The materials tetrahedron has a "digital twin"

Michael E. Deagen*, L. Catherine Brinson, Richard A. Vaia, and Linda S. Schadler*

For over three decades, the materials tetrahedron has captured the essence of materials science and engineering with its interdependent elements of processing, structure, properties, and performance. As modern computational and statistical techniques usher in a new paradigm of data-intensive scientific research and discovery, the rate at which the field of materials science and engineering capitalizes on these advances hinges on collaboration between numerous stakeholders. Here, we provide a contemporary extension to the classic materials tetrahedron with a dual framework—adapted from the concept of a "digital twin"—which offers a nexus joining materials science and information science. We believe this high-level framework, the materials–information twin tetrahedra (MITT), will provide stakeholders with a platform to contextualize, translate, and direct efforts in the pursuit of propelling materials science and technology forward.

Impact statement

This article provides a contemporary reimagination of the classic materials tetrahedron by augmenting it with parallel notions from information science. Since the materials tetrahedron (processing, structure, properties, performance) made its first debut, advances in computational and informational tools have transformed the landscape and outlook of materials research and development. Drawing inspiration from the notion of a digital twin, the materials–information twin tetrahedra (MITT) framework captures a holistic perspective of materials science and engineering in the presence of modern digital tools and infrastructures. This high-level framework incorporates sustainability and FAIR data principles (Findable, Accessible, Interoperable, Reusable)—factors that recognize how systems impact and interact with other systems—in addition to the data and information flows that play a pivotal role in knowledge generation. The goal of the MITT framework is to give stakeholders from academia, industry, and government a communication tool for focusing efforts around the design, development, and deployment of materials in the years ahead.

Introduction and background

Human technological evolution and innovations in materials share an interwoven history punctuated by advances in tools, instrumentation, and the exchange of knowledge. Today, as digitization and automation drive down the marginal cost of collecting, archiving, and sharing data, methods for extracting value from this abundance of digital information have proliferated. With a new paradigm of data-intensive scientific research and discovery underway,[1] the field of materials science and engineering aims to drastically reduce the overall time and cost to discover and develop new materials through efforts such as the Materials Genome Initiative.[2] While an ecosystem of data repositories and e-collaboration platforms offers numerous tools and resources around materials data,[3–7] facilitating interoperability and integration into the typical materials research workflow remains an ongoing challenge. To broaden the cross-fertilization of solutions across materials sub-fields and information technologies, a shared conceptual framework would enable experts in various subdisciplines to leverage, and contribute to, their combined abilities in order to solve challenges in an integrated, interoperable, open architecture. Fundamentally, we postulate that such a foundational framework lies within the complex interplay between materials and information science. We propose that an extension of the classic materials tetrahedron framework—inspired by the notion of a "digital twin"—provides such a scaffold as the materials–information twin tetrahedra (MITT) (**Figure 1**). Here, we deliver a perspective on the convergence of materials and information science through the lens of this new MITT

Michael E. Deagen, Department of Mechanical Engineering, The University of Vermont, USA; mdeagen@mit.edu
L. Catherine Brinson, Department of Mechanical Engineering and Materials Science, Duke University, NC 27708, USA
Richard A. Vaia, Air Force Research Laboratory, Wright-Patterson Air Force Base, USA
Linda S. Schadler, Department of Mechanical Engineering, The University of Vermont, USA; Linda.Schadler@uvm.edu
*Corresponding author
doi:10.1557/s43577-021-00214-0

© The Author(s) 2022

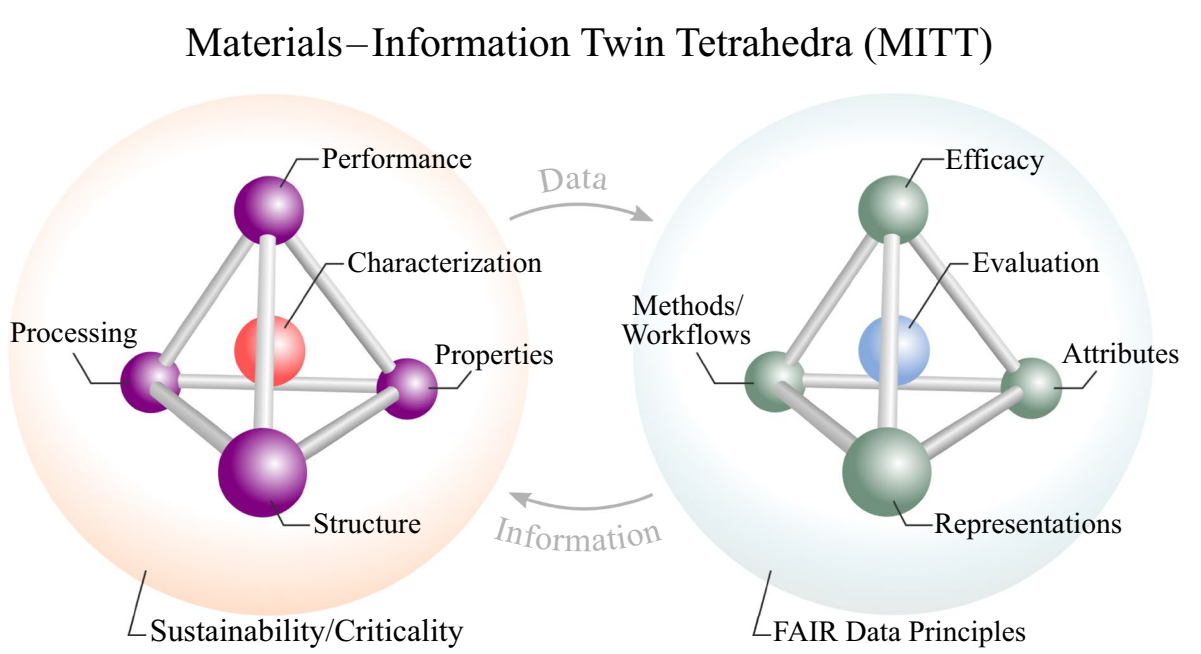

Figure 1. Materials–information twin tetrahedra (MITT) framework translates foundational concepts in materials science and engineering (from the materials tetrahedron) to parallel notions in information science (the "information tetrahedron"), highlighting the data and information flows that form a closed-loop for knowledge creation around the discovery, design, development, and deployment of materials.

