

Vision for Data and Informatics in the Future Materials Innovation Ecosystem

SURYA R. KALIDINDI,^{1,2} ANDREW J. MEDFORD,¹
and DAVID L. MCDOWELL¹

1.—Georgia Institute of Technology, Atlanta, GA, USA. 2.—e-mail: surya.kalidindi@me.gatech.edu

The high cost and time typically expended in the successful deployment of new materials into high-performance commercial products is attributable to multiple factors. The most significant of these include the heavy reliance on experiments, the persisting disconnect between multiscale experiments and multiscale models, the lack of a broadly accessible data and knowledge infrastructure that can support the implementation of a holistic systems approach, and the lack of a suitable framework for facilitating and enhancing the critically needed cross-disciplinary collaborations. The emerging discipline of materials data science and informatics (MDSI) promises to address these key technology gaps. The potential benefits to the materials innovation enterprise that could accrue from an aggressive adoption of the novel concepts and toolsets offered by MDSI are examined. A specific vision is expounded for the role of MDSI in bridging the large gap that exists between the multiscale materials experiments and the multiscale materials models.

CHANGING PARADIGM IN MATERIALS INNOVATION

It has become widely recognized in recent and ongoing national and international initiatives such as integrated computational materials engineering (ICME)^{1,2} and the U.S. Materials Genome Initiative (MGI)^{3–5} that the time to bring new and improved materials to market is simply too long to support competitive new manufactured product innovation. The gap between product design cycle time and materials development and certification is untenably large—by as much as 10–15 years in some high-value products in transportation, electronics, and other sectors (see Fig. 1). Indeed, the deployment of high-performance materials in cost-effective and scalable manufacturing processes is a key rate-limiting step in the successful commercialization of most advanced technologies. Success in these efforts will enable improved performance at reduced cost through the deployment of tailored and manufacturable material systems. This capability is central to many twenty-first century grand challenges identified in science and technology, including

light-weighting of transportation vehicles, low-cost sustainable energy, and improved health/quality of life, among others.

There are many reasons for the lag illustrated in Fig. 1. Historically, materials discovery has been largely serendipitous. Materials discovery and development has been largely based on experiments that, for the most part, can be characterized as relatively low throughput and of high quality. Indeed, ICME^{1,2,8,9} has raised the exciting prospect that the availability of a large suite of physics-based multiscale materials models would dramatically lessen the dependency on time-intensive experimental effort. However, the fact remains that the multiscale models have really not connected seamlessly^{10,11} with multiscale measurements. Although the reasons for this disconnect are many, one can point to the following major hurdles. Firstly, the physics-based multiscale models have a very large number of model parameters and/or alternate model form choices that need to be calibrated or somehow adjusted based on carefully designed multiscale experiments. As a specific example, plastic deformation in most structural materials occurs by a multitude of microscale mechanisms.^{12–15}

Although there exist several promising models for reliably capturing the salient details of any selected individual microscale mechanism in a physics-based model, there do not yet exist validated protocols that can de-convolve the relative action of multiple potential mechanisms in a sample subjected to a prescribed deformation condition. Secondly, one of the main reasons for the hurdle described above is that the data demands for systematically and rigorously exploring such verification and validation exercises is substantially large and complex. Often one needs multimodal observations requiring the simultaneous use of multiple sophisticated techniques, each requiring specialized hardware and software. The only practical way to address this overall task is through a coordinated community-wide effort to bring all these techniques to bear. Thirdly, a further complication arises from the fact that there is considerable uncertainty associated with the multiscale experimental observations, in addition to models. This can arise from unavoidable variations in processing of the

material and/or sample preparation for characterization (both structural and mechanical), resolution limits in the machines and techniques employed, fundamental limitations in currently available techniques (for example, most microscopy techniques are only capable of characterizing the material structure on the sample surface) and operator/machine error. In other words, one faces substantial challenges in the fusion and curation of the heterogeneous datasets. Consequently, it becomes clear that one should pursue model calibration and maturation only in a suitable statistical framework that accounts rigorously for the uncertainty associated with the available multimodal experimental observations (including its propagation through multiple scales wherever relevant).

