



Research
Materials Genome Engineering—Perspective

数据标识编码——连接材料基因组工程数据库与可传承集成智能制造的桥梁

王毅^a, 李佩璇^a, 林德焯^b, 唐斌^a, 王军^a, 管全梅^c, 叶谦^a, 代海星^d, 高军^c, 范晓丽^a, 寇宏超^a, 宋海峰^b, 周峰^{a,*}, 马纪军^{c,*}, 刘梓葵^e, 李金山^{a,*}, 刘维民^a

^a State Key Laboratory of Solidification Processing, Northwestern Polytechnical University, Xi'an 710072, China

^b CAEP Software Center for High Performance Numerical Simulation, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China

^c CRRC Tangshan Co., Ltd., Tangshan 063035, China

^d Beijing Star Travel Space Technology. Co. Ltd., Beijing 100013, China

^e Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:

Received 14 April 2019

Revised 23 April 2019

Accepted 16 October 2019

Available online 7 May 2020

关键词

数据标识符

数据库

数字孪生

集成计算材料工程

摘要

数据标识符 (DID) 是所有类型数据库中必不可少的标签, 尤其是与集成计算材料工程 (ICME)、可传承集成智能制造 (I³M) 和工业物联网有关的数据库。依据全球多个官方文件的展望蓝图的指引和先进材料的快速发展, 需要进行更多的研究工作来建立材料信息学的相关数据标准。本文提出了由一系列构建单元 (意义段) 组成的 DID 统一格式, 该格式与国际和国家标准 (如 ISO/IEC 29168-1:2000、GB/T 27766-2011、GA/T 543.2-2011、GM/T 0006-2012、GJB 7365-2011、SL 325-2014、SL 607-2018、WS 363.2-2011 和 QX/T 39-2005) 中标识符的经典格式一致。每个构建单元均由大写字母和数字组成, 不包含特殊符号。此外, 依据 ISO/IEC 10646 国际标准中统一编码标识符单元的格式, 对每个构建单元的总长度不做限制规定。基于这些规则, 本研究提出的 DID 具有灵活性, 便于在各种云平台之间进行扩展和共享。相应地, 智能手机或特定机器可以构造和精确识读这些常见的二维 (2D) 码, 包括汉信码 (Hanxin Code)、龙贝 (Lots Perception Matrix, LP) 码、快速反应 (Quick Response, QR) 码、网格矩阵 (Grid Matrix, GM) 码和数据矩阵 (Data Matrix, DM) 码。通过这些二维码作为一组与云平台相连的数据指纹, 人们可以自动跟踪成分-工艺-结构-性能-服役全流程中的进度和更新, 为加速先进材料的发现和制造以及提高研究产出、效能和协作铺平道路。

© 2020 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

随着人类进入集成智能制造时代[1-3], 集成计算材料工程 (ICME) 的数字孪生设计范式对于加快新型先进材料的发现和应用至关重要。大量官方文件和计划已概述了计算材料工程的独特挑战和机遇以及未来可传承

的集成智能制造 (I³M) 的战略蓝图, 如美国的材料基因组计划 (MGI) [4]、中国的材料基因工程 (MGE) 和人-信息-物理系统 (HCPS) [1]、德国的工业4.0 [2]、韩国的工业创新3.0 [5]。材料创新赋予了解决问题的新技术能力并促进社会的重大进步[6,7]。基于这些先进材料的设计范式, 材料发现和工程创新为技术进步开辟了新的

* Corresponding author.

E-mail address: zhouf@licp.cas.cn (F. Zhou), Sjc-majijun@tangche.com (J. Ma), ljsh@nwpu.edu.cn (J. Li).

2095-8099/© 2020 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

英文原文: Engineering 2020, 6(6): 612-620

引用本文: William Yi Wang, Peixuan Li, Deye Lin, Bin Tang, Jun Wang, Quanmei Guan, Qian Ye, Haixing Dai, Jun Gao, Xiaoli Fan, Hongchao Kou, Haifeng Song, Feng Zhou, Jijun Ma, Zi-Kui Liu, Jinshan Li, Weimin Liu. DID Code: A Bridge Connecting the Materials Genome Engineering Database with Inheritable Integrated Intelligent Manufacturing. *Engineering*, <https://doi.org/10.1016/j.eng.2020.05.001>

领域[6], 包括大数据[8]、数据挖掘[6]、机器学习[6,9]、人工智能、云计算、金属材料本体论和知识图谱。与大数据和机器学习相关的技术有助于将理论预测与微观自由度联系起来, 从而加速区域材料的设计和合成[9]。如集成实验衍射数据、对称性统计反馈、基于密度泛函理论(DFT)的优化算法, 第一原理辅助结构解析已成为一种新型的自动预测晶体结构的混合方法[10]。DFT计算和相图计算(CALPHAD)的集成被认为是材料基因和材料设计的主要部分[11-13], 并已建成一个强大的数据驱动的ICME材料开发框架[5,14], 该框架强调基于相物理性质的数据库。众所周知, ICME是通过计算工具获取、工程产品使役性能分析和制造加工仿真的材料相关信息的集成[5,15]。随着计算材料科学和计算能力的迅速发展, 多物理场模拟推动了热力学、动力学、结构、缺陷和性质在多尺度上的预测, 极大地加速了材料数据库或存储库的发展[5,13,16]。此外, 通过集成多尺度模拟, 设计策略可以跨越从电子到相甚至到产品的范围[12,14,17,18], 并且可以有效地确立筛选目标候选材料的原则和判据[5,13,19-21]。

从集成智能制造/工程的角度来看, 数据驱动的ICME支持数字孪生类型的设计/制造范式, 这突出了材料信息学的重要作用。尽管第三次工业革命(所谓的数字革命)仍在发展中, 但是第四次工业革命的重点是基础设施和数字科技, 并将非线性特征、数字技术和学科整合到虚拟材料和物理系统领域[3]。在智能设计时代, 通过计算可以提前确定新的目标、候选对象和技术方案, 从而有利于企业以较低的成本和较短的时间提供基于所获得的知识 and 模型的营销策略, 以便在全球竞争中生存。为了对目标设计、优化、计划和解决方案做出有效的决策, 解决公认的数据“4V”至关重要, 包括体量、速度、多样性和准确性[3]。据报道, 北美的大型制造商已经花费了近7万亿美元来升级以前带有传感器的设备, 使设备系统能够通过物联网(IoT)相互通信[3]。有效地利用这些运营数据进行商业决策是一个巨大的挑战或机遇, 然而目前仅有约1%的商业决策是这样做的[3]。因此, 不但应该关注促进数据集自动化收集、整理和分发的工具, 还应该定义并遵循标准化的数据和元数据格式[6]。例如, 可查找性、可访问性、互操作性和可重用性(FAIR)这四个基本原则被提出, 用于传统意义上的数据、算法、工具和工作流的指导原则。在收集和分享数据时应考虑所谓的FAIR指导原则[6,22]。