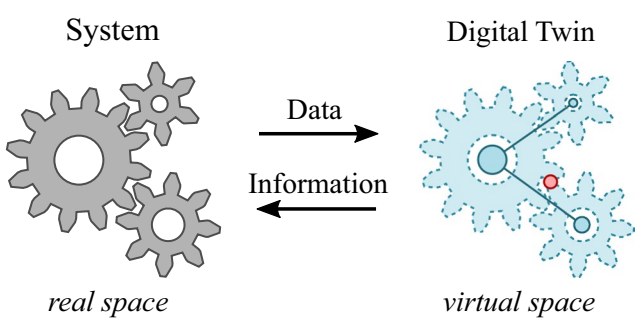

Figure 2. A digital twin comprises a virtual representation of a real system, linked by continual data and information flows throughout the system's life cycle.

framework, and in this perspective article we highlight select works and recent reviews as illustrative examples in this emerging and evolving space.

Since its inception, the materials tetrahedron has remained an enduring visual icon that illustrates the interdependence of key concepts in materials science and engineering. Published in 1989 by the National Research Council in its report on materials science and engineering for the 1990s,[8] the symbolic polyhedron depicts four foundational elements of materials science and engineering while emphasizing the edges that connect them. These four elements—processing, structure, properties, and performance—have also been arranged linearly as a three-link chain to highlight the forward "cause and effect" progression from processing to performance as well as its inverse "goal/means" counterpart.[9,10] In 2008, a depiction of the materials tetrahedron with characterization as an interstitial element at its center was added to the public domain by means of Wikipedia.[11] In 2019, Donahue made the case for appending the dimension of sustainability/criticality to the materials tetrahedron for a holistic perspective on the roles that materials play in the anthroposphere.[12] Donahue's reimagination of the materials tetrahedron paradigm extends beyond individual use cases of materials by recognizing factors such as raw material supply risk, price volatility, and the environmental implications of materials in their full life cycle.[†] At its essence, sustainability/criticality addresses the extent to which our materials systems, throughout their life cycle, impact and interact with other systems.

Over the past two decades, a concept has emerged from the manufacturing community that elegantly bridges the gap between physical systems and our virtual representations of them. Originally described by Grieves in the early 2000s as product life cycle management (PLM),[13] the term "digital twin" was coined by Vickers et al. in 2010 in a NASA technology area roadmap for materials, structures,

[†]While Donahue replaces the materials tetrahedron with a square pyramid depiction, we have taken the artistic license to represent the aspect of Sustainability/Criticality as an encapsulating sphere to lend a "global" visual cue and preserve the recognizability of the classic materials tetrahedron.

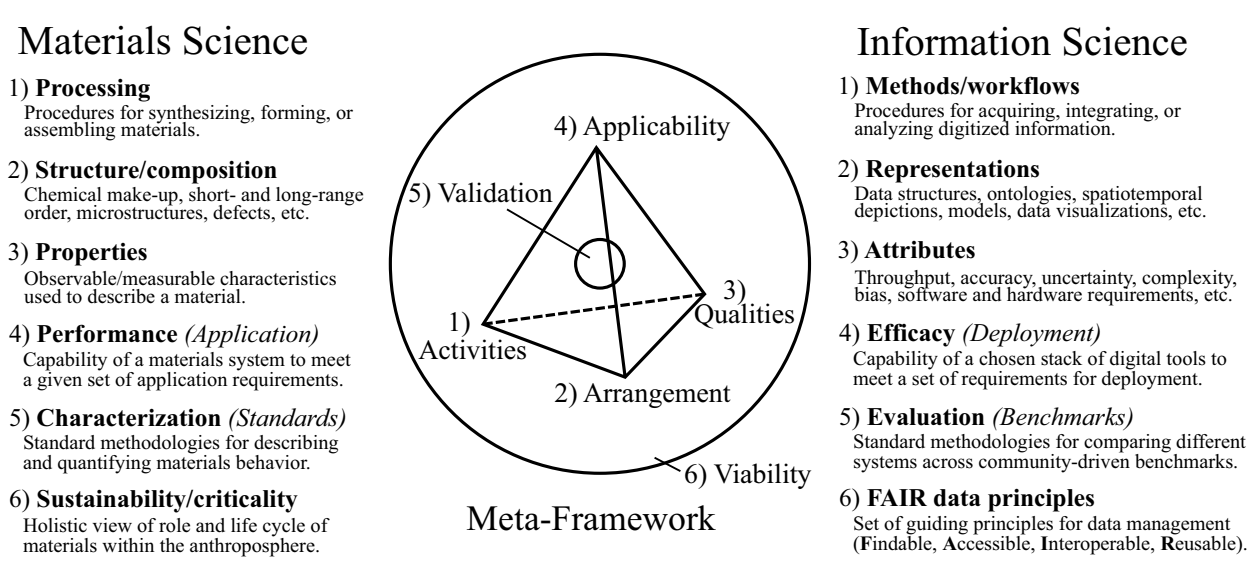

Figure 3. An underlying meta-framework captures the elements of the (extended) materials tetrahedron and relates these elements to counterparts in information science.

mechanical systems and manufacturing.[14] The digital twin—a virtual representation of a system that exists alongside its physical counterpart throughout its life cycle—includes (1) the sensor-equipped physical object or system itself; (2) a high-fidelity virtual representation, including historical data and predicted future performance; and (3) the data and information flows between the physical world and the digital twin, often through multi-modal sensing and probabilistic modeling at the instance and aggregate level (**Figure 2**).[15,16] The value proposition of a digital twin stems from its ability to manage and apply heterogeneous information, and the concept has gained significant traction in recent years as evidenced by citations in the literature.[17,18] Manifestations of digital twins have appeared in manufacturing,[19–21] aerospace,[15,22] healthcare,[23–25] transportation,[26–29] and the built environment including utilities and smart cities.[30–35] The concept will continue to gain traction as stakeholders adopt common data standards and models, improve data sharing practices, develop products and services around digital twins, and establish forums for experts in various disciplines (including materials science) to meet and collaborate.[36]

When we refer to a "digital twin" for the materials tetrahedron—a purely conceptual framework, as opposed to a physical system— we describe its interdependent counterpart in information science, hereby referred to as the "information tetrahedron." Given the data and information flows that connect these conceptual frameworks and provide the basis for knowledge creation, the digital twin analogy remains apropos. For reference, the terms data, information, and knowledge represent levels in the *data-information-knowledge-wisdom* (DIKW) hierarchy.[37]

In this perspective, we offer the MITT framework as a means to contextualize the interdisciplinary efforts across various stakeholders in furtherance of a robust infrastructure and workforce around materials data. In the following sections, we articulate the meta-framework underpinning MITT, describe the components of the information tetrahedron, and illustrate one possible application of the dual framework with existing and upcoming technologies. We believe that students, practitioners, educators, and policymakers armed with this framework can apply modern and future digital information capabilities to address grand materials challenges with the high-level perspective that has guided and benefited the field of materials science and engineering for more than three decades.

