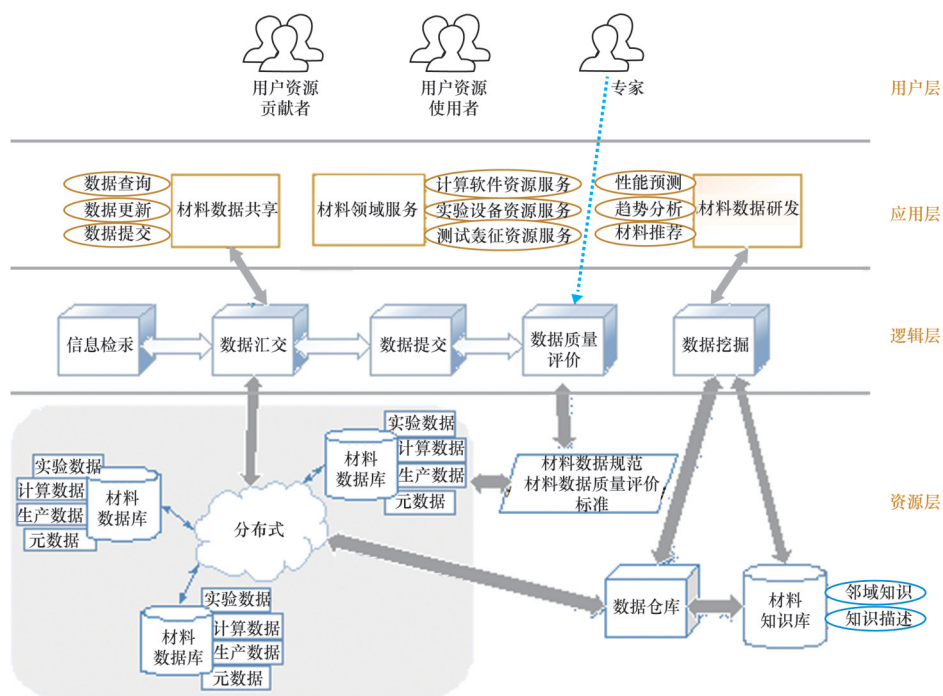


图 1 国家材料科学数据研发与服务公共平台的架构

Fig. 1 Architecture of the national public platform of materials data

材料科学数据研发与服务公共平台体系架构自上而下共分为 4 个层次,分别是用户层、应用层、逻辑层和资源层。4 个层次互相独立,却互为依托,下层为上层提供所需的数据基础和技术支撑。

用户层是 4 层体系架构的最高层,主要包含 3 类不同角色的用户,分别是资源生产者、资源使用者及专家。资源生产者是指材料科学数据的提供者,他们通过平台提供的数据共享功能,提交或更新材料科学数据。资源使用者即平台数

据资源、服务资源及研发资源的使用者。专家用户是平台的特殊角色,对平台数据资源进行质量评价。

应用层为用户层提供了平台的 3 大部分重要功能,分别是数据共享、数据研发及领域服务。其中,数据共享主要提供了材料科学数据相关的查询、提交和更新等功能支持,以方便材料数据交换;数据研发主要是基于数据挖掘技术封装而成的数据增值服务,主要实现的功能包括材料性能预测、趋势分析以及选材推荐等;领域服务集成了材料设计、计算、

实验与表征等多方面的资讯信息,为用户提供线上或线下的计算软件、实验设备以及测试表征的服务,以及最终实现材料的多层次跨尺度的材料设计,实现材料的快速设计研发。

逻辑层为应用层提供了各项功能的技术实现。为支撑应用层的数据共享、数据研发及领域服务3类平台功能,逻辑层需要实现的处理逻辑主要包括信息检索模块、数据汇交模块、数据提交模块、数据挖掘模块以及数据质量评价模块等。特别的,数据质量评价模块是依据材料数据评价标准而进行的数据质量评估操作,材料数据质量的影响因素很多,如专家意见、用户提交数据字段的完整程度,以及数据来源等等,以及用户对数据质量评价的资讯。这些处理逻辑各司其职,同时具有紧密的关联性和复杂的接口耦合。