Although frameworks for multilevel design and objective decision support have been developed and applied in multidisciplinary design optimization,^{16,17} the desired data in the materials space have not been available or at least not openly accessible. Furthermore, cross-disciplinary collaborations are essential

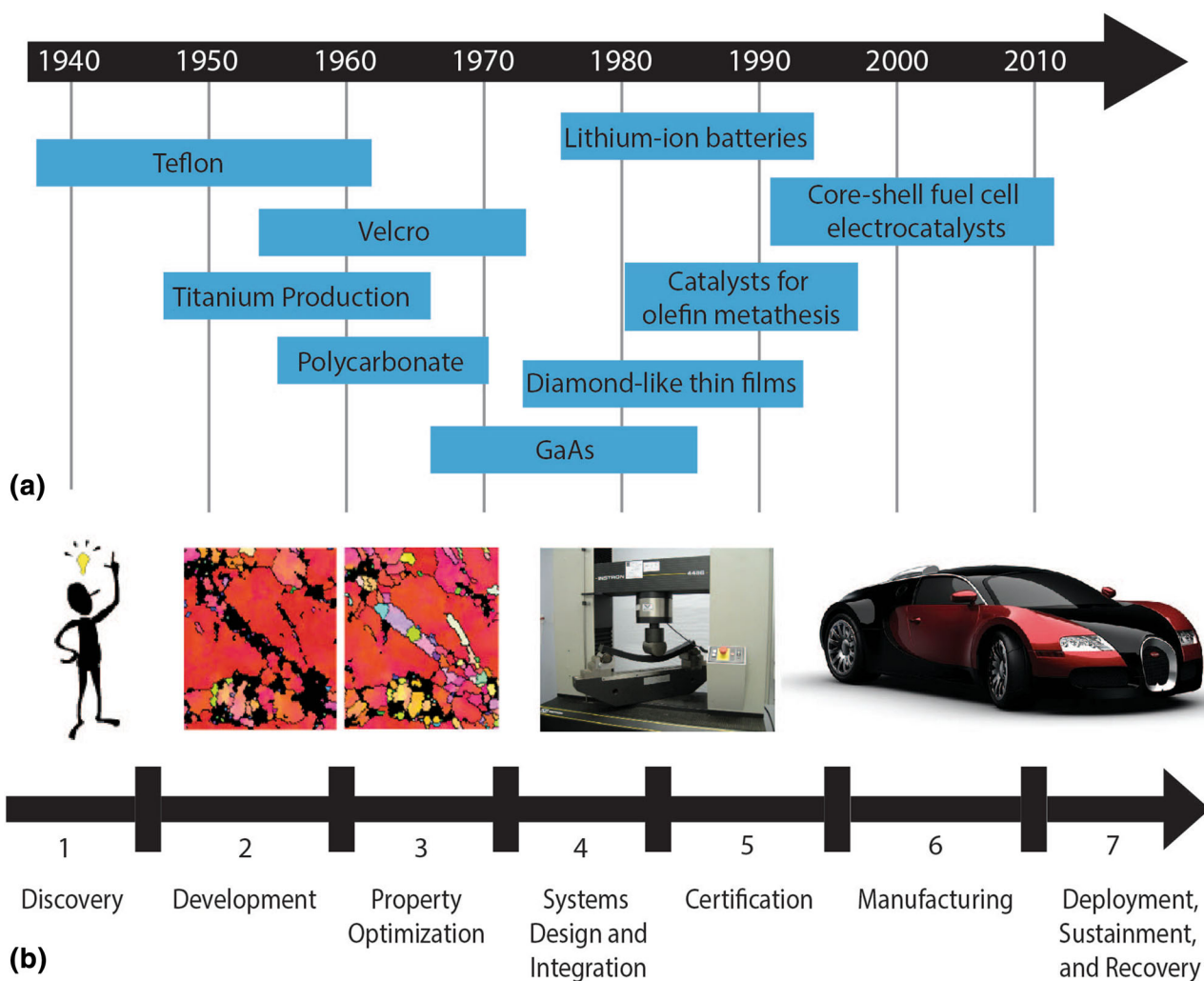


Fig. 1. (a) Illustration of the lag from discovery to application of new materials. (b) Illustration of the multiple, sequential, and time-consuming steps involved in the development and deployment of advanced materials into commercial products. Adapted with permission from J. Warren (NIST);^{3,6} information from Eagar and King.⁷

to realize the goal described above of accelerated and low-cost, scaled-up, materials innovation. Such collaborations have been very slow to set up and difficult to establish. There are silos within the materials community (for example, researchers working on different materials classes do not currently collaborate extensively). All of these factors have served to limit the rate of innovation in linking advances in materials to new and improved products. Given that data are the basic ‘currency’ of cross-disciplinary communication of knowledge, we envision materials data science and informatics playing a central role in the materials innovation ecosystem of the future.^{18,19} Organizing the future materials development workflows around platforms that enable seamless data capture, storage, analysis, and digital collaboration will enable teams of diverse stakeholders from industry, academia, and national laboratories to leverage materials data and expertise in order to accelerate the materials development and deployment process.

EMERGENCE OF MATERIALS DATA SCIENCE AND INFORMATICS

Recent advances in data science and informatics have the potential to offer innovative solutions for addressing several of the impediments listed earlier. Indeed, the MGI white paper³ had already identified ‘digital data’ as an important foundational element for the envisioned acceleration of materials development and deployment (see the intersecting circles of experimental tools, computational tools, and digital data in Fig. 2). The initial discussion of the role of data science in the MGI context was somewhat narrowly focused on the archival and sharing of the important materials datasets and databases (treated largely as digital data). Parallel discussion in the manufacturing community has brought forth innovative concepts such as the digital thread of manufacturing,^{20,21} some of which have been refined and adopted by the original equipment manufacturers (OEMs).^{22,23} In spite of the noteworthy advances already made, there still exists an immense gap between advanced materials and manufacturing.^{24,25} Several of the more recent roadmapping reports^{5,8,9,26,27} have articulated this gap, and have significantly broadened the anticipated role of data science and informatics. This broader definition points to the critical need for a new interdisciplinary field of study called materials data science and informatics (MDSI), whose focus will be on all technical and cultural aspects of the data- and cyber-infrastructure needed to streamline the efficient extraction of high-value materials knowledge (i.e., from all experiments and simulations conducted by the broader materials community, including both legacy and new efforts), and its seamless communication to the manufacturing industry. In this regard, the disparate elements of the materials innovation ecosystem in Fig. 2 convey the broad range of disciplines involved.

