A data identifier (DID) is an essential tag or label in all kinds of databases—particularly those related to integrated computational materials engineering (ICME), inheritable integrated intelligent manufacturing (I3M), and the Industrial Internet of Things. With the guidance and quick acceleration of the development of advanced materials, as envisioned by official documents worldwide, more investigations are required to construct relative numerical standards for material informatics. This work proposes a universal DID format consisting of a set of build chains, which aligns with the classical form of identifier in both international and national standards, such as 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, and QX/T 39–2005. Each build chain is made up of capital letters and numbers, with no symbols. Moreover, the total length of each build chain is not restricted, which follows the formation of the Universal Coded Character Set in the international standard of ISO/IEC 10646. Based on these rules, the proposed DID is flexible and convenient for extending and sharing in and between various cloud-based platforms. Accordingly, classical two-dimensional (2D) codes, including the Hanxin Code, Lots Perception Matrix (LP) Code, Quick Response (QR) code, Grid Matrix (GM) code, and Data Matrix (DM) Code, can be constructed and precisely recognized and/or decoded by either smart phones or specific machines. By utilizing these 2D codes as the fingerprints of a set of data linked with its cloud-based platforms, progress and updates in the composition–processing–structure–property–performance workflow process can be tracked spontaneously, paving a path to accelerate the discovery and manufacture of advanced materials and enhance research productivity, performance, and collaboration.

1. Introduction

As humanity moves into the integrated intelligent manufacturing era [1–3], the digital twin design paradigm of integrated computational materials engineering (ICME) is critical to accelerate the discovery and applications of novel advanced materials. The unique challenges and opportunities of computational materials engineering and the envisioned strategy for future inheritable integrated intelligent manufacturing (I3M) have been outlined in numerous official documents and plans, such as the Materials Genome Initiative (MGI) in the United States [4], Materials Genome Engineering (MGE) and human–cyber–physical systems (HCPSs) in China [1], Industry 4.0 in Germany [2], and Industry Innovation 3.0 in Republic of Korea [5]. Materials innovations enable new technological capabilities to solve problems and drive major societal advancements [6,7]. Based on these advanced materials design paradigms, materials discovery and engineering ingenuity open up new frontiers for technological advancement [6],...
such as big data [8], data mining [6], machine learning [6,9], artificial intelligence, cloud computation, metallic materials ontology, and graphic knowledge. The techniques related to big data and machine learning serve to link theoretical predictions with microscopic degrees of freedom, and thus accelerate the design and synthesis of regional materials [9]. For example, by integrating experimental diffraction data, statistical feedback of symmetry, and density functional theory (DFT)-based optimal algorithms, the first-principles-assisted structure solution has become a novel hybrid approach to automatically predict crystal structures [10]. The integration of DFT calculations and calculation of phase diagrams (CALPHADs) is considered to be the major part of the materials genome and material design [11–13], and has established a robust materials development framework of data-driven ICME [5,14], which emphasizes on phase-based properties databases. It is understood that ICME is the integration of materials information, captured by computational tools, with engineering product performance analysis and manufacturing process simulations [5,15].

With the quick development of computational materials science and computational power, multiphysics simulations drive the predictions of thermodynamics, kinetics, structures, defects, and properties at the multiscale, which dramatically accelerates the growth of material databases or repositories [5,13,16]. Moreover, by integrating multiscale simulations, design strategies can cross the range from electrons to phases and to products [12,14,17,18], and the principles and criteria for selecting targeted candidate materials can be addressed efficiently [5,13,19–21].