我们最近的短评文章[5]讨论了在当前数据驱动的

ICME的框架中, 数据库、工具包、平台、原理、基准和标准的主导作用。由于数据驱动的ICME的短期目标已基本完成, 长期目标正在进行当中, 其中包括培养下一代ICME生力军并建立基于网络的ICME基础架构, 以提高其全球工业竞争力和国家安全。据了解, 构建材料信息学的相关数字标准对于发展数据挖掘、深度/机器学习和人工智能, 验证有价值的数据和加速材料创新、发现和设计都是很有必要的。这还需要加强与工业伙伴的互动与合作[5,6,23]。为了解决技术从实验室转化为实际应用时出现的问题, 美国国家科学技术研究院(NIST)提出了关于性能和互操作性的标准, 以加速创新并将应用新型智能制造技术的风险降至最低[3,5]。此外, 最近提出的中国材料与试验学会(CSTM)标准草案《材料基因组工程数据通则》是对MGE数据内容进行标准化的首次尝试, 将对材料科学向数据驱动的科学体制的转变产生深远影响。

本研究提出了一种由一系列构建单元(意义段)组成的通用数据标识符(DID), 作为CSTM系统地建立MGE数据标准的一部分, 该标识符与国际和国家标准中使用的经典标识符格式保持一致。该DID具有灵活性, 便于在各种云平台之间进行扩展和共享。相应地, 典型的二维码可以被智能手机或特定机器生成并精确识别和解码。通过将这些二维码作为与云平台相连接的一组数据的指纹, 成分-工艺-结构-性能-服役(CPSPP)工作流程中的进展和更新将被自动跟踪, 从而为促进先进材料的开发及提高研究产出、成果和合作奠定基础。

2. 数字孪生智能制造时代

2.1. 设计时代的数字孪生制造

数字孪生是物理实体的计算机化伙伴, 是ICME时代的新型设计范式, 被用于飞机、火车和发动机的设计制造[3,5]。如图1所示[5], 先进材料的典型发现、设计、创新和制造链包括成分、工艺、微观结构、属性和性能。材料科学中的CPSPP关系或工作流过程在指导材料的发现和制造方面极为重要, 并且需要先进的技术, 包括高通量计算、增材制造、人工智能、数据挖掘、机器/深度学习等[2,5,24,25]。考虑到自下而上设计和自上而下工程[23], 实验环节和理论环节之间的数字孪生特征以不同的背景色突出显示。MGI强调实验工具、计算工具和数据库以及它们之间交互作用的主导地位, 而HCPS和MGE则强调HCPS的交互作用, 这预示了上述先进

技术的未来的应用。德国工业4.0强调通过信息-物理系统（CPS）来实现智能制造[8]。美国对未来智能制造业提出了两个目标：①展望并促进美国工业的前瞻性和领导力；②有助于生力军的发展[3]。这些目标也被中国的HCPS和MGE考虑在内。机器学习算法可以通过识别主要数据和有价值的关系来加快对材料和基础科学研究的本质理解，从而加强人-物理系统和人-信息系统的解释，并形成科学知识和模型[26]。在完成模型建立并根据材料参数预测性能之后，进一步分析训练模型中的梯度，可以确定主要的和有价值的数据关系，而这些关系无法通过人工检查或传统的统计分析方法建立[26]。在数字孪生智能制造时代（也称“设计时代”），通过将自下而上的理论设计与自上而下的实验路线相结合，利

用数字孪生智能制造技术，有望以较低的成本，更高效地设计和发现新的先进材料。这将加速MGI基础设施的发展、加强HCPS中的交互作用，并支持利用最新的先进技术同步构筑未来。

2.2. 生态系统中的设计

数据和数据基础结构是保证MGI/MGE、ICME和I³M成功取得进展的三个基础之一。所有这些术语都是基于信息学。信息学是一个广义术语，涵盖了数据驱动的设计阶段，如储仓、可视化以及统计学习算法的应用[27]。与MGI/MGE、ICME和I³M中强调的数字孪生智能制造范式相一致，通过所谓的模拟或数据驱动方法可以方便地解决选择和制造候选目标材料的准则、判据和