In the opinion of the authors, the focus on communicating materials knowledge to manufacturing and creating effective two-way couplings is an important guiding tenet for the success of the numerous national and international strategic initiatives mentioned earlier. Adopting this focus helps sharpen the definition of materials knowledge in terms of process–structure–property (PSP) linkages of high value to manufacturing. In other words, in an effort to move towards the goals listed earlier, we would strive to organize, formulate and express all materials insights (both legacy as well as new) into one of two forms: (1) process–structure (PS) linkages and (2) structure–property (SP) linkages. PS linkages aim to capture the details of material structure evolution as a function of the process parameters (capturing the process history), while SP linkages aim to express the properties (characteristics of materials response) as a function of the material structure. These linkages may take the form of a wide variety of equations or algorithms, but they must be quantitative, reproducible, and digitally captured. It should be noted that material structure plays an important role in both sets of linkages. Indeed, herein lies the main challenge for the task at hand. The mathematical descriptions of both “process” and “property” require relatively low-dimensional representations compared to the “material structure”. Accordingly, both PS and SP linkages are heavily biased towards high dimensionality of structure representation in terms of the numbers of input and output variables involved in formulating the desired linkages. As a result, it is tempting to bypass material structure and seek direct correlation between process route and properties;²⁸ however, this approach is not broadly applicable since it requires a complete description of the processing history, which is often unavailable or incomplete. The material structure captures all relevant aspects of the process history, and can be directly characterized, making it a central feature of PSP linkages. The very large number of variables involved in quantifying the material structure poses significant challenges to conventional approaches in formulating PSP linkages, and demands a new data-driven paradigm, illustrated schematically in Fig. 3. In this figure, the three main activities focused on synthesis and process route, hierarchical structure, and properties/responses appear in three large boxes that point to the current highly siloed disciplinary practices in materials processing and manufacturing sciences, materials sciences, and the mechanical design sciences, respectively. The PSP linkages described above aim to connect the high-value knowledge accumulated in these disciplinary efforts in a consistent manner that provides high value and transparency to the overall effort and, more importantly, facilitate use of systems-level integrated design and optimization methods.¹⁶