Translating the materials tetrahedron to information science

In a prototypical implementation of the MITT framework, materials systems provide data to information systems, which in turn generate information that guide the further improvement of these materials systems. Iteratively, this cycle aims for knowledge creation around discovery, design, development, and deployment of materials systems. Humans necessarily remain in the loop in terms of strategic guidance and implementation of this process, but the reciprocal exchange of data and information should take advantage of available automated workflows.

To define the information tetrahedron, we identified six dimensions underpinning the materials tetrahedron and translated these dimensions to notions from information science (**Figure 3**). These dimensions—activities, arrangement,

Table I. Contextualized examples, including recent progress in materials data and informatics, related to each element of the information tetrahedron of the MITT framework.

Activities	*Arrangement*	*Qualities*	*Applicability*	*Validation*	*Viability*
Methods/ Workflows	Representations	Attributes	Efficacy	Evaluation	FAIR Data Principles[39]
Inverse design[40] Multiphysics simulations[41] Autonomous experiments[42–44] Interpretable ML methods[45,46] Open-source toolkits[47–49] Correlative characterization[50] Mixed-initiative user interaction[51]	Atomic or molecular data structures[52,53] Spatiotemporal depictions (pixelated, voxelated, graph-based) Physical descriptors[54] Schemas, taxonomies, controlled vocabularies[55,56] Workflow representations[57] Ontologies[58] Low-dimensional embeddings Data visualizations	Complexity Throughput Accuracy Bias Uncertainty Usability Software dependencies Hardware requirements Cost	Clearly defined scope and requirements Extent to which system meets requirements Suitability of system for the task at hand Time and cost savings over alternatives	Benchmark data sets and tasks[59,60] Objective tests and measures for comparison Metrics for data "FAIRness" UI/UX assessment Validation of predictions	Findable Accessible Interoperable Reusable AI-ready[61] Sustained life cycle efficacy

qualities, applicability, validation, and viability—form a generalizable meta-framework at the heart of the proposed MITT framework.

The meta-framework classifies the various aspects of these frameworks broadly, such that the specification and implementation of instantiations of this meta-framework may follow the best practices of their respective fields. Thus, the meta-framework enables both the materials tetrahedron and information tetrahedron to describe individual aspects of each domain without relying on one-to-one mappings. For example, the dimension "Arrangement" refers to materials structure and composition in the materials domain and to digital representations including data and metadata structures in the information domain. The generalizability of this meta-framework may enable the tetrahedron concept to percolate into other domains, providing a useful organizational and translational tool. In the following section, we expound further on the components of the information tetrahedron of the MITT framework, focusing on recent progress in materials data and informatics.

Components of the information tetrahedron

Strictly speaking, a digital twin represents a virtual instantiation of a physical system. To translate the essence of a "digital twin" to the materials tetrahedron—a purely conceptual framework to begin with—we describe the MITT framework as a paradigm that considers both materials science and information science side-by-side, connected by data and information flows. Development of materials data and informatics systems involves many complex tradeoffs and nuances that accompany the various strategic and design decisions made by software architects and engineers.[38] Striving for a general-purpose digital information system while managing scope to stay within development constraints mirrors many of the challenges faced by materials engineers in designing or optimizing materials systems. **Table I** summarizes the following paragraphs, which relate the information science components of the MITT paradigm to recent progress in materials data and informatics.

Methods/workflows

Acceleration of materials discovery and development presupposes the accompanying digital methods and workflows for collecting, curating, integrating, and analyzing data. Methods and workflows may include algorithms for achieving "inverse design" of processing parameters given a set of performance goals,[40] multiphysics simulations,[41] high-throughput methods that automate or semi-automate experimentation and data collection,[42–44] interpretable machine learning methods,[45,46] retrieval and management of large data sets facilitated by open-source toolkits,[47–49] methods for bridging length scales and imaging modalities via correlative characterization,[50] or mixed-initiative user interfaces that leverage automation to support better decision-making through human–computer interaction.[51] When applying the MITT paradigm to a materials data and informatics, one should consider how these methods and workflows combine to transform fragmented data into actionable information.

Representations

As symbolic representations of information, data derive their utility only when placed within broader models or contexts that impart meaning. In the absence of a preexisting model of the world, machines do not have the innate capacity to interpret data without some form of metadata ("data describing the data"), necessitating efforts from the community to develop and adopt common standards around digital representations for materials data and metadata. These representations may include: data structures for atomic or molecular arrangements;[52,53] pixelated, voxelated, or graph-based spatiotemporal depictions at various hierarchical scales; quantitative physical descriptors;[54] schemas, taxonomies, and controlled vocabularies;[55,56] or

standardized representations of materials processing or computational workflows.[57] Metadata structures include ontologies,[58] which comprise machine-interpretable networks of linked concepts, as well as structured representations of data provenance. Data that are inherently uninterpretable (e.g., a trained neural network model) may have metadata describing their inputs and outputs or low-dimensional embeddings serving as representations. To display collected or analyzed data, advances in interactive data visualization can provide an interface between these systems and human decision-makers. When applying the MITT framework, one should carefully consider (meta)data representations as they relate to deployment requirements, with a preference for representations aligned with the FAIR data principles (findable, accessible, interoperable, reusable).[39]

Attributes

The attributes of a digital information system should provide an objective view of the system and its standing among comparable systems. These attributes include technical specifications of the software and hardware, as well as software libraries or data corpora that these systems rely on. Complexity, throughput, accuracy, bias, uncertainty, usability, and other relevant system attributes should be accompanied by community-driven benchmarks that enable researchers to objectively measure them so that system architects and end users can select the best stack of digital tools given their budget and scope. In the context of MITT, attributes reflect the various tradeoffs of particular methods, workflows, and representations at the individual and systems level.

Efficacy

The efficacy of an information system rests on its architecture, integration, and environment in the context of a given deployment setting. In addition to the system's internal ability to process and analyze data, efficacy dictates how well the system can interface with the real world by incorporating new data from external signals and presenting actionable information to decision-makers. Tradeoffs between system components must be considered when optimizing system efficacy.

Evaluation

Community-driven standards and procedures for objective validation of digital information systems lie at the heart of the information tetrahedron. Benchmark data sets and tasks facilitate consistent and systematic evaluation of the myriad digital information methods and systems that will emerge in the coming years.[59,60] Evaluation protocols and benchmarks for FAIR data management as well as human usability should be considered among these assessments.