资源层作为平台体系架构的最底层,为逻辑层提供了数据资源,具体包括物理分布逻辑统一的材料分布式数据库节点,材料数据仓库,材料知识库及标准规范。其中,分布式数据库节点中存储了材料科学数据,分为实验数据、计算数据和生产数据、元数据信息,这些数据通过逻辑层的数据汇交满足用户对材料科学数据的跨库检索需要;材料数据仓库是为满足数据挖掘的需要按照既定的模式对材料数据进行的整合和集成,数据挖掘的数据直接从材料数据仓库中获取;材料知识库是对材料学科知识的建模与描述,如材料科学数据的本体,记录并维护了材料学科概念及联系,同数据仓库一起为数据挖掘提供数据源和基础关联规则;材料数据规范规定了材料科学数据汇交的模式;质量评价标准为数据质量评价功能提供算法依据。

4.3 国家材料数据的研发与服务公共平台的建设内容

材料领域包含的内容广,覆盖面大,针对中国材料数据的发展现状,有目标有针对性地开展以下几个方面的工作。

4.3.1 加大材料数据收集和整合的力度,扩大中国材料数据资源,建设国家材料数据研发与服务公共平台的数据仓库

充分发挥材料科学数据共享网这一国家级平台的基础,加强材料资源的数字化建设,扩大数据收集整理与整合的范围,提高数据整合的精度,并结合数据时代对大数据的要求,确保材料数据整合的标准化,以及系统性、准确性、完整性、一致性、及时性和重用性等方面的应用需求,同时确保数据库的更新维护和科学管理。

4.3.1.1 适用于MGI的数据组织体系的研究确立

目前国际上除中国的国家材料科学数据共享网及日本在2014年开始建设的国家材料数据中心外,无其他国家级材料数据库。目前存在的材料数据分类体系主要包括按照材料的尺度以及材料分类体系这两个维度进行划分。两种分类方法各有特点,而材料领域的各个学科差异很大,因此,材料数据研发与服务公共平台的数据分类方法无法复制现有体系,需要通过组织材料领域各学科的专家、高校、科研院所的科研人员和潜在用户的研讨、论证,确立适用于MGI新材料创新与应用的数据分类体系,同时吸纳更多建设单位和个人的参与。

4.3.1.2 材料数据的调研与入库评估

国内目前除一些国家级材料数据库外,还有一定量的数据资源,尤其是近几年国家对科研的大量投入,产生的数据量激增,材料数据研发与服务公共平台的民用数据库应该对全部公众开放数据上传窗口,让所有拥有数据资源的人都参与到材料数据的整合中。因此,需要对国内数据资源进行全面深入的调研,并对所有数据资源系统进行入库评估。

数据调研的内容包括:1) 所属材料分类及其类型;2) 生产方式;3) 完整性和一致性;4) 更新状态;5) 与数据源的一致性;6) 可获取性。

入库评估的主要内容包括:1) 数据来源状况与数据的完整性;2) 数据的可获取性;3) 数据的成熟度。

完整性和一致性是数据挖掘与分析对数据的基本要求。在早期收集的实验数据中,由于当时有限的研究条件,大量实验数据只有最终性能而无制备工艺,又因为成分有一定波动范围,导致不同源的同种材料的性能存在明显差异,各国国家标准的数据也不同,因此数据应具有明确的来源,并与数据属性要求保持一致。生产方式(计算、实验、表征等)及其详细的条件参数清晰,如计算数据的条件参数包括计算或模拟软件的名称、版本号、建立的模型、模拟计算的初始条件、边界条件和各类假设等,实验数据应包括材料的成分、制备加工方法、工艺参数、微观组织和国家标准号等,表征数据应包括表征方法、检测设备型号、检测单位、样品详细信息等。数据的成熟性与数据的质量有密切关系。在数据的评估中重点关注两类数据,一类是前沿热门方向的材料数据,标准化程度低,数据大多为实验数据,但如果满足一定的形式要求,可以入库,另一类为成熟材料的早期数据,对于缺少的信息需根据情况进行处理。