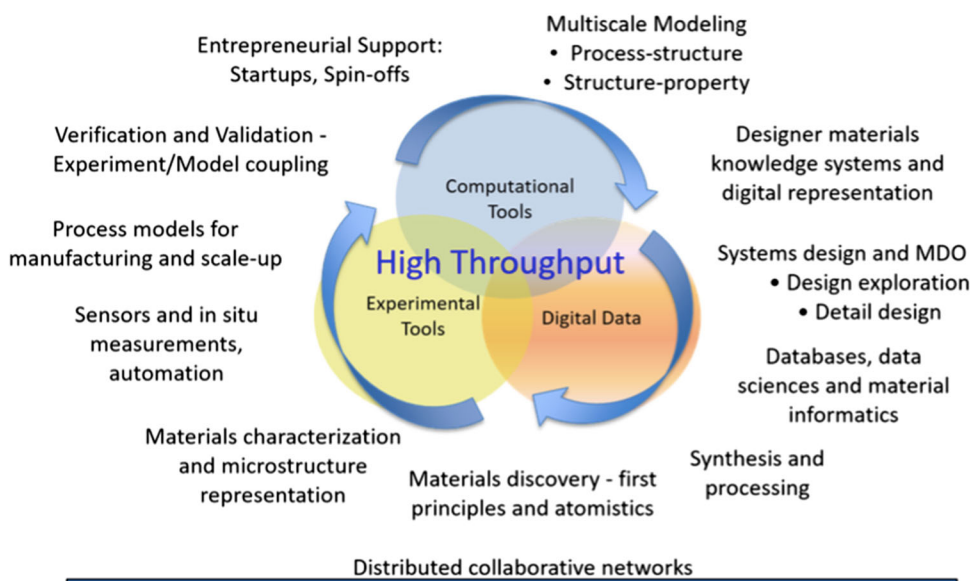


Fig. 2. Elements of the materials innovation ecosystem, as outlined by the Georgia Tech Institute for Materials, generalizing the central theme of the U.S. Materials Genome Initiative of coupling computational tools, experimental tools and digital data with an emphasis on high throughput methods and direct engagement of various stakeholder sub-disciplines that supplement the foundational materials sciences (characterization, representation, discovery, synthesis and process) with systems design, databases and data science, multidisciplinary design optimization, computational multiscale modeling, uncertainty quantification, verification and validation, automation, in situ measurements and scale-up manufacturing processes, as well as support for entrepreneurship and rapid innovation. Reprinted with permission from D.L. McDowell and S.R. Kalidindi.¹⁸