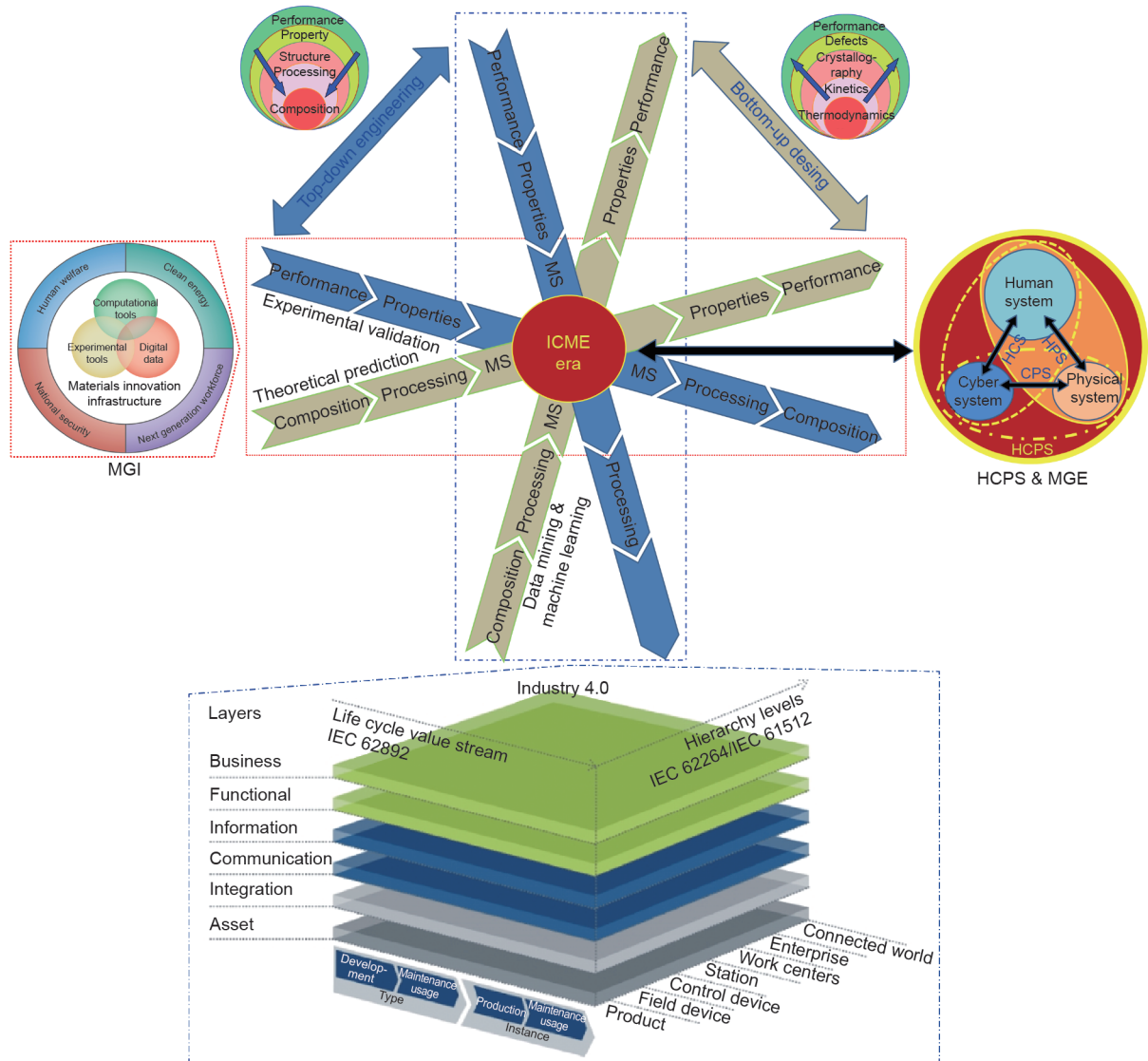


图1. ICME时代的数字孪生设计范式示意图，参考美国的MGI、中国的HCPS和MGE以及德国的工业4.0 [5]。MS：微观结构；HCS：人-信息系统；HPS：人-物理系统；CPS：信息-物理系统。

策略问题[5,19,21,26,28–31]。数据是材料4.0的基本资源，它是工业4.0下I³M的一种形式，如图2所示[8]。材料信息学将是未来工业制造的首要工作站，它包括材料加工和性能的大数据、机器学习算法、多尺度建模、虚拟合成和表征、原型测试和验证以及生命周期评估。例如，数字制造与设计创新研究所（DMDII）作为一个与学术界、行业和政府合作伙伴建立公私合作伙伴关系的机构，承担了通过数字技术提高美国制造业竞争力的责任[3]。DMDII的目标是将整个生命周期过程中的数据进行数字化并将其集成起来，获得更好的解决方案和决策，并在这个领域成为一个卓越的全球性组织[3]。迄今为止，DMDII的5年合作协议已取得了一些极其重要的经验、教训和问题，其中包括：①通过数字制造改善组织间的运作；②加速数字技术的创新；③多方合作促成创新解决方案；④解决数字制造技术中的“死亡谷”（valley of death）问题[3]。类似地，计算材料设计中心（Center for Computational Materials Design）是ICME的前身和催化剂，由美国国家科学基金会产业/大学合作研究中心于2005年成立[23]，是将学术界、行业和政府联系起来的基础。该中心可在一系列CPSPP关系中促进计算材料科学和力学的发展，其重点是培养和训练计算材料设计的未来生力军[23]。良好的数据管理和知识集成、数据发布过程后相关团体的再次利用以及知识探索和创新都是至关重要的[22]。

鉴于当前大数据及其使用的趋势，数据正呈指数级

增长，其生成方式正变得全球化，并正在向新兴市场转移[3]。自2011年以来，MGI已投资超过2.5亿美元，用于软件工具的开发、收集和报道实验数据的标准化方法的建立，重点大学的计算材料科学中心的建设以及大学与企业特定应用研究方面的合作等方面[32,33]。一个可持续的数据生态系统由一套机制（即标准）组成，这些机制的功能类似于自然界中的溪流，它们克服障碍，将数据从个体存储库传输到“数据海洋”（ocean of data），然后再循环将数据传回个体存储库[13]，如图3（a）所示[13]。FAIR原则可以被描述为一个生态系统，这个生态系统包括湖泊（各种数据存储库）、水流（相互连接）、渗流（私有数据）、海洋（收集）以及冷凝和降水（再利用）。这种运输和循环的驱动力或动力并非基于全部有用数据。因此，为了揭示数据在空间和时间上可能的联系及其与其他项目的关系，应将可操作或有价值的标记与元数据链接[3]，这是我们提出通用DID格式的初衷。

图3（b）[12,13,32]总结了当前材料基因组系统的技术、工具、模型和数据库基础，并强调未来基于相特性和结构调控显著提升系统功能。在材料基因组计划中需要更多的基础投资，重点是改进适用参数设计模型和构建高质量的数据库[32]。人们认为，可扩展的、自动优化的相平衡基础设施（ESPEI）将是建立“数据海洋”和开发具有多种缺陷的多组分材料性能数据库的重要组成部分[13]。在CPSPP关系[32,34,35]的指导下，图