4.3.1.3 面向MGI需求的材料数据完整性的分析与补充建设

现有材料数据的获得往往是研究者关注和研究的相关方面积累的数据,而研究者无暇研究的其他方面可能存在潜在的突出性能,实际生产和生活中往往有很多“无心插柳成荫”的事例,在新药的使用中就有用于治疗某一种病的药物被发现用于另一种病的治疗上效果更佳,在材料领域也不乏这样的例子,如高速钢以其在高速切削加工时仍保持良好的红硬性而得名,但其做为动态冲击状态下的冷作模具材料显示出非常优越的性能和高寿命。因此,开发现有材料的未知潜能,将现有材料的多方面的性能等数据均发掘出来,作为数据库的重要补充部分,将为支持高通量计算和模拟数据的验证,现有材料的新性能发现和性能提升,尤其是新性能的发现,提供有力的支撑。结合MGI和国家经济高速发展和国防建设的迫切需求,先以能源材料、高温超导、高温合金、磁性材料、催化材料、结构材料等作为先导,逐步全面开展现有材料的数据的手收集和积累。

4.3.1.4 数据整合和计算模拟软件、模型的嵌入

材料数据的研发与服务平台,数据的整合是核心,但该平台还需要满足材料研发的更进一步的需求,尤其是多层次

跨尺度的材料计算及高通量实验表征的要求。因此,在整合数据的同时,将材料计算相关的计算模拟软件、模型以及经验公式等同时整合入库,兼顾其与数据的关联,发现不同尺度的计算及验证用关键数据,实现不同尺度、不同层次间的关联,将材料研究中碎片式的知识与数据有机联系成为一个整体,最终实现材料从设计到应用的过程的高效低成本的控制,形成材料数据的知识库和工具库。

4.3.1.5 数据更新

数据库的更新是数据库可持续发展的重要保证。在中国早期开发的材料数据库中,多数因长期缺乏有效更新,逐步消亡。而一个数据库能否保持数据的准确性和及时性,很大程度上依赖于数据库的更新功能的强弱与实时。数据库的更新通常包括插入、删除、修改三种操作,材料数据库更多的是插入和修改,对于早期用相对落后的技术和制备手段得到的数据,对现在的研究仍有较高的研究价值,予以保留。

4.3.1.6 材料数据质量评价方法及其应用研究、材料数据质量评价体系的建立

数据质量是衡量数据库的质量和数据库应用的核心内容和根本保证。从技术和制度两方面,搭建国家材料数据研发与服务公共平台的标准质量评价体系,包括机器检查、专家评估与同行评议等^[12]。目前国际上尚无材料数据质量评价方法的标准,甚至连完整的材料数据质量评价的方法论都是空白。因此,在建设国家材料数据公共平台时,需要学习其他领域在数据质量评价上的研究经验,并结合材料数据的特点,开展材料数据质量评价方法及其构建与应用的研究。目前常用的数据质量评价方法包括:演绎推算、内部验证、与更高精度的独立原始资料对比、独立抽样检查、多边形叠加检查、有效值检查等。对于材料数据质量评估,梳理出材料数据的质量元素是质量评价的第一步。

然而,仅从技术层面上解决材料数据质量评估是不够的,还需要研究建立并形成材料数据质量评估标准,从制度层面上,搭建中国材料数据质量评价标准体系,形成材料数据质量评价的标准化、制度化、常态化的态势。对所有库内数据进行质量评估,排除错误数据对计算和模拟结果的验证的干扰,避免将材料创新引入歧途,并为今后国际间的材料数据库的合作做好充分的准备。

材料数据质量评价方法及体系,是对材料数据进行标准化的检测和评估,对材料数据质量评价方法及体系研究,正如很多领域所做的大量研究一样^[13-16],将成为今后材料数据科学的一个研究方向。

4.3.1.7 建设民用材料数据库和军用材料数据库,形成中国自主材料数据资源系统

国家级民用材料数据库的建设将采用与国家材料科学数据共享网相同的分布式数据库的模式,形成物理上分布、逻辑上统一的材料数据仓库,各数据拥有单位和个人可以建设自己的小型数据库并存在本地服务器,通过分布式数据库的建设整合在一起,通过材料数据对象唯一标识符(DOI)体系对各自所有的数据资源命名而获得清晰的知识产权归

属,通过标准化的接口技术实现跨库查询和检索。对于用户来讲,这个数据仓库是一个体系和逻辑完整的材料数据库,与集中式数据库无任何差异。

军用材料数据库的整合建设采用与民用材料数据库相同的材料数据体系,遵守相同的材料数据库相关标准,以便在应用过程中可以将民用材料数据库作为补充资源,理解无门槛,调用无障碍。网络安全(cyber security)是全球关注的热点问题,目前最安全的数据库存储方式仍是线下的。因此,兼顾数据安全和材料数据的全面性要求,军用材料数据库采用集中式数据库形式。