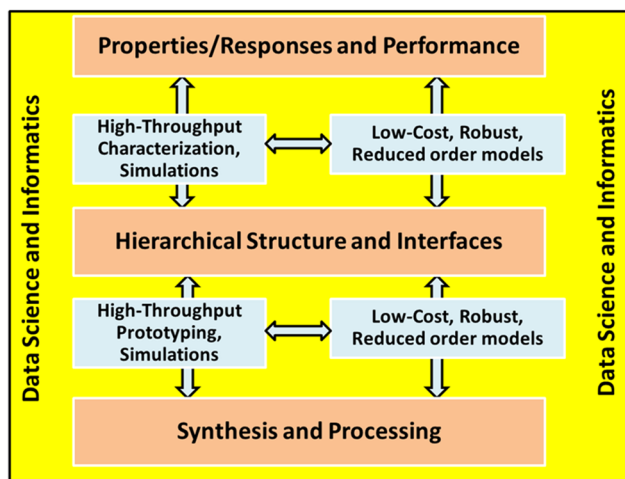


Fig. 3. Envisioned data-driven paradigm in curation of PSP linkages. The study of materials processing, structure, and performance is typically siloed within manufacturing, materials, and mechanical design sciences, respectively. Materials data science and informatics seeks to connect these disciplines through high-throughput approaches and “reduced order” process–structure and structure–property models.

Probing a bit deeper into the desired PSP linkages central to materials innovation, as mentioned earlier, the high dimensionality of the hierarchical material structure presents the central challenge. Consequently, it should be of no surprise that establishing the desired PSP linkages requires a large amount of multiscale data (i.e., results from experiments, models, or both). This requirement

naturally drives materials innovation efforts toward high-throughput strategies (see examples in the functional and biological materials^{29–33} for inspiration). Such high-throughput strategies are still under development in the field of structural materials.^{34–43} In fact, given the complexity involved, it is argued that the development, curation, and dissemination of the “best” high-throughput strategies should be undertaken within a suitable supporting data science and informatics infrastructure (the yellow-colored background) that serves as the “glue” to connect all the overlaid components in Fig. 3. Moreover, data transactions conducted in an open (or open to selected collaborators on a specific project) environment facilitate transparency and promote long-term utility of the knowledge aggregated in any team effort. An anticipated benefit is that the materials innovation efforts will gain significantly (both in reduction of cost and time) from the adoption of the emerging data science and informatics toolsets by eliminating or reducing the unintended redundant effort, focusing the team effort on high-value tasks, and ensuring the highest levels of transferability of the knowledge gained to new problems/challenges.

As a specific example of the potential benefits of adopting data science toolsets, consider the challenges involved with rigorous quantification of the hierarchical structure of a material. The primary challenges in this task arise from (1) the need to describe the hierarchical structure spanning a multitude of length/structure scales (ranging from atomistic to macroscale), (2) the need to adopt a

statistical description that allows quantification of variance and natural insertion into established composite theories (e.g., homogenization theories, localization theories; see also Ref. 44), and (3) the desired versatility to be applicable in a consistent manner to a very broad range of material systems encountered in advanced technology applications. Historically, experts in materials science and engineering have employed mostly intuitive, low-order, microstructure measures such as the overall (averaged) elemental compositions, elemental compositions of constituent phases, phase volume fractions, crystal structure descriptors, average chord lengths (or grain sizes) for constituent phases, average precipitate size/spacing, orientation distribution function, and grain boundary character distribution function, among several others. These explorations have not yet identified clear ‘winners’ for a broad and consistent adoption by the entire materials and manufacturing communities. One approach that has shown tremendous promise to lead to a systematic and comprehensive framework for microstructure measures is based on the formalism of n -point spatial correlations (also simply referred as n -point statistics).^{45–53} In this paradigm, one probes systematically the statistics of what one might find in the neighborhood of every randomly selected point in the material structure.