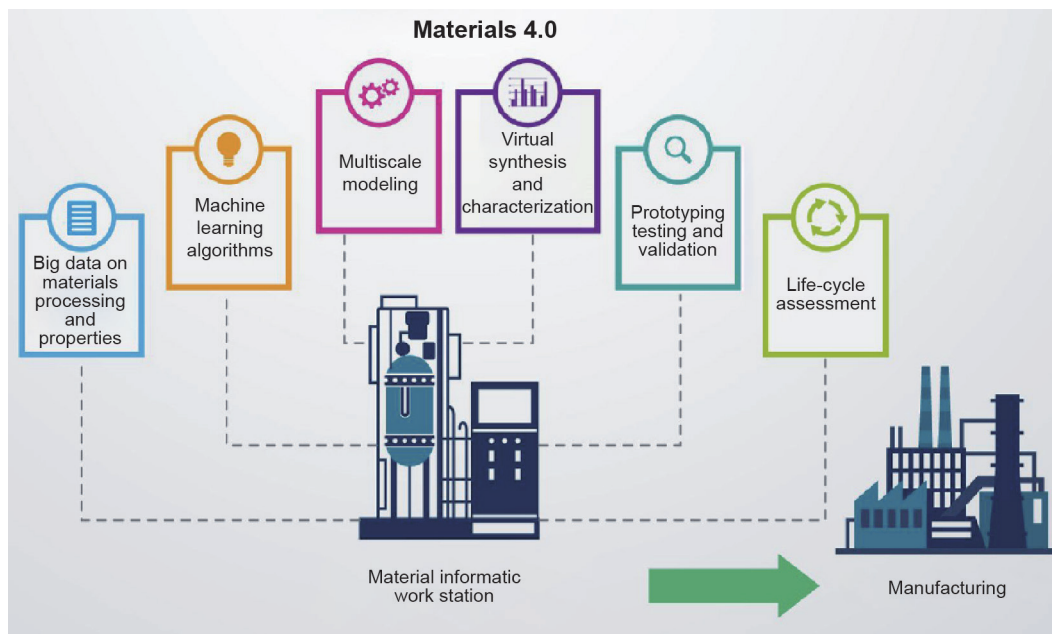
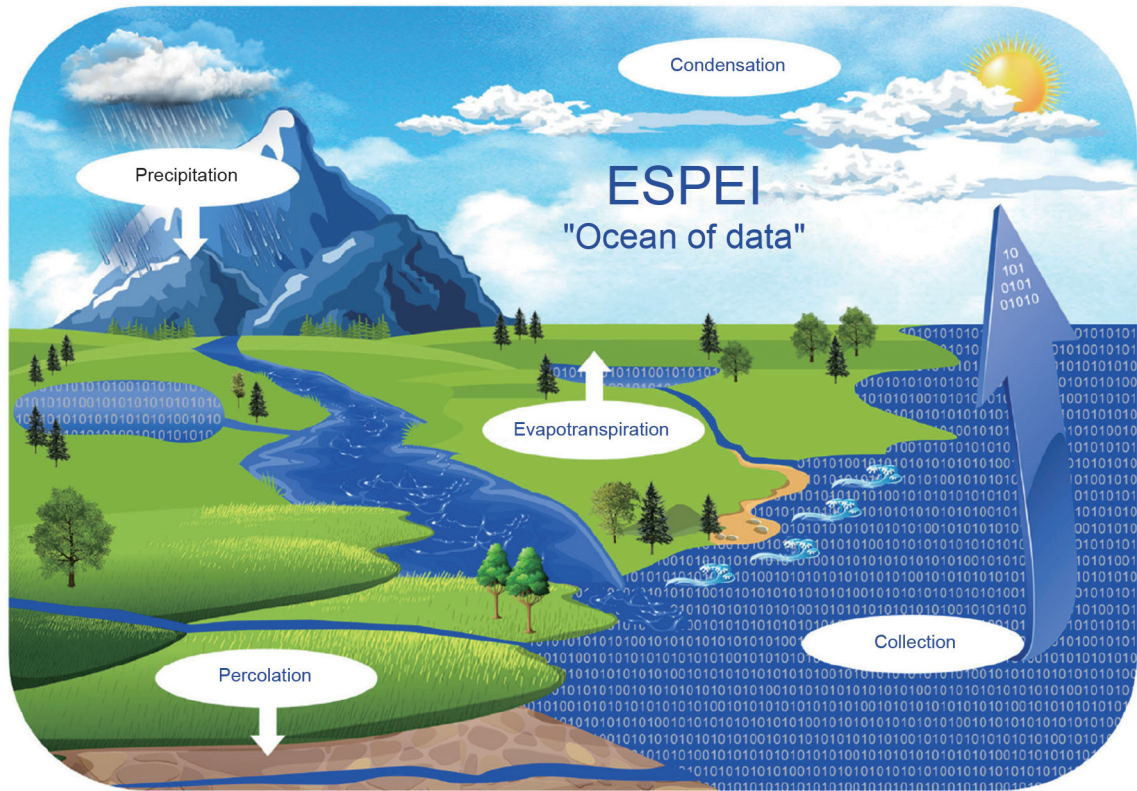
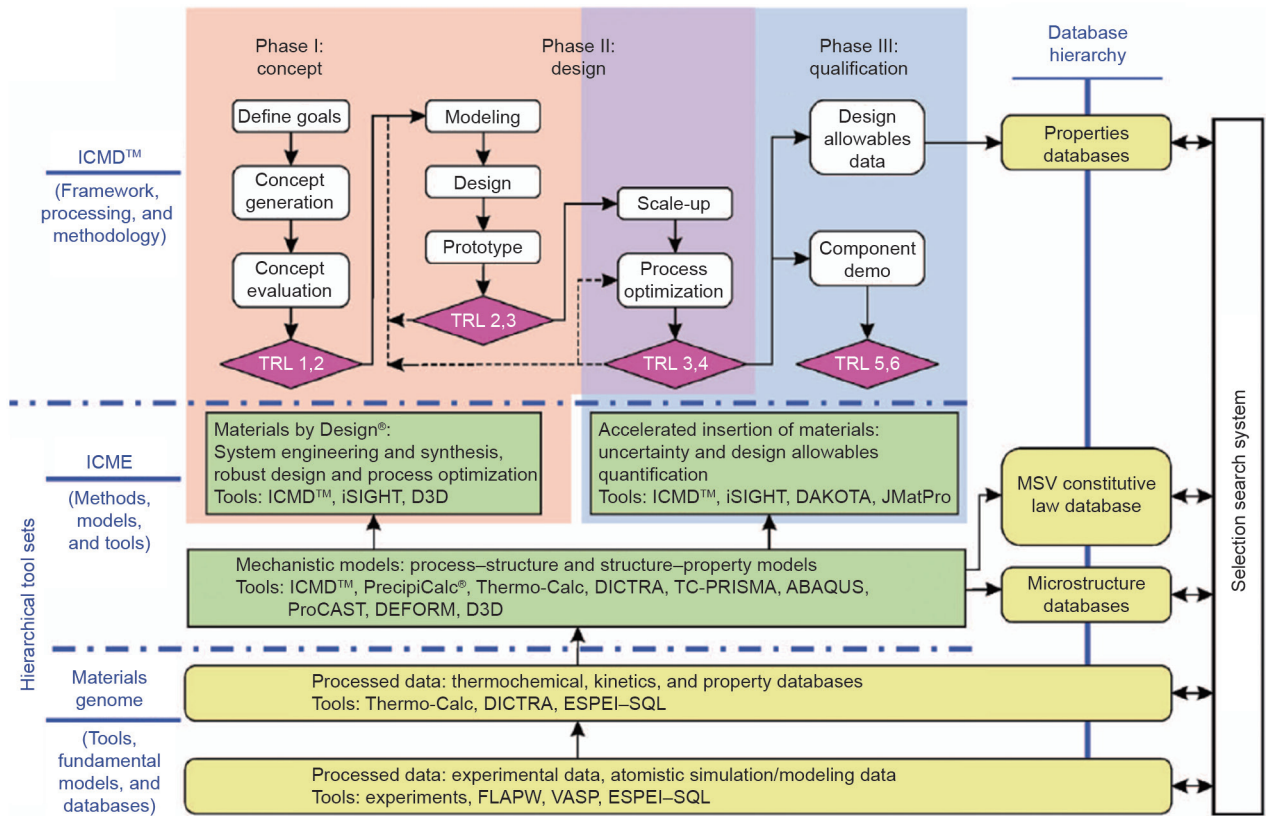


图2. 基于网络的材料大数据平台（材料4.0）的概念[8]。



(a)



(b)