FAIR data principles

Conversations around the life cycle and impact of data and information remain pivotal in order to ensure continual growth and reuse of information resources. The four guiding principles for data management and stewardship that comprise FAIR (findable, accessible, interoperable, reusable) recognize the need among stakeholders in academia, industry, funding agencies, and publishers for an infrastructure that maintains the value of data beyond initial publication.[39] Ongoing, community-wide commitments to FAIR data and metadata standards aim to reduce the labor-intensive data pre-processing and cleaning steps that historically kick off data science projects, making data "AI-ready."[61] Some data may remain proprietary or confidential, but any data advertised as "open" should contain adequate metadata. In addition to mandates or incentive structures,[62] the prevalence of FAIR data depends on automated methods and workflows that enable researchers to seamlessly manage and share their data. As a supplement to the FAIR principles for data, materials data and informatics platforms should incorporate configuration management activities to maintain continued efficacy and integrity throughout the projected life cycle of the platform. When considering an information system's viability, one must also take into account internal factors such as reliance on specific software and hardware infrastructures in addition to external factors such as funding agency mandates and new expectations driven by transformations in data culture.

Highlighting recent progress and reviews

The topics of informatics, data science, and machine learning have appeared with increasing prevalence in materials-related works in the academic literature—most notably in the past decade—as reflected in the quantity of publications and citations (**Figure 4**). This exponential growth persists even when normalized by the steady annual growth of all materials-related publications. The progression shows the extent to which the field of materials science has evolved in the three decades since the initial conception of the materials tetrahedron in 1989 and the impact of integration and digitization on the landscape of materials research and development. For example, the availability and capability of computational tools initiated the discipline of integrated computational materials engineering (ICME),[63] combinatorial materials synthesis and rapid characterization techniques led to high-throughput experimental (HTE) methodologies,[64] and model-based concurrent design, development, and deployment for materials and systems form the basis of NASA's recent "Vision 2040" roadmap.[65]

Systems for autonomous experimentation (AE) exemplify the closed-loop coupling of materials science and information science as laid out in a recent community perspective article by Stach et al.[66] Owing to advances in mechanical automation and robotics, rote tasks that often consume researcher time and effort can be delegated to machines that prepare, characterize, and test samples. Aided by machine learning, collected data inform future hypothesis testing for efficient navigation of complex, high-dimensional design spaces. Researchers may instead focus on higher-level

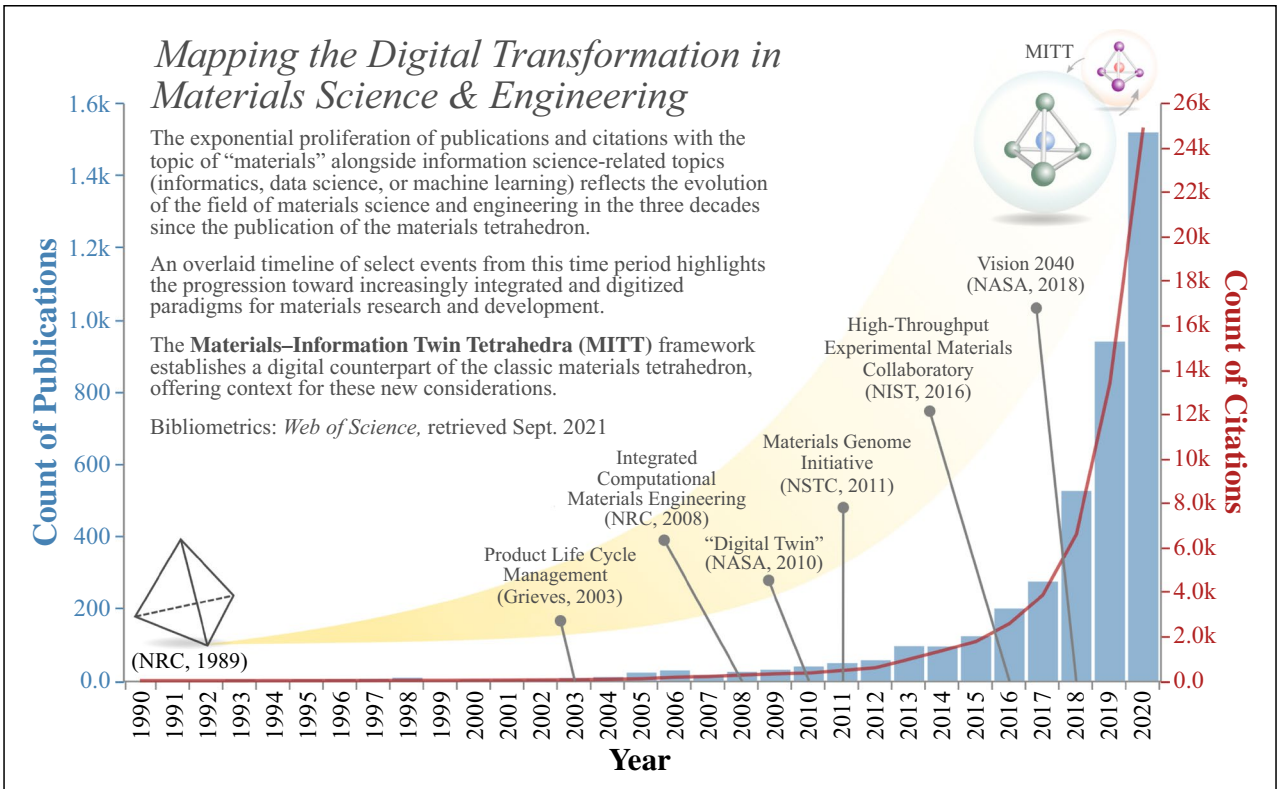

Figure 4. Bibliometric data from *Web of Science* show the count of publications (blue bars) and citations (red line) at the intersection of the topic of "materials" with any of the topics of "informatics," "data science," or "machine learning" in the years 1990–2020. The timeline highlights select examples from this progression toward increased integration and digitization in materials research and development.

guidance such as providing insights, context, and objectives to the AE campaign. Stach et al. describe the benefits of AE in terms of multiplying the productivity of individual researchers in addition to network effects when these distributed systems share learned insights. Importantly, they present areas for investment in hardware, software, data management, and workforce education in the coming years to realize the potential of AE systems for materials research and development. Similar opportunities and challenges for closed-loop automation exist outside the field of materials science as well, such as drug discovery,[67] healthcare,[68] supply chain management,[69] architecture,[70] and chemistry.[71,72]

Although an exhaustive review falls outside the scope of this perspective, we urge the interested reader to consult recent review articles and perspectives that explore the intersection of materials science with one or more of the topics in information science outlined in this perspective. These resources include discussions of data ecosystems and infrastructures for metadata management;[73–75] high-throughput library generation and characterization;[76,77] successes and challenges of Materials Genome Initiative research, machine learning methods, and computational materials databases;[78–80] methods for linking materials characterization and computation across length scales;[81,82] methods, representations, and applications in the field of polymer informatics;[83] materials discovery for energy applications;[84] and integrated systems for next-generation microscopy.[85]