From the perspective of integrated intelligent manufacturing/ engineering, data-driven ICME supports a digital twin type of design/manufacturing paradigm, which highlights the important role of materials informatics. While the Third Industrial Revolution (the so-called Digital Revolution) is still under development, the Fourth Industrial Revolution focuses on infrastructures and digital technologies, and integrates nonlinearity and a reappearance of digital technologies and disciplines into the area of virtual materials and physical systems [3]. In the age of design aided by intelligence, new targets, candidates, and strategies will be identified by computational methods ahead of time, making it beneficial for companies to deliver solutions based on the obtained knowledge and models at lower cost and in a shorter time, in order to survive global competition. To make effective decisions about target design, optimization, plans, and solutions, it is extremely important to address the recognized “four Vs” of data, which include volume, velocity, variety, and veracity [3]. It has been reported that almost $7$ trillion USD has been spent by big manufacturers in North America to upgrade previous equipment with sensors in order to enable systems to communicate with each other through the Internet of Things (IoT) [3]. It is a great challenge—or opportunity—to efficiently use this operational data to make business decisions, which is only being done for about $1\%$ of business decisions at present [3]. Not only should there be concerted investigation into tools that promote the automated collection, curation, and distribution of datasets, but standardized data and metadata formats should also be defined and followed [6]. For example, the four foundational principles of findability, accessibility, interoperability, and reusability (FAIR) have been proposed as a guideline for data in the conventional sense, algorithms, tools, and workflows. The so-called FAIR guiding principles should be taken into account during the collection and sharing of data [6,22,24].

Our recent brief review [5] discussed the dominant roles of databases, toolkits, platforms, principles, benchmarks, and standards in the recent framework of data-driven ICME. Since the short-term goals of data-driven ICME have almost been completed, the long-term goals are under way, and include educating the next-generation ICME workforce and establishing a web-based ICME infrastructure, in order to improve global competitiveness for industry and national security. It is understood that constructing relative numerical standards for material informatics is necessary in the development of data mining, deep/machine learning, and artificial intelligence; in the verification of valuable data; and in the acceleration of materials innovations, discovery, and design. This also requires increased interaction and collaboration with industrial partners [5,6,23]. With the aim of addressing the issues that occur when translating technologies from laboratory to practical applications, the National Institute of Science and Technology (NIST) has proposed both performance and interoperability standards in order to accelerate innovation and minimize the risk involved in the application of novel smart manufacturing technologies [3,5]. In addition, the recently proposed Chinese Society of Testing and Materials (CSTM) standard draft titled General rule for materials genome engineering data is a first attempt to standardize the content of MGE data, which will have deep implications for the transformation of materials science into a data-driven scientific regime.

In the present work, a universal format of data identifier (DID) consisting of a set of build chains is proposed as a part of a CSTM effort to systematically establish an MGE data standard, which aligns with the classical form of identifier that is utilized in both international and national standards. The proposed DID is flexible and convenient for extending and sharing in and between various cloud-based platforms. Accordingly, classical two-dimensional (2D) codes can be constructed and precisely recognized and decoded by either smart phones or specific machines. By utilizing these 2D codes as the fingerprints of a set of data linked with its cloud-based platforms, progress and updates in the composition–processing–structure–property–performance (CPSPP) workflow process will be tracked spontaneously, paving a path to promote the development of advanced materials and enhance research productivity, performance, and collaboration.

2. The digital twin intelligent manufacturing era

2.1. Digital twin manufacturing in the age of design

Digital twins are computerized companions of physical assets; they are a novel design paradigm in the ICME era and are used for aircraft, trains, and engines [3,5]. As shown in Fig. 1 [5], the typical discovery, design, innovation, and manufacturing chain of advanced materials consists of composition, processing, microstructure, properties, and performance. The CPSPP relationship or workflow process in materials science is extremely important in guiding the discovery and manufacturing of materials, and requires advanced technologies, including high-throughput computations, additive manufacturing, artificial intelligence, data mining, machine/deep learning, and so on [2,5,24,25]. In view of bottom-up design and top-down engineering [23], the features of digital twin between the experimental and theoretical chains are highlighted in different background colors. While the MGI emphasizes the dominant roles of experimental tools, computational tools, and databases, along with their interactions, the HCPSS and MGE emphasize the interactions of HCPSSs, which hint at the future applications of the advanced technologies described above. Industry 4.0 emphasizes that intelligent manufacturing is enabled through cyber–physical systems (CPSSs) [8]. Two aims have been addressed in future perspectives on intelligence manufacturing: (1) to envision and promote boldness and leadership in industry in the United States; and (2) to benefit workforce development [3]. These aims are also considered within HCPSSs and the MGE in China. Machine learning algorithms can accelerate the generation of fundamental insights into materials and basic science research through the identification of dominates and valuable data.
relationships to strengthen human–physical and human–cyber interpretations and yield scientific knowledge and models [26]. After a model has been established to predict the performance from the parameters of materials, further analysis of gradients in the training model can determine the main and valuable data relationships that cannot be readily established by means of human inspection or traditional statistical analysis [26]. In the digital twin intelligent manufacturing era (also called the “age of design”), by integrating theoretical bottom-up design with experimental top-down engineering routines, it is expected that novel advanced materials will be designed and discovered more efficiently at lower cost using the digital twin intelligent manufacturing approach. This will accelerate the growth of MGI infrastructures, strengthen the interactions in HCPSs, and support a novel approach to simultaneously fabricate the future by utilizing the latest advanced technologies.