图3. ESPEI在“数据海洋”中起着关键作用。(a)“数据海洋”的可持续生态系统示意图[13];(b)基于MGI/MGE的ICME方法应用的方法、工具、技术和数据库的整体层次结构[12,13,32]。MSV:多尺度变量;TRL:技术准备水平;FLAPW:全势线性化缀加平面波;VASP:维也纳从头算模拟软件包;ESPEI-SQL:可扩展的、自我优化的相位平衡基础结构-结构化查询语言;DICTRA:扩散控制转换;D3D:Direct3D。

3 (b) [12,13,32]显示了当前所有的计算材料设计方法都是基于处理过的数据,而处理过的数据又依赖于原始数据,ESPEI在这两种数据收集形式中都扮演着重要的角色,也就是说,可以假定这两组数据孕育了所有其他属性[13]。这个概念与图3 (a) [13]所示的“数据海洋”的概念非常吻合。此外,通过结合ICME的级联结构或基于MGI数据库的所谓的集成计算材料设计(ICMD),ICME/ICMD机械设计模型可以加速创新,最终将研究室理念转化为工业制造。

3. 标准系统

3.1. DID 编码原理及应用

集成物联网相关技术和基于云的技术将促进先进材料的发现和设计,并提高研究效率、性能和协作。通常而言,一系列基于云的工具(统称为物联网)可以集成实验室中的一切,从研究协议和设备到发表和数据存储[36]。这种数字实验室管理将远远优于当前的科学工作流程范式,甚至可能产生前所未有的研究方法,使今天所做的工作相形见绌[36]。例如,已经建立的一个名为nanoHub.org的网络平台,为全球172个国家/地区的24万多名用户提供计算纳米技术的网络支持[37]。由于它是一个科学云,nanoHub平台的用户可以设计和运行其工具,而无需安装或仅需最少的基础设施;因此,该平台以用户友好的方式把这些工具提供给全球[37]。

我们认为,以条形码或二维码形式存储的有价值 and 有用数据的标签将被视为一组数据的指纹,并将与云平台链接。因此,被标记的数据将有可能自动跟踪CPSPP工作流程中的进展和更新,这将加速新型先进材料的发现和制造,并提高研究效率、性能和协作(图4)。值得一提的是,建立这一云平台(即正在开发的www.MGE-TriD.com平台)的优先任务是在摩擦学研究领域发现新型先进润滑材料的协同应用。参与这个平台的所有研究人员都可以在全球范围内自发地合作。一旦从理论上或实验上生成大量数据,云基础设施和低成本存储设备就可以直接支持这个平台,并将这些数据推向可能感兴趣的研究人员[36]。

图5所示的是一种由一系列构建单元(意义段)组成的通用DID格式,该格式与国际和国家标准中使用的标识符的经典格式一致,如ISO/IEC 29168-1:2000 [38]、GB/T 27766-2011 [39]、GA/T 543.2-2011 [40]、GM/

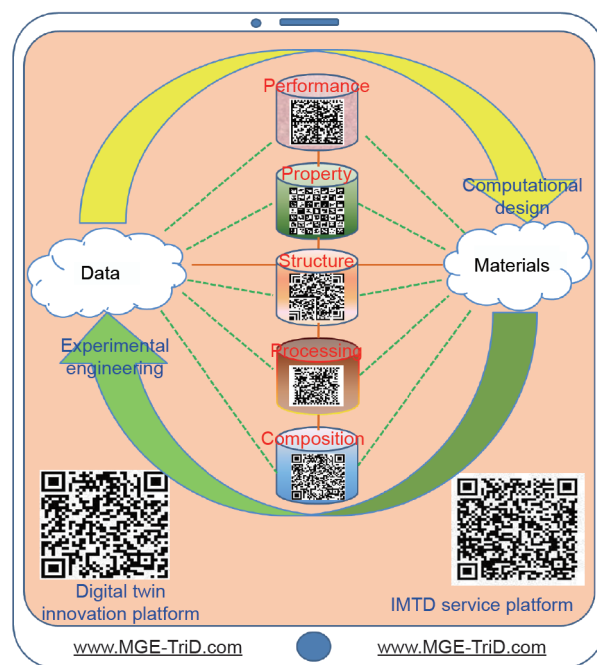


图4. 云平台示意图,展示了DID编码介导的数字孪生创新/制造范式,为自发跟踪CPSPP设计和发现程序铺平了道路。IMTD: 智能制造技术数据。

T 0006-2012 [41]、GJB 7365-2011 [42]、SL 325-2014 [43], SL 607-2018 [44]、WS 363.2-2011 [45]和QX/T 39-2005 [46]。在这里,每个构建单元均由大写字母和数字组成,没有符号,并且可以从成熟的云平台构建或转换。例如,智能制造技术数据(IMTD)服务平台已使用的DID原则已经被考虑,并集成到我们正在开发的平台(www.MGE-TriD.com)中。此外,每个构建单元的总长度不受限制,这遵循了ISO/IEC 10646国际标准中通用编码字符集格式。基于这些规则,本研究提出的DID具有灵活性,便于在各种云平台之间进行扩展和共享。因此,传统的二维码,包括汉信码(Hanxin Code)、龙贝(Lots Perception Matrix, LP)码、快速反应(Quick Response, QR)码、网格矩阵(Grid Matrix, GM)码和数据矩阵(Data Matrix, DM)码,可以被智能手机或特定机器构造并精确识别和解码。将这些二维码作为一组与其云平台链接的数据指纹,可以自动跟踪CPSPP工作流程中的进度和更新。

此外,DID编码不仅在云平台上提供数据指纹或一组数据的记录,而且还支持构建I³M核心的未来技术,包括先进材料、大数据分析、云计算、工业互联网和移动设备(图6)。DID编码将改善ICME时代数字孪生设计范式中HCS、HPS、CPS的交互。例如,数据驱动的智能ICME强调先进技术,如增材制造、机器学习和大