The most basic n -point statistic are the 1-point spatial correlations (i.e., $n = 1$). These statistical measures of microstructure capture the probability of finding a specified local state of interest at any spatial point selected randomly within the microstructure. In other words, they only capture the information on the volume fraction of various local states (i.e., distinct microstructural constituents) encountered in the material’s internal structure, and capture absolutely no information regarding the surrounding neighborhoods encountered. At the next higher level, the 2-point statistics quantify the neighborhood by looking at one other spatial location relative to the first randomly selected spatial point. As a specific example, f_r^{np} denotes the joint probability of finding local state n at the first randomly selected spatial point in the microstructure, while also finding the local state p at a spatial point that is r away from the first spatial point. It is important to treat r as a vector (that has both a direction and a magnitude) in this definition. Note that f_r^{np} denotes one statistical measure of the microstructure for selected the combination of values of n , p , and r . In general, one utilizes a set of 2-point statistics to quantify any given material structure. The set of 2-point statistics can lead to a very high dimensional representation of the microstructure. Note also that the treatment above can be extended easily, at least conceptually, to higher-order spatial correlations (i.e., 3-point statistics and higher), but with added cost and perhaps diminishing return based on the value of information conveyed.

Principal component analysis (PCA)^{54,55} provides a linear transformation of high-dimensional data in a new orthogonal frame in which the axes are ordered according to the observed variance among the elements of the dataset. Consequently, a truncated PCA representation provides an objective (data-driven) reduced-order representation of the original data. Applying PCA on 2-point spatial correlations of the microstructure has been shown to be remarkably efficient in not only obtaining objective low-dimensional measures of the microstructures but also in establishing high-fidelity PSP linkages (as metamodels or surrogate models to replace numerically expensive models).^{54–57}

The overall fidelity of PSP linkages indeed depends on a number of factors, including (1) the quality and quantity of experimental data utilized, (2) the quality and quantity of physically-based modeling/simulation data utilized, (3) the efficacy and suitability of the analytics performed, and (4) the degree of verification and validation conducted. Consequently, it is quite natural that different PSP linkages formulated for a given phenomenon of interest might exhibit vastly different levels of fidelity and robustness. In the terminology of data science, it is very convenient to think of these data transformations (the process of extracting high value information from data) on a graded scale as data \rightarrow information \rightarrow knowledge \rightarrow wisdom. In the context of the communication of materials knowledge to manufacturing processes, the different levels of data transformations can be benchmarked as shown in Fig. 4.^{11,53} Careful evaluation of the currently available PSP linkages would invariably lead to the realization that a predominant number of them could only be characterized as information, with very few moving up to the knowledge category. This is mainly because the data demands for the validation and verification of multiscale materials models can only be realistically met with carefully organized large scale efforts. Additionally, such an activity requires intimate collaborations between a multitude of disciplines (covering the relevant length and time scales of interest) and approaches (i.e., computations, experiments, analytics, statistics, applied mathematics).

The emerging field of MDSI addresses the critical needs described above with three main interrelated thrusts: (1) Data management, (2) Data Analytics, and (3) e-Collaborations. Data management broadly addresses all aspects of the *datafication*⁵⁸ of materials data, which includes automated capture of data and metadata, robust and reliable storage, aggregation, archival, retrieval, and sharing protocols. Obviously, this is a necessary first component of any MDSI effort, as all other components critically hinge on this one. Some of the challenges in this task arise from the use of different formats used for the files generated by the different techniques employed in multiscale materials characterization.

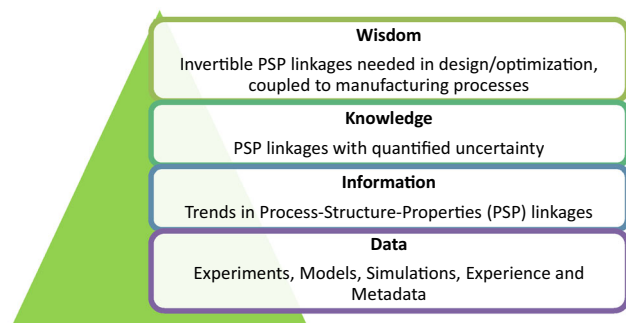


Fig. 4. A graded scale to benchmark the different stages in the data transformations that occur in the process of extracting materials knowledge and expressing it in the forms most useful to manufacturing processes.