Demonstrating the value of materials data and informatics

Just as the classic materials tetrahedron does not fully capture, for example, the subtleties of precipitation hardening of aluminum alloys, the MITT framework does not claim to solve the challenges ahead in creating a robust infrastructure around materials data and informatics. Instead, the framework presents a holistic view of certain high-level, interdependent elements such that experts across disciplines can make this infrastructure a reality. The broader materials community will accept and adopt materials data and informatics resources as these resources demonstrate more efficient digital representations and methods, offer robust cyber-physical infrastructures that establish trust, and provide training of the workforce in the utilization of these tools.[86] Making these systems and tools relevant to the broader materials community will require clear articulation of the value proposition of these resources, continued investment in the areas laid out in the MITT framework, and the refinement of shared visions for the future of materials research and development.

Data and informatics have the potential to transform the conventional landscape of materials research and development. By extrapolating from existing technologies, one can speculate on manifestations of the materials research landscape in the near future:

> *...A researcher monitors the incoming data on her tablet. In the background, an autonomous research system chirps and hums as it passes cassettes of samples between preparation stages for characterization. She, along with a geographically distributed network of colleagues, designs and develops materials systems for electrically insulating coatings—her materials system of choice features functional molecules tethered to nanoscopic ceramic particles and dispersed in self-healing polymers. The boundless configurations of possible materials components make exploration of the design space both invigorating and daunting. To manage the complex design tradeoffs, she and her team develop methods that consolidate computational and experimental research—augmented with troves of data from several online data repositories through APIs—to predict the electrical, mechanical, thermal, and degradation behavior of these materials systems. To validate and improve model predictions along the Pareto frontier, she runs combinatorial experiments with gradient libraries that vary mix ratios and process parameters to reveal and quantify the effects on nanoparticle dispersion in these composite systems. On some days, carrying out an experiment involves a few simple gestures on her tablet—unified data formats enable the autonomous research system to interoperate with the workflow planning system, lab inventory, characterization and image analysis tools, and her team's Bayesian models—and each material sample receives a globally unique identifier with semantic links to its detailed processing history and characterization results. As the data populate her customized dashboard of interactive charts in real-time, she recalls stories of data antiquity—misplaced USB drives, critical experimental parameters scrawled in the margins of laboratory notebooks, and hours spent manually aggregating data from PDF files into spreadsheets. When she considers how she cites and publishes work today, everything just feels a lot more—what's the right word?* Natural, *she mutters to herself. With a few touches on her tablet, she pushes the newly acquired data to her team's shared knowledge base and alerts the subscribers to her data streams who may review and independently verify the results. If this particular materials system shows promise as a viable electrical insulator—and even if it does not—the data collected today will contribute to the growing online research data repositories from which her team developed their first model, a small act of solidarity in the modern scientific endeavor...*

Table II elaborates upon this hypothetical scenario in the context of the MITT framework. Achieving such visions will require persistent, well-coordinated efforts among a variety of stakeholders including materials researchers, software developers, original equipment manufacturers, and funding agencies. As a prerequisite, connecting systems-level expertise across disciplines requires shared language and communication tools. Domain-specific terminology and jargon tend to form lexical barriers that hinder cross-disciplinary collaboration and may obscure higher-level interdisciplinary commonalities. The classic materials tetrahedron paradigm presents a cogent visual depiction of the field of materials science, and revealing the underlying meta-framework enables the translation of this paradigm to information science.

As a general approach, the meta-framework could apply to other systems-focused disciplines that couple with materials science to form closed feedback loops. For example, one can consider the coupling of materials science and bioengineering that occurs in bio-integrated materials.[87] At a high-level, the systems-level meta-framework would apply to biomaterials engineering through consideration of activities (e.g., pathways), arrangement (e.g., geometry, organization), qualities (e.g., characteristics, traits), applicability (e.g., goodness of fit), validation (e.g., bioanalysis), and viability (e.g., longevity, compatibility) in terms of the biological system of interest. Instead of digital data and information flows, these interfaces might be mediated by sensing and modulation via biophysical or biochemical signals and actuation. However, we concentrate here on the convergence of materials and information science, leaving other combinatorial examples of joint frameworks open to future discussions.

With its high-level vantage point on materials and information science, the MITT framework provides a pedagogical launchpad and a groundwork for cross-disciplinary communication pertaining to infrastructure built around materials data. Faced with troves of data to potentially analyze, the challenge becomes how to organize researcher attention and computational resources around extraction of high-value information and the ultimate translation of this information into lasting solutions. The MITT framework extends the systems-oriented paradigm of the classic materials tetrahedron to encourage proficiency in tools that augment and accelerate materials discovery, design, development, and deployment through data science and informatics. By applying this framework to contextualize, translate, and direct various efforts in materials data and informatics, present and future generations of materials scientists and engineers will find themselves well-equipped to tackle the multifold challenges that arise in an increasingly complex, data-driven world.

Acknowledgments

The authors gratefully acknowledge support from the NSF CSSI program (OAC-1835648, OAC-1835782), NSF DMREF program (CMMI-1818574, CMMI-1729743,

Table II. To illustrate one of many possible applications of the MITT framework, aspects of the above narrative, specifically to the design of nanoparticle-polymer material systems for electrically insulating coatings, have been organized into the framework's various dimensions.

Dimensions of the meta-framework					
Activities	*Arrangement*	*Qualities*	*Applicability*	*Validation*	*Viability*
Materials tetrahedron					
Processing	Structure	Properties	Performance	Characterization	Sustainability/criticality
Nanoparticle synthesis Particle surface modification Solution mixing Twin-screw extrusion Serial sectioning	Volume fraction Dispersion Interfacial area Graft density Entanglement Charge distribution	Dielectric breakdown strength Dielectric constant Dielectric loss Glass transition temperature Melting temperature Melt viscosity Charge mobility	Integration into encapsulation for high-voltage power transmission Reliability via enhanced charge trapping and self-healing mechanisms	X-ray diffraction (XRD), rheometry, thermogravimetric analysis (TGA), dynamic mechanical analysis (DMA), differential scanning calorimetry (DSC), broadband dielectric spectroscopy (BDS), electrostatic force microscopy (EFM), transmission electron microscopy (TEM), pulsed electro-acoustic (PEA) measurement	Designed recyclability and repurposing Assessed value of material selection within system design Life cycle and cost analysis Prognostics and prediction of performance lifetime Raw material traceability
Information tetrahedron					
Methods/workflows	Representations	Attributes	Efficacy	Evaluation	FAIR data principles
High-throughput experimentation Gradient libraries Image analysis Multi-scale simulations Bayesian optimization Data streams from APIs	Molecular data structures Binarized TEM images Voxelated reconstructions Stochastic models Processing, structure, characterization, and provenance metadata linked by ontology Interactive data visualization	Bias and uncertainty estimation Open-source software dependencies IoT-integrated hardware Data security and verifiability Intuitive user interface design	N-fold increase in research data output compared to traditional design space exploration Reduced time to integrate and scale up manufacturing	Validation of models through experiment Lab equipment integration Comparison against shared benchmarks Human–computer interaction assessment	Persistent identifiers Tiered access, data security Lab equipment, inventory, modeling DBs, analysis tools interoperate Reconfigurable, agile network of nodes Open-access, open-source integration