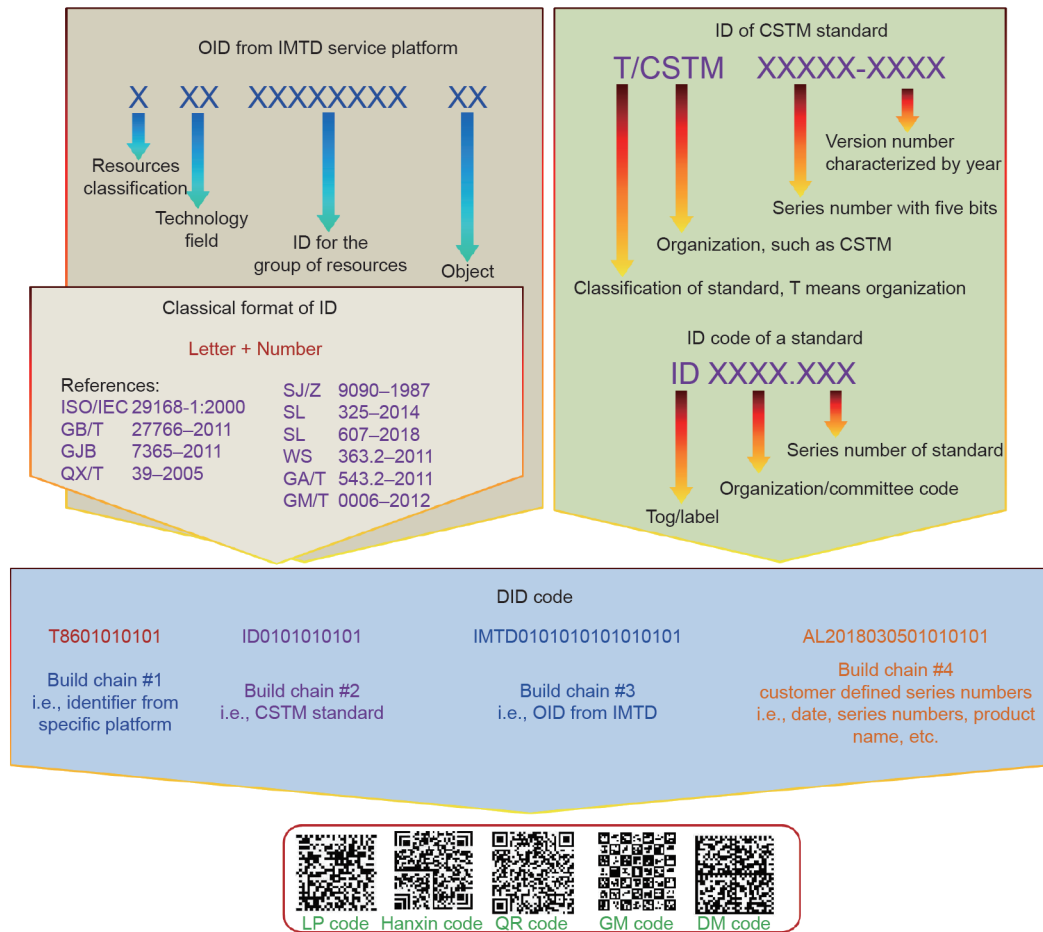


图5. 根据现有的国际、国家和组织标准，展示的DID编码的推荐构造规则示意图。OID：对象标识符；ID：标识符；LP：龙贝；QR：快速反应；GM：网格矩阵；DM：数据矩阵。

数据分析（或数据挖掘）。这些技术可以被认为是多尺度建模和基于模拟的材料和系统设计在航空和运输业中的催化剂[24,47]。机器学习正受到越来越多的关注，并且在先进材料的发现和设计过程中，无论是在时间效率还是在预测精度方面都取得了巨大的进步[9,48-52]。虚拟现实和增强现实不仅增强了人机交互，而且为数字和现实世界中的协同工作奠定了基础。在传统的可传承智能制造（I²M）[53]中，“可传承”指的是数字孪生设计范式的一个新特点，即核心技术、通用模型或基本原理不会随着时间或其他更新或变化而改变。正如华为技术有限公司在2019年的白皮书中称，集成化、智能化和可传承性（I³）功能被认为是未来基于数字云平台的三大基本功能。一方面，我们提出的构建DID编码的原则与云平台的数据指纹或数据记录类似，这符合I³的特性，可以为I³M服务。另一方面，可传承的特性体现了材料发现、设计和制造中的材料基因的原始理念。相信所有这些成果和进步都会降低制造成本，提高产品质量。

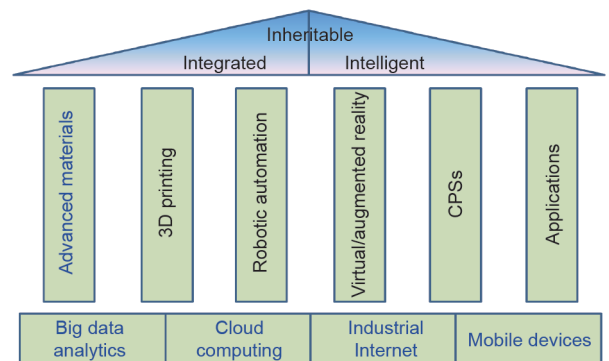
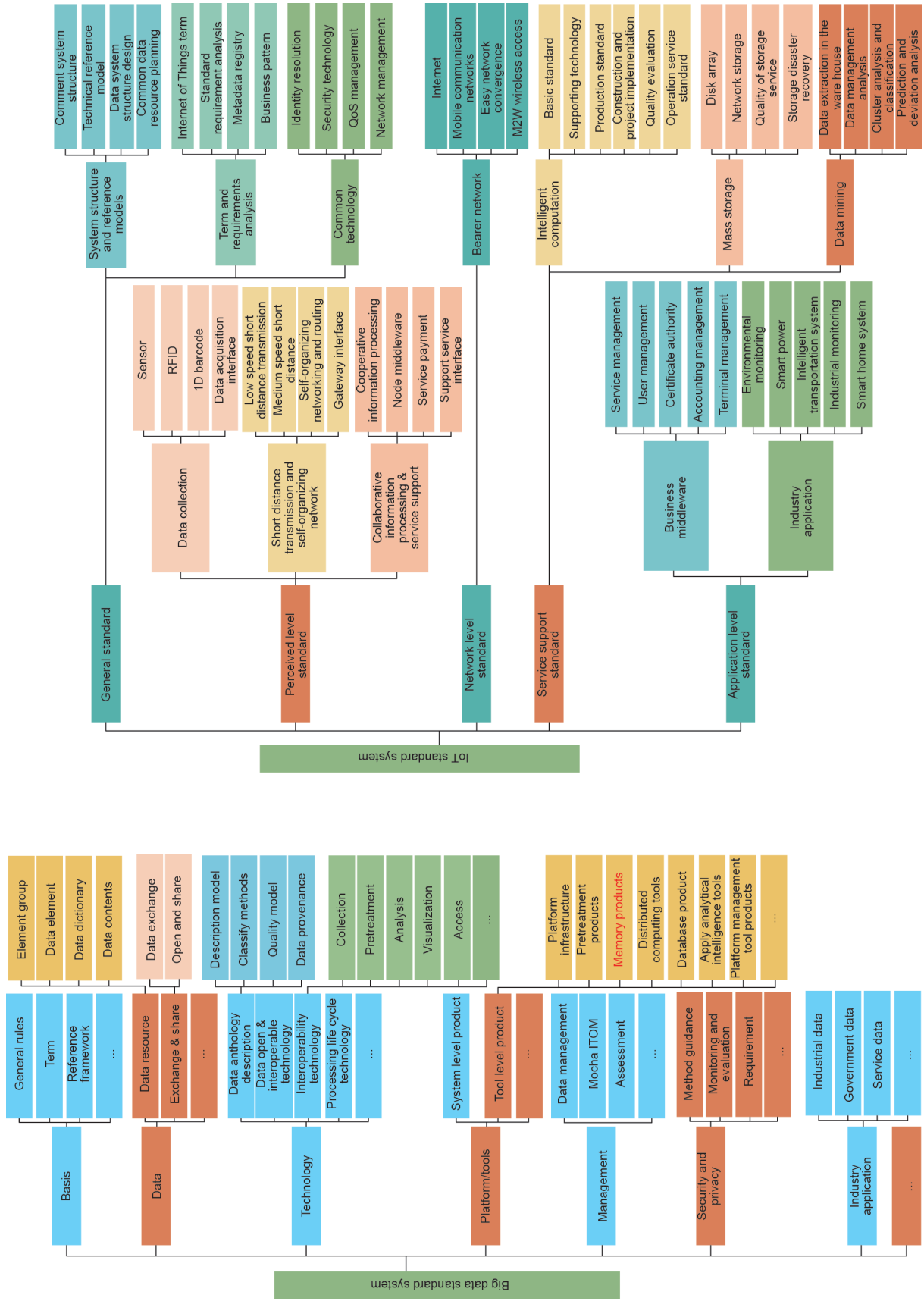