为保证材料数据在数据科学的研究中具有互操作性和数据应有价值的体现,民用材料与军用材料数据库所整合的数据均采用统一的标准化的材料数据提交格式。

4.3.1.8 专题材料数据库的建设

民用材料数据库和军用材料数据库的建设,是一项长期的工作,是中国材料领域资源的数字化储备。而面对当前激烈的国际竞争,应对经济发展放缓,满足现阶段材料高效创新和国家科技和经济创新与快速发展需求,国家科技重大专项和国防建设急需、支撑高新技术领域发展和解决经济社会重大紧迫问题,需要集中力量开发建设一批专题数据库。

主要建设以下几个专题数据库,包括以下几方面。

1) 重点新材料的不同尺度计算和验证用基础数据库。

由于在材料设计计算时,功能材料与结构材料在不同尺度的材料设计及其对数据的需求上表现出差异显著,故将材料分为功能材料与结构材料两类讨论。两者对数据的需求表现如图2所示。

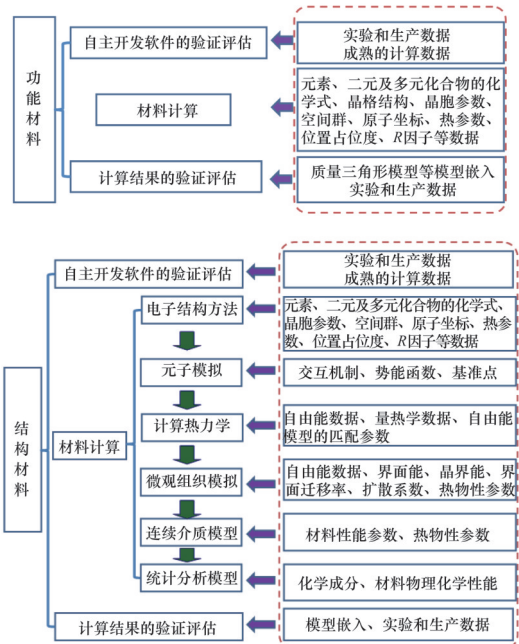


图2 功能材料与结构材料在软件开发、计算、计算结果验证等方面对材料数据的需求

Fig. 2 Requirement of functional and structural materials on materials data in aspects of software development, computation and testing

2) 高性能结构材料专题数据库。包括基础数据、成分数据、微观组织数据、制备工艺过程数据、性能等数据,以及计算模型、软件及性能预测功能。

3) 清洁能源材料专题数据库。目前主要针对储能材料、光伏光热材料、LED材料等材料,包括材料设计、合成、制备、测试和分析、产品性能等多方面的基础、过程、微观组织、性能数据等。

4) 催化材料专题数据库。包括化工催化材料、光催化材料等的材料计算数据、物相结构、制备与性能数据等。

5) 钢铁材料专题数据库。包括涉及多尺度、多层次耦合下的成分及物相设计计算与制备的基础数据、过程数据、性能数据等,以及多尺度信息与数据的支撑与传递系统,同时建设汽车用钢、海洋工程材料、能源材料专题数据库及钢铁材料生产加工过程专题数据库等。

6) 航空材料专题数据库。包括航空航天用材料的基础数据、成分、微观组织、制备和生产过程数据、性能数据,以及材料设计用多种软件和模型的嵌入等。

7) 3D打印专题数据库。包括适用于3D打印的金属等材料的化学成分、打印设备、打印工艺的宏观和微观表征数据等。

8) 应急预案专题数据库。在对天宫一号的关键部件在短时间内提供准确的服役和寿命评估的事例证明,材料数据的积累对应国家重大关键项目的突发事件的材料问题是极其重要的。应急预案专题数据库包括空天飞行器材料专题数据库、核材料专题数据库等,同时将根据国家发展的需求随时增加新类型材料的数据库的建设。

建成的国家材料数据研发与服务公共平台,可以与材料计算平台和测试表征平台构成相互关联、紧密协作、互高效循环运转的从新材料设计到创新产品应用的协同创新体系。数据库平台为高通量计算平台和分析测试平台提供全面、系统、科学的数据体系,是MGI中最基础的支撑部分。