2.2. Designing in an ecosystem

Data and data infrastructures are one of the three foundations that guarantee the successful progress of MGI/MGE, ICME, and I\textsuperscript{4}M. All these terms are based on informatics, which is a broad term encompassing data-driven design stages, such as warehousing, visualization, and the application of statistical learning algorithms [27]. In line with the digital twin intelligent manufacturing paradigm emphasized in the MGI/MGE, ICME, and I\textsuperscript{4}M, the principles, criteria, and strategies in selecting and manufacturing candidate target materials will be addressed conveniently by so-called simulation- or data-driven approaches [5,19,21,26,28–31]. Data is the elementary resource for Materials 4.0, which is one form of I\textsuperscript{3}M under Industry 4.0, as shown in Fig. 2 [8]. Material informatics will be the future primary workstation for industrial manufacturing; it consists of the big data on materials processing and properties, machine learning algorithms, multiscale modeling, virtual synthesis and characterization, prototyping testing and validation, and life-cycle assessment. For example, the Digital Manufacturing and Design Innovation Institute (DMDII), a public–private partnership with academic, industry, and government partners, has assumed the responsibility to improve the competitiveness of US manufacturing through digital technologies [3]. The goal of the DMDII is to be a preeminent worldwide organization for digitizing data across the life-cycle.

Fig. 1. Schematic diagram of the digital twin design paradigm in the ICME era, referring to the MGI in the United States, the HCPS and MGE in China, and Industry 4.0 in Germany [5]. MS: microstructure; HCS: human–cyber system; HPS: human–physical system; CPS: cyber–physical system.
processes and integrating it in order to obtain better solutions and decisions [3]. Several extremely important experiences, lessons, and questions have been obtained thus far in the DMDII’s five-year cooperative agreement, including: ① improving the operations among organizations through digital manufacturing; ② accelerating innovations in digital technologies; ③ multi-party collaboration enabling innovative solutions; and ④ solving the “valley of death” problem in digital manufacturing technologies [3]. Similarly, the Center for Computational Materials Design, a predecessor of and catalyst for the ICME, was founded by the National Science Foundation Industry/University Cooperative Research Center in 2005 [23], and serves as a basis for coupling academia, industry, and government to advance the state of computational materials science and mechanics across a portfolio of CPSPP relationships, with an emphasis on the education and training of the future workforce in computational materials design [23]. Good data management is crucial to subsequent data and knowledge integration and reuse by the community after the data publication process, as well as to knowledge exploration and innovation [22].

Given the current trend of big data and its use, data is increasing exponentially and its generation is becoming global and is shifting toward emerging markets [3]. Since 2011, the MGI has invested more than $250 million USD in software tools, standardized methods to collect and report experimental data, centers for computational materials science and mechanics across a portfolio of CPSPP relationships, with an emphasis on the education and training of the future workforce in computational materials design [23]. Good data management is crucial to subsequent data and knowledge integration and reuse by the community after the data publication process, as well as to knowledge exploration and innovation [22].