图6. 有助于I³M的10项关键技术。蓝色突出显示的技术可能涉及DID编码。

3.2. 基于云技术的标准系统

最后，有必要强调的是，与材料信息学相关的数字标准对于机器学习算法、数据挖掘和人工智能的发展以及材料创新、发现和设计的加速至关重要。例如，NIST正在开发若干面向制造系统技术和两个颠覆性制造领域



(a)

(b)

图7. 中国提出的大数据和物联网系统标准框架。ITOM: 信息技术运营管理; RFID: 射频识别; QoS: 服务质量。

(即机器人系统和增材制造)的测量标准[3]。图7给出了中国大数据和物联网系统标准的框架,这将是I³M的基础。值得期待的是,我们提出的数据标识编码标准有望成为这两个标准体系的一部分。下一代科学标准的构建正在进行,这将改变我们对未来生力军的培养和训练。

4. 结论

本文简要介绍了设计时代ICME数字孪生设计范式的几种观点,并提出了一种由一组构建单元组成的通用DID格式,且该格式与国际和国家标准中使用的经典标识符格式一致。本文所提出的DID具有灵活性,便于在各种云平台之间进行扩展和共享。值得一提的是,建立MGE云平台的主要目的是发现新型先进材料的协同应用。这个平台上的所有研究人员可以实时在全球范围内合作。一方面,我们提出的构建DID编码的原则与I³特性相一致,可以为I³M服务;另一方面,支持未来有助于I³M的技术,包括高级材料、大数据分析、云计算、工业互联网和移动设备,可以改善HCS、HPS和CPS之间的相互作用。遗传特征起源于材料发现、设计和制造过程中“材料基因组”的原始概念。由于我们提出的构建DID编码的原则可作为I³M云平台上数据记录的指纹,因此可以期待,这些原则可能会被融入或整合进中国的大数据和物联网标准体系。

致谢

本研究得到了国家重点研究发展计划(2018YFB-0703801、2018YFB0703802、2016YFB0701303和2016YFB0701304)和中国中车唐山机车车辆有限公司(201750463031)的资金支持。特别感谢上海交通大学汪洪教授的富有成效的讨论及建设性的建议。

Compliance with ethics guidelines

William Yi Wang, Peixuan Li, Deye Lin, Bin Tang, Jun Wang, Quanmei Guan, Qian Ye, Haixing Dai, Jun Gao, Xiaoli Fan, Hongchao Kou, Haifeng Song, Feng Zhou, Jijun Ma, Zi-Kui Liu, Jinshan Li, and Weimin Liu declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] Zhou J, Li P, Zhou Y, Wang B, Zang J, Meng L. Toward new-generation intelligent manufacturing. *Engineering* 2018;4(1):11–20.
- [2] Zhong RY, Xu X, Klotz E, Newman ST. Intelligent manufacturing in the context of Industry 4.0: a review. *Engineering* 2017;3(5):616–30.
- [3] National Academies of Sciences, Engineering, and Medicine. The Fourth Industrial Revolution: proceedings of a workshop—in brief. Washington, DC: National Academies Press; 2017.
- [4] National Science Technology Council. Materials genome initiative for global competitiveness. Washington, DC: National Science and Technology Council; 2011.
- [5] Wang WY, Li JS, Liu WM, Liu ZK. Integrated computational materials engineering for advanced materials: a brief review. *Comp Mater Sci* 2019;158:42–8.
- [6] de Pablo J, Jackson NE, Webb MA, Chen LQ, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. *NPJ Comp Mater* 2019;5(1):41.
- [7] Jain A, Persson KA, Ceder G. Research update: the materials genome initiative: data sharing and the impact of collaborative *ab initio* databases. *APL Mater* 2016;4(5):053102.
- [8] Jose R, Ramakrishna S. Materials 4.0: materials big data enabled materials discovery. *Appl Mater Today* 2018;10:127–32.
- [9] Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. *J Mater* 2017;3(3):159–77.
- [10] Meredig B, Wolverton C. A hybrid computational–experimental approach for automated crystal structure solution. *Nat Mater* 2013;12(2):123–7.
- [11] Kaufman L, Agren J. CALPHAD, first and second generation—birth of the materials genome. *Scr Mater* 2014;70:3–6.
- [12] Olson GB, Kuehmann CJ. Materials genomics: from CALPHAD to flight. *Scr Mater* 2014;70:25–30.
- [13] Liu ZK. Ocean of data: integrating first-principles calculations and CALPHAD modeling with machine learning. *J Phase Equilib Diffus* 2018;39(5):635–49.
- [14] Zhou BC, Wang WY, Liu ZK, Arróyave R. Electrons to phases of magnesium. In: Horstemeyer MF, editor. *Integrated computational materials engineering (ICME) for metals: concepts and case studies*. New York: John Wiley & Sons; 2018.
- [15] Committee on Integrated Computational Materials Engineering, National Research Council. *Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security*. Washington, DC: National Academies Press; 2008.
- [16] Pollock TM. Alloy design for aircraft engines. *Nat Mater* 2016;15(8):809–15.
- [17] Krajewski PE, Hector LG, Qi Y, Mishra RK, Sachdev AK, Bower AF, et al. Atoms to autos: a multi-scale approach to modeling aluminum deformation. *JOM* 2011;63(11):24–32.
- [18] Modeling across scales: a roadmapping study for connecting materials models and simulations across length and time scales [Internet]. Warrendale: The Minerals, Metals & Materials Society; c2015 [cited 2019 Apr 23]. Available from: <https://www.tms.org/multiscalestudy>.
- [19] Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, et al. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat Nanotech* 2018;13(3):246–52.
- [20] Sendek AD, Yang Q, Cubuk ED, Duerloo KAN, Cui Y, Reed EJ. Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. *Energy Environ Sci* 2017;10(1):306–20.
- [21] Cubuk ED, Ivancic RJS, Schoenholz SS, Strickland DJ, Basu A, Davidson ZS, et al. Structure-property relationships from universal signatures of plasticity in disordered solids. *Science* 2017;358(6366):1033–7.
- [22] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
- [23] Liu ZK, McDowell DL. The Penn State–Georgia Tech CCMD: ushering in the ICME Era. *Integr Mater Manuf Innovation* 2014;3:409–28.
- [24] Liu X, Furrer D, Koster J, Holmes J. Vision 2040: a roadmap for integrated, multiscale modeling and simulation of materials and systems. Technical paper. Washington, DC: NASA; 2018. No.: NASA/CR-2018-219771.
- [25] Aspuru-Guzik A, Persson K, Tribukait-Vasconelos H. Materials acceleration platform—accelerating advanced energy materials discovery by integrating high-throughput methods with artificial intelligence [Internet]. Mission Innovation; 2018 Jan [cited 2019 Apr 23]. Available from: <http://missioninnovation.net/wp-content/uploads/2018/01/Mission-Innovation-IC6-Report-Materials-Acceleration-Platform-Jan-2018.pdf>.
- [26] Umehara M, Stein HS, Guevarra D, Newhouse PF, Boyd DA, Gregoire JM. Analyzing machine learning models to accelerate generation of fundamental materials insights. *NPJ Comp Mater* 2019;5:34.
- [27] Broderick SR, Santhanam GR, Rajan K. Harnessing the big data paradigm for ICME: shifting from materials selection to materials enabled design. *JOM* 2016;68:2109–15.
- [28] Xia Z. Hydrogen evolution: guiding principles. *Nat Energy* 2016;1(10):16155.
- [29] Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. *NPJ Comp Mater* 2018;4(1):25.