4.3.2 国家材料数据研发与服务公共平台的服务功能的建设

为满足材料基因组计划的需求,不断完善材料数据在MGI应用中与材料计算、实验与表征的协同,以应用实践作为检验材料数据工作的标准。开展基础设施建设,包括数据标准、软件标准、标准化的数据管理体系、数据安全等方面的研究。在整合数据入库的同时,兼顾软件与模型在平台中的嵌入及与数据的关联,发现不同尺度的计算及验证用关键数据,实现不同尺度、不同层次间的关联,将材料研究中碎片式的知识与数据有机联系成为一个整体,最终实现材料从设计到应用过程的高效低成本的控制^[7],并以材料数据为管理部门、材料领域各学科、材料生产等提供服务。

4.3.2.1 材料数据库研发与服务公共服务平台的基础结构建设

1) 标准体系建设。标准化建设是大数据的要求,同时更是分布式异构材料数据库的基本要求,是保证今后数据库的交互协同、数据的共享、分析与挖掘的基础和前提。需要建

立数据标准、软件标准和标准化的数据管理体系,其中数据标准包括材料元数据标准、材料数据模型标准、材料数据更新标准等,软件标准包括软件接口标准、软件集成标准及软件测试标准等,标准化的数据管理体系包括互操作标准、数据传输、发布、存取控制标准等,并最终形成相关国家标准。

2) 运行管理机制的建设。运行管理机制建设,是平台可持续发展的保障。运行管理机制包括材料数据共享机制、材料数据更新机制、材料数据专家委员会、材料数据专家咨询服务系统、以及人才队伍建设和人才培养机制等方面。

这里尤为重要的是材料数据共享、引用和出版机制。开放共享是自由整合和数据应用的前提,但材料数据的开放共享一直是困扰材料科学数据共享网建设的重要因素,但数据的开放共享不仅仅是材料领域的问题,而是全球各个领域的问题,美国、德国、英国、欧盟等发达国家从上世纪末和本世纪初就纷纷出台各类数据共享的法律法规,但执行的结果则差强人意^[8]。因此,急需建立一个良性发展的材料数据共享生态系统。目前较为可行之举是采取双赢的举措,使数据共享者能够从其共享行为获益。DOI体系的建立,解决了材料数据的知识产权所属问题,为推进材料数据共享迈出了第一步。但关于材料数据共享和知识产权保护,材料数据的引用与出版,还需要在法律上开展更为深入的探讨。

材料数据专家委员会对平台的发展方向、建设模式、运行服务机制、工作重点等方面进行系统研究,对平台的总体运行提出决策意见,材料数据专家咨询服务系统为数据库的各类用户提供远程咨询服务,对材料创新活动中的各类问题进行快速相应,提供应急解决预案,并协助组织多领域专家咨询会,对疑难问题进行综合会诊。

3) 软件运行服务系统建设。采用适应数据挖掘与分析的数据库模式,注重交互性,不仅实现用户与网络服务器间的交互,而且同一网站的不同用户间、不同网站间的交互。建立用户开放窗口,作为平台建设中的创新活动,建立内容更丰富、联系性更强、工具性更强的用户服务平台。1) 数据的存储管理。目前在对数据的存储管理上主要有并行数据库、NoSQL数据库和NewSQL数据库等3类。NewSQL是对各种新的可扩展/高性能数据库的简称,不仅具有NoSQL对海量数据的存储管理能力,还保持了传统数据库支持ACID和SQL等特性,使得关系型数据库系统在处理海量数据时仍能获得很好的性能,既提供SQL数据库的质量保证,也能提供NoSQL数据库的可扩展性。材料数据库研发与服务公共服务平台,将结合材料数据的分布、异构、关系复杂等特点,研究NewSQL数据库模式的应用;2) 材料数据安全管理体系的建设。除考虑数据库的开放度和用户权限的设定外,需要在从数据库的支持软件和服务硬件上把关,使数据库管理与服务安全、高效地进行;3) 建设交互协同平台软件。开发数据透明交互软件及其标准化设计,解决数据库间交互的问题,确保中心数据库与分布数据库的联合交互、外部数据库的元数据的识别与交换,以及多层次、跨尺度的材料计算作业间数据的分布式调用^[8],使材料的智能化制造成为可能。