Fig. 2. The concept of a web-based materials big data platform, or Materials 4.0 [8].
Fig. 3. The extensible, self-optimizing phase-equilibrium infrastructure (ESPEI) mediates a key role in the “ocean of data.” (a) Schematic diagram of the sustainable ecosystem of the “ocean of data” [13]; (b) overall hierarchical architecture of the methods, tools, techniques, and databases for the application of ICME methods based on MGI/MGE [12,13,32]. MSV: multiscale variable; TRL: technology readiness level; FLAPW: full-potential linearized augmented plane wave; VASP: Vienna ab-initio simulation package; ESPEI–SQL: extensible, self-optimizing phase-equilibrium infrastructure-structured query language. DICTRA: diffusion controlled transformation; D3D: Direct3D.
to the development of unprecedented methods of research that would eclipse what can be done today [36]. For example, a cyber infrastructure named nanoHub.org has been established and supports the Network for Computational Nanotechnology for more than 240,000 users in over 172 countries worldwide [37]. As it is a scientific cloud, the users of the nanoHub platform can design and run their tools with no installation or minimal infrastructure requirements; thus, the platform provides a worldwide service of these tools through a user-friendly approach [37].

In our opinion, the tag in the form of bar codes or 2D codes for valuable and useful data will be considered the fingerprints for a set of data and will be linked with its cloud-based platforms. Accordingly, it will be possible to spontaneously track progress and updates in the CPSPP workflow process, which will accelerate the discovery and manufacture of novel advanced materials and enhance research productivity, performance, and collaboration (Fig. 4). It is worth mentioning that the main priority in setting up this cloud-based platform (i.e., the ongoing platform of www.MGE-TriD.com) is its collaborative applications in discovering novel advanced lubricating materials in the research area of tribology. All of the researchers involved in this platform can work together globally and spontaneously. Once a vast amount of data is generated theoretically or experimentally, the cloud infrastructure and low-cost storage equipment can directly support it and push it toward researchers who may be interested in it [36].

As shown in Fig. 5, a universal DID format consisting of a set of build chains is proposed, which aligns with the classical form of identifier that is utilized in both international and national standards. This format provides a path for spontaneously tracking CPSPP design and discovery procedures. IMTD: intelligent manufacturing technology data.

**Fig. 4.** Schematic diagram of a cloud-based platform presenting the DID coding-mediated digital twin innovation/manufacturing paradigm, which paves a path for spontaneously tracking CPSPP design and discovery procedures. IMTD: intelligent manufacturing technology data.

**Fig. 5.** Schematic diagram presenting the recommended constructing rules for the DID code in line with available international, national, and organization standards. OID: object identifier; ID: identifier; LP: Lots Perception Matrix; QR: Quick Response; GM: Grid Matrix; DM: Data Matrix.
standards, such as ISO/IEC 29168:1-2000 [38], GB/T 27766–2011 [39], GA/T 543.2–2011 [40], GM/T 0006–2012 [41], GJB 7365–2011 [42], SL 325–2014 [43], SL 607–2018 [44], WS 363.2–2011 [45], and QX/T 39–2005 [46]. Here, each build chain is made up of capital letters and numbers with no symbols, and can be constructed or transformed from well-established cloud-based platforms. For example, the DID principles utilized in the intelligent manufacturing technology data (IMTD) service platform have been considered and integrated into our ongoing platform (www.MGE-Trid.com). Moreover, the total length of each build chain is not restricted, which follows the formation of the Universal Coded Character Set in the international standard of ISO/IEC 10646. Based on these rules, the proposed DID is flexible and convenient for extending and sharing in and between various cloud-based platforms. Accordingly, classical 2D codes, including the Hanxin Code, Lots Perception Matrix (LP) Code, Quick Response (QR) code, Grid Matrix (GM) code, and Data Matrix (DM) Code, can be constructed and precisely recognized and decoded by either smart phones or specific machines. By utilizing these 2D codes as the fingerprints of a set of data linked with its cloud-based platforms, progress and updates in the CPSPP workflow process can be tracked spontaneously.