- [30] Ulissi ZW, Medford AJ, Bligaard T, Nørskov JK. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat Commun* 2017;8(1):14621.
- [31] Wang WY, Tang B, Lin DY, Zou C, Zhang Y, Shang SL, et al. A brief review of data-driven ICME for intelligently discovering advanced structural metal materials: insight into atomic and electronic building blocks. *J Mater Res* 2020;35:872–89.
- [32] Xiong W, Olson GB. Cybermaterials: materials by design and accelerated insertion of materials. *NPJ Comp Mater* 2016;2(1):15009.
- [33] Nosengo N. The material code. *Nature* 2016;533(7602):22–5.
- [34] Olson GB. Designing a new material world. *Science* 2000;288(5468):993–8.
- [35] Frankel D, Hatcher N, Snyder D, Sebastian J, Olson GB, Vernon G, et al. Improving manufacturing quality using integrated computational materials engineering. In: *Proceedings of the 4th World Congress on Integrated Computational Materials Engineering*; 2017 May 21–25; Ypsilanti, MI, USA; 2017. p. 23–32; 2017.
- [36] May M. Companies in the cloud: digitizing lab operations. *Science* 2017;355(6324):532–4.
- [37] Madhavan K, Zentner L, Farnsworth V, Shivarajapura S, Zentner M, Denny N, et al. nanoHUB.org: cloud-based services for nanoscale modeling, simulation, and education. *Nanotechnol Rev* 2013;2(1):107–17.
- [38] The Joint Technical Committee. ISO/IEC JTC 1. ISO/IEC 29168-1:2011: Information technology—open systems interconnection—part 1: object identifier resolution system. ISO standard. Geneva: International Organization for Standardization; 2011.
- [39] Zhang W, Zhang YP, Liu B, Zhang DY, Fan XC. GB/T 27766–2011: Twodimensional barcode—grid matrix code. Chinese standard. Beijing: General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration; 2011. Chinese.
- [40] Ma WY, Zhang SJ, Tang YJ, Chi SC, Luo Z, Xu WH, et al. GA/T 543.2–2011: The data element of public security (2). Chinese standard. Beijing: The Ministry of Public Security of the People's Republic of China; 2011. Chinese.
- [41] Liu P, Liu XD, Kong FY, Li YZ, Xu Q, Liu ZS, et al. GM/T 0006–2012: Cryptographic application identifier criterion specification. Chinese standard. Beijing: State Cryptography Administration; 2012. Chinese.
- [42] Fan XC, Geng GP, Pan H, Huang XL, Zhang YP, Zhang G, et al. GJB 7365–2011: Grid matrix code. Chinese standard. Beijing: Chinese PLA General Armament Department; 2011. Chinese.
- [43] Mao XW, Peng H, Pan MM, Gao JJ, Li YT, Cheng YL, et al. SL 325–2014: Structure and identifier for water quality database. Chinese standard. Beijing: Ministry of Water Resources of the People's Republic of China; 2014. Chinese.
- [44] Jiang B, Liu Q, Xu P, Xia P, Yu YC, Shao YT, et al. SL 607–2013: Structure and identifier in water resources literature database. Chinese standard. Beijing: Ministry of Water Resources of the People's Republic of China; 2013. Chinese.
- [45] Meng Q, Liu LH, Li LJ, Xu KJ, Ren QQ, Wang Y. WS 363.2–2011: Health data element dictionary—part 2: identification. Chinese standard. Beijing: Ministry of Health of the People's Republic of China; 2011. Chinese.
- [46] Wang GF, Wu ZQ. QX/T 39–2005: Core metadata content of meteorological dataset. Chinese standard. Beijing: China Meteorological Administration; 2005. Chinese.
- [47] Shapira P. Making the future. *Science* 2017;358(6366):1007.
- [48] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machinelearning-assisted materials discovery using failed experiments. *Nature* 2016;533(7601):73–6.
- [49] Kavvas ES, Catoi E, Mih N, Yurkovich JT, Seif Y, Dillon N, et al. Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nat Commun* 2018;9(1):1–9.
- [50] Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N. Deep learning and process understanding for data-driven Earth system science. *Nature* 2019;566(7743):195–204.
- [51] Gray J, Banchi L, Bayat A, Bose S. Machine-learning-assisted many-body entanglement measurement. *Phys Rev Lett* 2018;121:150503.
- [52] Kim K, Ward L, He J, Krishna A, Agrawal A, Wolverson C. Machine-learningaccelerated high-throughput materials screening: discovery of novel quaternary Heusler compounds. *Phys Rev Mater* 2018;2(12):123801.
- [53] Chen Y. Integrated and intelligent manufacturing: perspectives and enablers. *Engineering* 2017;3(5):588–95.