4) 建设材料科学数据服务云。(1) 数据的分析处理。Map Reduce 模式和 Hadoop 框架的并行处理模式可以应对 MGI 活动中的数据分析处理。采用 Map Reduce 模式, 可避免大量数据的传输, 提高数据处理效率, 同时具有良好的容错性, 一部分节点的 down 机对集群的正常工作不会造成影响, 借助 Hadoop 这一开源分布式并行编程框架编写程序, 运行于计算机集群上, 从而实现对数据库平台的海量材料数据的处理。(2) 材料数据云服务及其运行模式建设。利用云计算的思想, 通过分布式计算, 将服务任务分解多个计算机或服务器上来进行, 可以大大提高解决问题的能力, 缩短服务时间。在 MGI 体系下, 云数据服务的模式尚待不断研究、探索和实践。

4.3.2.2 开发满足 MGI 需求的高效查询检索与定制服务

大数据的高效检索是数据库平台应率先实现的功能之一, 高效检索软件的开发与应用, 对数据库应用平台的服务至关重要。搜索引擎的出现在一定程度上解决了信息筛选问题, 但当用户无法准确描述自己的需求时, 搜索引擎的筛选效果将大打折扣。推送技术的基础思想在于将浏览器主动查询信息改为服务器主动发送信息, 是内容提供者向用户传递消息的一种服务。定制服务包括定制产品以及相应的服务双重内容, 专题数据库的建设是大型定制产品, 而移动互联思维下的定制服务更包括了针对所有用户提供的定制产品及服务, 材料数据推送服务将成为目前材料数据库应用平台的主要服务模式。

4.3.3 基于材料科学数据的材料信息学和材料数据科学的研究与应用

从 20 世纪 50 年代信息学诞生, 到组合芯片方法^[9]在材料快速筛选中的应用, 材料信息学的诞生, 到数据科学是知识发现的第 4 种手段^[17,19], 在这样数据时代和 MGI 的双重背景下, 数据科学在材料领域应用而形成材料数据科学, 必将与脑数据学、行为数据学、生物数据学、气象数据学、金融数据学、地理数据学等一样, 通过认识数据的各种类型、状态、属性及变化形式和变化规律, 揭示材料的现象和规律, 最终成为材料领域的一个全新的学科。

对比图 3 中 IT 时代与传统产业的发展模式, 不难看出, 两者具有明显的相似之处, 可以预见, 今后材料数据科学的

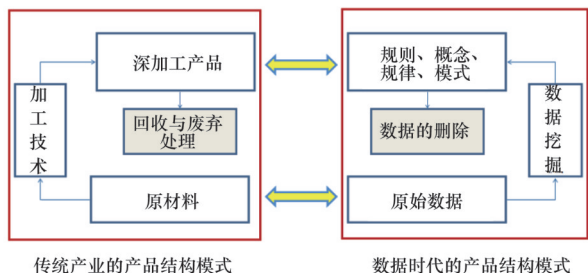


图3 数据时代与传统产业的产品结构的类比

Fig. 3 Analogy of product structures between the data era and conventional industries

发展将同样遵从科学和社会的发展规律。

材料数据科学, 作为材料科学领域的新兴学科, 应包括以下几方面的研究内容。1) 材料数据的基本特征及其评价体系研究与建立, 以及材料大数据的准确性、完整性、及时性、一致性要求的理论、技术与手段研究; 2) 材料信息学。研究材料数据挖掘、分析与可视化等技术的开发, 主要包括数据挖掘和可视化。数据挖掘表现为采用分类、聚类、回归分析、特征分析、关联规则、变化和偏差分析等数据挖掘方法, 从大量材料数据中搜索和提取隐含的、未知的、有潜在价值的信息或模式, 从而发现和获得材料的本征参数及其复杂关系。可视化表现为利用计算机图形学等直观的方式表示计算模拟结果中的大量非数字型数据信息, 创建视觉图像, 帮助用户理解关系错综复杂而数字规模庞大的概念和结果; 3) 材料数据学。材料数据学是将数据学^[20]应用于材料数据, 开展多层次跨尺度计算的全自动操作研究^[21,22], 开展满足材料智能制造的全流程的材料数据及其支持系统的研究与开发建设; 针对高通量并行计算、组合实验方法等产生的海量流数据, 开展数据的系统存储与高效利用。