CMMI-1729452), ongoing collaborations with NIST, CHiMaD (NIST 70NANB19H005) at Northwestern University, and support from the Air Force Research Laboratory, Materials and Manufacturing Directorate.

Author contributions

Following an initial conversation between M.E.D. and L.S.S., M.E.D. developed the manuscript for this perspective with significant guidance and feedback from L.C.B., R.A.V., and L.S.S.

Conflict of interest

The authors declare no competing financial interests.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0.

References

1. T. Hey, S. Tansley, K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, Redmond, WA, 2009)
2. National Science and Technology Council (US), "Materials Genome Initiative for Global Competitiveness" (Executive Office of the President, National Science and Technology Council, Washington, DC, 2011)
3. J. Hill, G. Mulholland, K. Persson, R. Seshadri, C. Wolverton, B. Meredig, Materials science with large-scale data and informatics: Unlocking new opportunities. *MRS Bull.* **41**(5), 399 (2016)
4. S. Ramakrishna, T.-Y. Zhang, W.-C. Lu, Q. Qian, J.S.C. Low, J.H.R. Yune et al., Materials informatics. *J. Intell. Manuf.* **30**(6), 2307 (2018)
5. K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. *Nature* **559**(7715), 547 (2018)
6. L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: Status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019)
7. A. Lopez-Bezanilla, P.B. Littlewood, Growing field of materials informatics: Databases and artificial intelligence. *MRS Commun.* **10**(1), 1 (2020)
8. Solid State Sciences Committee and National Research Council, *Materials Science and Engineering for the 1990s: Maintaining Competitiveness in the Age of Materials* (National Academies Press, Washington, DC, 1989)
9. G.B. Olson, Computational design of hierarchically structured materials. *Science* **277**(5330), 1237 (1997)
10. G.B. Olson, Designing a new material world. *Science* **288**(5468), 993 (2000)
11. User: Dhatfield, Materials science tetrahedron. Wikimedia Commons (2008). https://commons.wikimedia.org/wiki/File:Materials_science_tetrahedron;structure,_processing,_performance,_and_proprerties.svg. Accessed 29 Mar 2021
12. C.J. Donahue, Reimagining the materials tetrahedron. *J. Chem. Educ.* **96**(12), 2682 (2019)
13. M. Grieves, PLM-beyond lean manufacturing. *Manuf. Eng.* **130**(3), 23 (2003)
14. R. Piascik, J. Vickers, D. Lowry, S. Scotti, J. Stewart, A. Calomino, *Technology Area 12: Materials, Structures, Mechanical Systems, and Manufacturing Road Map* (NASA Office of Chief Technologist, Washington, DC, 2010)
15. E.J. Tuegel, A.R. Ingraffea, T.G. Eason, S.M. Spottswood, Reengineering aircraft structural life prediction using a digital twin. *Int. J. Aerosp. Eng.* **2011**, 154798 (2011)
16. E. Negri, L. Fumagalli, M. Macchi, A review of the roles of digital twin in CPS-based production systems. *Procedia Manuf.* **11**, 939 (2017)
17. F. Tao, H. Zhang, A. Liu, A.Y.C. Nee, Digital twin in industry: State-of-the-art. *IEEE Trans. Industr. Inform.* **15**(4), 2405 (2019)
18. D. Jones, C. Snider, A. Nassehi, J. Yon, B. Hicks, Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **29**, 36 (2020)
19. R. Söderberg, K. Wärmefjord, J.S. Carlson, L. Lindkvist, Toward a Digital Twin for real-time geometry assurance in individualized production. *CIRP Ann.* **66**(1), 137 (2017)
20. F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, F. Sui, Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Technol.* **94**(9–12), 3563 (2017)
21. G. Zhou, C. Zhang, Z. Li, K. Ding, C. Wang, Knowledge-driven digital twin manufacturing cell towards intelligent manufacturing. *Int. J. Prod. Res.* **58**(4), 1034 (2019)
22. E. Glaessgen, D.S. Stargel, *53rd Structures, Structural Dynamics, and Materials Conference* (American Institute for Aeronautics and Astronautics, Honolulu, April 23–26, 2012), p. 1818
23. K. Bruynseels, F. Santoni de Sio, J. van den Hoven, Digital Twins in health care: Ethical implications of an emerging engineering paradigm. *Front. Genet.* **9**, 31 (2018)
24. Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, M. Jamal Deen, A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access* **7**, 49088 (2019)
25. B. Björnsson, C. Borrebaeck, N. Elander, T. Gasslander, D.R. Gawel, M. Gustafsson, R. Jörnsten, E.J. Lee, X. Li, S. Lilja, D. Martínez-Enguita, A. Matussek, P. Sandström, S. Schäfer, M. Stenmarker, X.F. Sun, O. Sysoev, H. Zhang, M. Benson, Digital twins to personalize medicine. *Genome Med.* **12**(1), 4 (2020)
26. M. Baumann, S. Rohr, M. Lienkamp, 2018 *2018 Thirteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)* (Monte Carlo, Monaco, April 10–12, 2018), pp. 1–6
27. A. Rassõlkin, T. Vaimann, A. Kallaste, V. Kuts, *2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)* (Riga, Latvia, October 7–9, 2019), p. 1
28. L. Merkle, A.S. Segura, J.T. Grummel, M. Lienkamp, *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)* (Taipei, Taiwan, May 5–9, 2019), p. 155
29. Z. Wang, X. Liao, X. Zhao, K. Han, P. Tiwari, M.J. Barth, G. Wu, *2020 IEEE 91st Vehicular Technology Conference* (*VTC2020*-Spring) (Antwerp, Belgium, May 25–28, 2020), p. 1
30. C. Brosinsky, D. Westermann, R. Krebs, *2018 IEEE International Energy Conference (ENERGYCON)* (Limassol, Cyprus, June 3–7, 2018), p. 1
31. T. Ruohomäki, E. Airaksinen, P. Huuska, O. Kesäniemi, M. Martikka, J. Suomisto, *2018 International Conference on Intelligent Systems (IS)* (Funchal, Portugal, September 25–27, 2018), p. 155
32. P. Conejos Fuertes, F. Martínez Alzamora, M. Hervás Carot, J. Alonso ampos, Building and exploiting a Digital Twin for the management of drinking water distribution networks. *Urban Water J.* **17**(8), 704 (2020)
33. M. Farsi, A. Daneshkhah, A. Hosseinian-Far, H. Jahankhani, *Digital Twin Technologies and Smart Cities* (Springer, Cham, Switzerland, 2020)
34. S. Shirowzhan, W. Tan, S.M.E. Sepasgozar, Digital Twin and CyberGIS for improving connectivity and measuring the impact of infrastructure construction planning in smart cities. *ISPRS Int. J. Geo-Inf.* **9**(4), 240 (2020)
35. C. Fan, Y. Jiang, A. Mostafavi, Social sensing in disaster city Digital Twin: Integrated textual–visual–geo framework for situational awareness during built environment disruptions. *J. Manag. Eng.* **36**(3), 04020002 (2020)
36. F. Tao, Q. Qi, Make more digital twins. *Nature* **573**, 490 (2019)
37. J. Rowley, The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* **33**(2), 163 (2007)
38. M. Aykol, J.S. Hummelshøj, A. Anapolsky, K. Aoyagi, M.Z. Bazant, T. Bligaard, R.D. Braatz, S. Broderick, D. Cogswell, J. Dagdelen, W. Drisdel, E. Garcia, K. Garikipati, V. Gavini, W.E. Gent, L. Giordano, C.P. Gomes, R. Gomez-Bombarelli, C. Balaji Gopal, J.M. Gregoire, J.C. Grossman, P. Herring, L. Hung, T.F. Jaramillo, L. King, H.-K. Kwon, R. Maekawa, A.M. Minor, J.H. Montoya, T. Mueller, C. Ophus, K. Rajan, R. Ramprasad, B. Rohr, D. Schweigert, Y. Shao-Horn, Y. Suga, S.K. Suram, V. Viswanathan, J.F. Whitacre, A.P. Willard, O. Wodo, C. Wolverton, B.D. Storey, The materials research platform: Defining the requirements from user stories. *Matter* **1**(6), 1433 (2019)
39. M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016)
40. A. Zunger, Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2**(4), 0121 (2018)
41. D.E. Keyes, L.C. McInnes, C. Woodward, W. Gropp, E. Myra, M. Pernice et al., Multiphysics simulations. *Int. J. High Perform. Comput. Appl.* **27**(1), 4 (2013)