Furthermore, DID code not only provides data fingerprints or a record for a set of data on a cloud-based platform, but also supports future technologies for constructing the core of IM, including advanced materials, big data analytics, cloud computing, the Industrial Internet, and mobile devices (Fig. 6). DID code will improve HCS, HPS, CPS interactions in the digital twin design paradigm in the ICME era. For example, advanced technologies such as additive manufacturing, machine learning, and big data analytics (or data mining) have been highlighted in the data-driven intelligent ICME, and can be considered to be catalysts for multiscale modeling and the simulation-based design of materials and systems in the aerospace and transportation industries [24–47]. Machine learning is attracting increasing attention and has achieved great progress in the discovery and design of advanced materials in terms of both time efficiency and prediction accuracy [9,48–52]. Virtual reality and augmented reality not only enhance human–computer interactions, but also pave a new path for collaborative work in both the digital and real worlds. In traditional inheritable intelligent manufacturing (IM) [53], “inheritable” refers to a novel feature of the digital twin design paradigm; namely, that the central technologies, universal models, or fundamental principles are not changed over time or with other updates or variations. The integrated, intelligent, and inheritable (I3) features are considered to be the three basic features of future digital cloud-based platforms, as reported in a white paper from Huawei in 2019. On the one hand, our proposed principles for constructing the DID code are just like data fingerprints or a data record for a cloud-based platform; this aligns with the I3 features and can serve the I3M. Furthermore, the inheritable feature indicates the original concept of the material genome in materials discovery, design, and manufacturing. It is believed that all of these achievements and improvements will result in a reduction in manufacturing cost and an enhancement of product quality.

3.2. Standard systems for cloud-based technologies

In the end, it is necessary to emphasize that relative numerical standards for material informatics are essential in the development of machine learning algorithms, data mining, and artificial intelligence, and in the acceleration of materials innovation, discovery, and design. For example, several measurement standards in manufacturing systems-oriented technologies and in two disruptive manufacturing areas (i.e., robotic systems and additive manufacturing) are being developed by NIST [3]. Fig. 7 presents the frame-work of the proposed systematic standards of big data and the IoT in China, which will be the foundations of I3M. It is expected that our proposed principles for constructing the DID code may become a part of these two standard systems. The process of implementing the next-generation science standards is under way, and will change the education and training of our future workforce.

4. Conclusions

This work briefly introduces several perspectives of the digital twin design paradigm of ICME in the age of design. A universal DID format consisting of a set of build chains is proposed, which aligns with the classical form of identifier that is utilized in both international and national standards. The proposed DID is flexible and convenient for extending and sharing in and between various cloud-based platforms. It is worth mentioning that the main purpose in setting up the MGE cloud-based platform is its collaborative applications for discovering novel advanced materials. All of the researchers involved in this platform can work together simultaneously and globally. Our proposed principles for constructing the DID code align with the I3 features and can serve the I3M. Supporting future technologies that contribute to the I3M, including advanced materials, big data analytics, cloud computing, the industrial Internet, and mobile devices, can improve the interactions of HCSs, HPSs, and CPSs. The inheritable feature refers to the original concept of the “material genome” in materials discovery, design, and manufacturing. As our proposed principles for constructing the DID code will act as the fingerprints of the data record in the cloud-based platform for I3M, it is expected that our principles may be involved or combined in standard systems for big data and the IoT in China.

Acknowledgements

This work was financially supported by the National Key Research and Development Program of China (2018YFB0703801, 2018YFB0703802, 2016YFB0701303, and 2016YFB0701304) and CRRC Tangshan Co., Ltd. (201750463031). Special thanks to Professor Hong Wang at Shanghai Jiao Tong University for the fruitful discussions and the constructive suggestions/comments.

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 Weinmin Liu declare that they have no conflict of interest or financial conflicts to disclose.
Fig. 7. Framework of the proposed systematic standards of (a) big data and (b) the IoT in China. ITOM: information technology operations management; RFID: radio frequency identification; QoS: quality of service.
References