4.3.4 积极开展国际合作与交流, 使中国材料数据的影响从亚洲扩大到全球

亚洲材料数据会议 (AMDS) 是两年一次的材料数据学术会议, 2010 年中国承办了第二届亚洲材料数据库会议, 扩大了中国在材料数据库建设上在亚洲和世界上的影响, 亚洲材料数据会议系列会议上, 材料数据工作者与国内外同行共同分析了材料数据库建设的突出问题和发展趋势, 并充分交流了材料数据库建设的经验与成果。同时, 北京科技大学发起成立了亚洲材料数据库委员会 (AMDC), 曲选辉任首任主席。AMDS 和 AMDC 使中国的材料数据在亚洲形成了主导地位。近年来, 国际数据委员会 (CODATA) 的活动也吸引了材料数据工作者的参与, 同其他领域学习了很多先进的、系统的理念, 能够以全球的视角看待材料数据的问题, 以及解决思路和发展方向。CODATA 中国委员会每年组织并承办“科学数据大会”, 材料数据及其应用是该会议的分会之一。CODATA 的各项活动, 将中国材料数据的发展提升了一大步, 使材料数据与材料大数据、材料基因组工程也逐渐进入其他领域研究者的视野。

5 材料数据发展存在的主要问题

1) 知识产权与数据共享问题。这一问题目前已经成为全球各研究领域中的数据共享的共性问题^[23], 对数据的知识产权属性尚无法律界定, 因此即使数据的获得来自国家项目的资助, 数据生产者也不愿将数据无偿共享, 这样直接影响并大大降低了材料数据库的价值。目前较为可行之举是采取双赢的举措, 使数据共享者能够从其共享行为获益。材料数据对象唯一标识符 (DOI) 体系的建立, 解决了材料数据的知识产权所属问题, 为推进材料数据共享迈出了第一步。法律、软件开发。

2) 材料数据相关工作的评价。目前材料数据的整理整合工作对于科研工作者来讲基本上是奉献性的,所做工作因难于发表文章等原因,无法计入单位的考核体系,从事材料数据库建设的研究者往往被边缘化,严重打击了对材料数据的工作热情,年轻人更是不愿触及相关工作。而材料领域涉及的材料种类多,数据覆盖面广、信息量大,而且数据资源是动态激增的,因此材料数据库的建设,是一项长期的科研活动,成立专门的部门并且每年有固定的经费支持,是国外一些单位的做法。

6 结论

材料是文明发展的基石,更是国家经济发展和安全的基础和根本保障。材料数据是材料基因组计划的3大工具之一,同时材料数据做为数据时代中材料科学资源的保存方式,其发展具有前所未有的重要意义。然而中国在材料数据的收集整理、数据库建设及其应用上与国外发达国家的差距,以及在材料计算用数据库上一片空白的现状亟待改变。依托国家材料科学数据共享网等材料数据资源,加大材料数据整合力度,建设材料数据服务平台,开展材料数据的分析与挖掘,开展材料信息学与材料数据科学的研究,并通过专题数据库的建设,为社会各类用户提供定制性服务,最终实现将材料从发现到应用的进程缩短一倍以上,成本降低一半以上的目标。

材料数据相对与其他很多科学领域起步较晚,但在数据时代大数据风起云涌之时,材料基因组计划的提出与实践为材料数据的发展提供了强大的动力,材料数据库的建设和材料数据科学的发展,有望在今后迅猛发展,成为材料领域的重要学科和方向,并为进一步推动材料数据为加快材料创新进程、实现材料智能制造而显示出巨大的价值。

致谢 感谢中国工程院课题咨询研究项目和美国肯纳公司对本课题的帮助。

参考文献(References)