42. P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, B. Maruyama, Autonomy in materials research: A case study in carbon nanotube growth. *npj Comput. Mater.* **2**(1), 16031 (2016)
43. D.P. Tabor, L.M. Roch, S.K. Saikin, C. Kreisbeck, D. Sheberla, J.H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C.J. Brabec, B. Maruyama, K.A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**(5), 5 (2018)
44. J.H. Montoya, K.T. Winther, R.A. Flores, T. Bligaard, J.S. Hummelshøj, M. Aykol, Autonomous intelligent agents for accelerated materials discovery. *Chem. Sci.* **11**(32), 8517 (2020)
45. J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**(1), 83 (2019)
46. S.M. Moosavi, K.M. Jablonka, B. Smit, The role of machine learning in the understanding and design of materials. *J. Am. Chem. Soc.* **142**(48), 20273 (2020)
47. L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60 (2018)
48. C. Draxl, M. Scheffler, NOMAD: The FAIR concept for big data-driven materials science. *MRS Bull.* **43**(9), 676 (2018)
49. D. Mrdjenovich, M.K. Horton, J.H. Montoya, C.M. Legaspi, S. Dwaraknath, V. Tshitoyan, A. Jain, K.A. Persson, Propnet: A knowledge graph for materials science. *Matter* **2**(2), 464 (2020)
50. T.L. Burnett, P.J. Withers, Completing the picture through correlative characterization. *Nat. Mater.* **18**(10), 1041 (2019)
51. E. Horvitz, "Principles of Mixed-Initiative User Interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, PA, 1999), p. 159
52. S.R. Hall, F.H. Allen, I.D. Brown, The crystallographic information file (CIF): A new standard archive file for crystallography. *Acta Crystallogr. A* **47**(6), 655 (1991)
53. T.S. Lin, C.W. Coley, H. Mochigase, H.K. Beech, W. Wang, Z. Wang, E. Woods, S.L. Craig, J.A. Johnson, J.A. Kalow, K.F. Jensen, B.D. Olsen, BigSMILES: A structurally-based line notation for describing macromolecules. *ACS Cent. Sci.* **5**(9), 1523 (2019)
54. L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **114**(10), 105503 (2015)
55. B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, I. Foster, The materials data facility: Data services to advance materials science research. *JOM* **68**(8), 2045(2016)
56. H. Zhao, Y. Wang, A. Lin, B. Hu, R. Yan, J. McCusker, W. Chen, D.L. McGuinness, L. Schadler, L.C. Brinson, NanoMine schema: An extensible data representation for polymer nanocomposites. *APL Mater.* **6**(11), 111108 (2018)
57. M. Govoni, M. Munakami, A. Tanikanti, J.H. Skone, H.B. Runesha, F. Giberti, J. de Pablo, G. Galli, Qresp, a tool for curating, discovering and exploring reproducible scientific papers. *Sci. Data* **6**, 190002 (2019)
58. L. Takahashi, K. Takahashi, Visualizing scientists' cognitive representation of materials data through the application of ontology. *J. Phys. Chem. Lett.* **10**(23), 7482 (2019)
59. A. Dunn, Q. Wang, A. Ganose, D. Dopp, A. Jain, Benchmarking materials property prediction methods: The Matbench test set and automatminer reference algorithm. arXiv preprint (2020). http://arxiv.org/2005.00707
60. C.L. Clement, S.K. Kauwe, T.D. Sparks, Benchmark AFLOW data sets for machine learning. *Integr. Mater. Manuf. Innov.* **9**(2), 153 (2020)
61. B. Mons, E. Schultes, F. Liu, A. Jacobsen, The FAIR principles: First generation implementation choices and challenges. *Data Intell.* **2**(1–2), 1 (2020)
62. B. Mons, Invest 5% of research funds in ensuring data are reusable. *Nature* **578**(7796), 491 (2020)
63. National Research Council, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security* (National Academies Press, Washington, DC, 2008)
64. M.L. Green, C.L. Choi, J.R. Hattrick-Simpers, A.M. Joshi, I. Takeuchi, S.C. Barron, E. Campo, T. Chiang, S. Empedocles, J.M. Gregoire, A.G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. Van Duren, A. Zakutayev, Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Appl. Phys. Rev.* **4**(1), 011105 (2017)
65. X. Liu, D. Furrer, J. Kosters, J. Holmes, *Vision 2040: A Roadmap for Integrated, Multiscale Modeling and Simulation of Materials and Systems* (NASA/CR—2018-219771, NASA STI Program, Hampton, VA, 2018)
66. E. Stach, B. DeCost, A.G. Kusne, J. Hattrick-Simpers, K.A. Brown, K.G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, Autonomous experimentation systems for materials development: A community perspective. *Matter* **4**(9), 2702 (2021)