- [1] Hey T. The fourth paradigm: Data-intensive scientific discovery[M]. Redmond, WA: Microsoft Research, 2009: 109-130.
- [2] 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代: 生活、工作与思维的大变革[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013: 2-18.
- [3] Mayer-Schönberger V, Cukier K. Big data: A revolution that will transform how we live, work and think[M]. Sheng Yangyan, Zhou Tao, trans. Hangzhou: Zhejiang People's Publishing House, 2013: 2-18.
- [4] Xiang X D, Sun X, Briceno G, et al. A combinatorial approach to materials discovery[J]. Science, 1995, 268(5218): 1738-1740.
- [5] Liu Z K, Chen L Q, Raghavan P, et al. An integrated framework for multi-scale materials simulation and design[J]. Journal of Computer-Aided Materials Design, 2004, 11(2-3): 183-199.
- [6] Olsen G B. Pathways of discovery designing a new material world[J]. Science, 2000, 228(12): 933-998.
- [7] Yang Y, Lin T, Weng X L, et al. Data flow modeling, data mining and QSAR in high-throughput discovery of functional nanomaterials[J]. Computers & Chemical Engineering, 2011, 35(4): 671-678.
- [8] 王盼盼, 杨迪. 阿里巴巴拟以大数据推动智能制造[EB/OL]. 2014-11-01. http://news.xinhuanet.com/fortune/2014-11/01/c_1113074082.htm.
- [9] Wang Panan, Yang Di. Alibaba intends to promote smart manufacturing with big data [EB/OL]. 2014-11-01. http://news.xinhuanet.com/fortune/2014-11/01/c_1113074082.htm.
- [10] 师昌绪. 材料大辞典[M]. 北京: 化学工业出版社, 1994.
- [11] Shi Changxu. Materials comprehensive dictionary[M]. Beijing: Chemical Industry Press, 1994.
- [12] 36 大数据. 中国大数据行业面临的五大挑战以及应对策略[EB/OL]. 2014.11.21. <http://www.36dsj.com/archives/17137>.
- [13] 36 big data. The strategies for the five challenges the big data industry are facing[EB/OL]. 2014.11.21. <http://www.36dsj.com/archives/17137>.
- [14] Kietzmann J, Pitt L, Berthon P. Disruptions, decisions, and destinations: Enter the age of 3-D printing and additive manufacturing[J]. Business Horizons, 2014. <http://dx.doi.org/10.1016/j.bushor.2014.11.005>.
- [15] Chou D T, Wells D, Hong D, et al. Novel processing of iron-manganese alloy-based biomaterials by inkjet 3-D printing[J]. Acta Biomaterialia, 2013, 9(10): 8593-8603.
- [16] 胡良霖. 科学数据资源的质量控制和评估[J]. E-Science, 2009(1): 50-55.
- [17] Hu Lianglin. The quality control and evaluation of scientific data resources[J]. E-Science, 2009(1): 50-55.
- [18] Sargent P. Data quality in materials information systems[J]. Computer-Aided Design, 1992, 24(9): 477-490.
- [19] Bhagwat S A, Patterson K Y, Holden J M. Validation study of the USDA's data quality evaluation system[J]. Journal of Food Composition and Analysis, 2009, 22(5): 366-372.
- [20] 罗芳. 土地利用数据综合结果的质量评价[D]. 武汉: 武汉大学, 2013.
- [21] Luo Fang. Quality evaluation for land use data generalization[D]. Wuhan: Wuhan University, 2013.
- [22] 苗海芳. 多元交通信息数据质量评价和控制方法研究[D]. 长春: 吉林大学, 2014.
- [23] Miao Haifang. Research on evaluation and control methods on multi-source traffic information data[D]. Changchun: Jilin University, 2014.
- [24] Curtarolo S, Morgan D, Persson K, et al. Predicting crystal structures with data mining of quantum calculations[J]. Physical Review Letters, 2003, 91(13): 135503.
- [25] 黎建辉. 科学数据共享的新机制与新趋势[C]//第507次香山会议. 北京, 2014-10-22.
- [26] Li Jianhui. New mechanism and trend of scientific data sharing[C]// No. 507 Xiangshang Science Conference, Beijing, 2014-10-22.
- [27] Morgan D, Ceder G, Curtarolo S. High-throughput and data mining with ab initio methods[J]. Measurement Science and Technology, 2005, 16(1): 296.
- [28] 朱扬勇, 熊赞. 数据学[M]. 上海: 复旦大学出版社, 2009.
- [29] Zhu Yangyong, Xiong Yun. Dataology[M]. Shanghai: Fudan University Press, 2009.
- [30] National Research Council (US). Committee on Integrated Computational Materials Engineering. Integrated computational materials engineering: A transformational discipline for improved competitiveness and national security[M]. Washington, DC: National Academies Press, 2008.
- [31] Robinson L. New TMS study tackles the challenge of integrating materials simulations across length scales[J]. JOM, 2014, 66(8): 1356-1359.
- [32] Sally T, David L M, Amanda B, et al. Sharing data in materials science [J]. Nature, 2013, 503: 463-464.

(编辑 田恬)