67. P. Schneider, W.P. Walters, A.T. Plowright, N. Sieroka, J. Listgarten, R.A. Goodnow Jr., J. Fisher, J.M. Jansen, J.S. Duca, T.S. Rush, M. Zentgraf, J.E. Hill, E. Krutoholow, M. Kohler, J. Blaney, K. Funatsu, C. Luebkemann, G. Schneider, Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug. Discov.* **19**(5), 353 (2020)
68. J. Li, J.Y. Liang, S.J. Laken, R. Langer, G. Traverso, Clinical opportunities for continuous biosensing and closed-loop therapies. *Trends Chem.* **2**(4), 319 (2020)
69. G. González Rodríguez, J.M. Gonzalez-Cava, J.A. Méndez Pérez, An intelligent decision support system for production planning based on machine learning. *J. Intell. Manuf.* **31**(5), 1257 (2019)
70. M. Bechthold, J.C. Weaver, Materials science and architecture. *Nat. Rev. Mater.* **2**(12), 17082 (2017)
71. R.W. Epps, M.S. Bowen, A.A. Volk, K. Abdel-Latif, S. Han, K.G. Reyes, A. Amassian, M. Abolhasani, Artificial chemist: An autonomous quantum dot synthesis bot. *Adv. Mater.* **32**(30), e2001626 (2020)
72. M. Christensen, L.P.E. Yunker, F. Adedeji, F. Häse, L.M. Roch, T. Gensch, G. dos Passos Gomes, T. Zepel, M.S. Sigman, A. Aspuru-Guzik, J.E. Hein, Data-science driven autonomous process optimization. *Commun. Chem.* **4**(1), 112 (2021)
73. B. Blaiszik, L. Ward, M. Schwarting, J. Gaff, R. Chard, D. Pike, K. Chard, I. Foster, A data ecosystem to support machine learning in materials science. *MRS Commun.* **9**(4), 1125 (2019)
74. A. Medina-Smith, C.A. Becker, R.L. Plante, L.M. Bartolo, A. Dima, J.A. Warren, R.J. Hanisch, A controlled vocabulary and metadata schema for materials science data discovery. *Data Sci. J.* **20**(1), 18 (2021)
75. R.L. Plante, C.A. Becker, A. Medina-Smith, K. Brady, A. Dima, B. Long, L.M. Bartolo, J.A. Warren, R.J. Hanisch, Implementing a registry federation for materials science data discovery. *Data Sci. J.* **20**(1), 15 (2021)
76. R.K. Vasudevan, K. Choudhary, A. Mehta, R. Smith, G. Kusne, F. Tavazza, L. Vlcek, M. Ziatdinov, S.V. Kalinin, J. Hattrick-Simpers, Materials science in the artificial intelligence age: High-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics. *MRS Commun.* **9**(3), 821 (2019)
77. D.B. Miracle, M. Li, Z. Zhang, R. Mishra, K.M. Flores, Emerging capabilities for the high-throughput characterization of structural materials. *Annu. Rev. Mater. Res.* **51**(1), 131 (2021)
78. J.J. de Pablo, N.E. Jackson, M.A. Webb, L.-Q. Chen, J.E. Moore, D. Morgan, R. Jacobs, T. Pollock, D.G. Schlom, E.S. Toberer, J. Analytis, I. Dabo, D.M. DeLongchamp, G.A. Fiete, G.M. Grason, G. Hautier, Y. Mo, K. Rajan, E.J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, J.-C. Zhao, New frontiers for the materials genome initiative. *npj Comput. Mater.* **5**(1), 41 (2019)
79. D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science. *Annu. Rev. Mater. Res.* **50**(1), 71 (2020)
80. M.K. Horton, S. Dwaraknath, K.A. Persson, Promises and perils of computational materials databases. *Nat. Comput. Sci.* **1**(1), 3 (2021)
81. S. Hong, C.H. Liow, J.M. Yuk, H.R. Byon, Y. Yang, E. Cho, J. Yeom, G. Park, H. Kang, S. Kim, Y. Shim, M. Na, C. Jeong, G. Hwang, H. Kim, H. Kim, S. Eom, S. Cho, H. Jun, Y. Lee, A. Baucour, K. Bang, M. Kim, S. Yun, J. Ryu, Y. Han, A. Jetybayeva, P.-P. Choi, J.C. Agar, S.V. Kalinin, P.W. Voorhees, P. Littlewood, H.M. Lee, Reducing time to discovery: Materials and molecular modeling, imaging, informatics, and integration. *ACS Nano* **15**(3), 3971 (2021)
82. J. Fish, G.J. Wagner, S. Keten, Mesoscopic and multiscale modelling in materials. *Nat. Mater.* **20**(6), 774 (2021)
83. L. Chen, G. Pilania, R. Batra, T.D. Huan, C. Kim, C. Kuenneth, R. Ramprasad, Polymer informatics: Current status and critical next steps. *Mater. Sci. Eng. R Rep.* **144**, 100595 (2021)
84. Z. Lu, Computational discovery of energy materials in the era of big data and machine learning: A critical review. *Mater. Rep. Energy* **1**(3), 100047 (2021)
85. S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S.V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.-C.D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, Towards data-driven next-generation transmission electron microscopy. *Nat. Mater.* **20**(3), 274 (2021)
86. B. DeCost, J.R. Hattrick-Simpers, Z. Trautt, A.G. Kusne, E. Campo, M.L. Green, Scientific AI in materials science: A path to a sustainable and scalable paradigm. *Mach. Learn. Sci. Technol.* **1**, 033001 (2020)
87. T.R. Ray, J. Choi, A.J. Bandodkar, S. Krishnan, P. Gutruf, L. Tian, R. Ghaffari, J.A. Rogers, Bio-integrated wearable systems: A comprehensive review. *Chem. Rev.* **119**(8), 5461 (2019) □