A flexible schema is highly desirable. Furthermore, some of the important metadata regarding some of the experimental data are rarely digitally captured and linked with the actual dataset; such data often reside in laboratory notebooks of the experimentalists and are lost with time. A further limitation of the existing modes of data curation is that they seldom capture the complete history of the curation efforts associated with the data—this is especially important to ensure longevity and high utilization of the data. Ideally, one would capture and permanently associate the entire prior history of successes and failures associated with the use of the specific data. This contextual information is critical for the user to develop sufficient confidence in the use of the available databases. For standardizing the data archiving format, XML schema provides an ideal structure for capturing material science knowledge, because it is scalable, modular, and transformable for hierarchical data systems.^{59–62} Building on these concepts, NIST's Materials Data Curation System⁵⁹ allows user-customized capture of a broad variety of materials datasets, along with the relevant metadata. The incorporation of uncertainty quantification with data and/or metadata is a highly desirable advance in future curation approaches.

Once the data and metadata are captured and organized to facilitate easy discoverability and access, one might explore the application of a large number of available data analytic tools. The earlier discussion on the objective low-dimensional representation of material structure and its usage in mining high-fidelity, low-computational cost, PSP linkages provides a good illustration of how one might employ data analytics in materials innovation efforts. In general, this component takes advantage of high-performance computing toolsets based on techniques such as noise filtering, data fusion, uncertainty quantification, statistical analyses, dimensionality reduction, pattern recognition, regression analysis, machine learning, and statistical learning, among others. A large number of these tools can be conveniently accessed through source

code repositories such as R,⁶³ SciPy,⁶⁴ NumPy,⁶⁵ Scikit-learn,⁶⁶ StatsModels,⁶⁷ and Pandas,⁶⁸ as well as through commercial packages such as MATLAB.⁶⁹ The coupling of data analysis with multiscale modeling and experiments at various scales of structure offers important means of calibrating and validating data science methods. In this regard, it is noteworthy that the framework described earlier for establishing PSP linkages based on spatial correlations and PCA can be accessed from the open access, open source, repository, PyMKS.⁷⁰ This repository also provides several case studies illustrating the versatility and utility of the high-level APIs (application program interfaces) provided in PyMKS. These case studies address a broad range of materials systems (metals, polymer composites, etc.) and a broad range of materials phenomena (mechanical loading, molecular dynamics, spinodal decomposition, etc.). A much broader set of case studies demonstrating the versatility and power of the MDSI concepts and toolsets mentioned here can be seen in numerous open access, open source, research blogs^{71,72} disseminated as a part of coursework in the innovative graduate program FLAMEL⁷³ at Georgia Tech. All of these examples provide a clear testament to the transformative role of MDSI in the materials innovation arena.

Recent advances in computer and information technologies have elevated the prospect for dramatically scaling up collaborations through the use of online tools. Called *e-collaboration tools*, these new tools have the potential to team-up diverse expertise (Fig. 2) transcending generational, geographical and organizational barriers, and to direct the combined efforts of a team towards solving important scientific and technological problems. Such e-collaboration platforms provide online access to team and/or project management tools facilitating a wide variety of communications between team members,^{74,75} a suite of discussion and annotation tools, and, perhaps most importantly, workflow capture and management tools for PSP linkages and to couple with manufacturing (e.g., KNIME^{76,77}). Over the past few years, there have been several ongoing efforts at integrating all of the e-collaboration toolsets in a single online platform that will provide easy and convenient access to groups of domain scientists (such as materials scientists). One such effort, called MATIN,⁷⁸ has been in development over the past year at the Georgia Tech's Institute for Materials (GT-IMAT).⁷⁹ MATIN utilizes the open source HUBzero⁸⁰ as an infrastructural foundation, and has built various value-added components on top of this foundation (see Fig. 5).

The rapidly changing landscape of materials innovation driven by the emergence of MDSI presents a significant quandary to industry engaged in materials innovation and deployment in high performance products. If industry does not make the necessary adjustments in their innovation

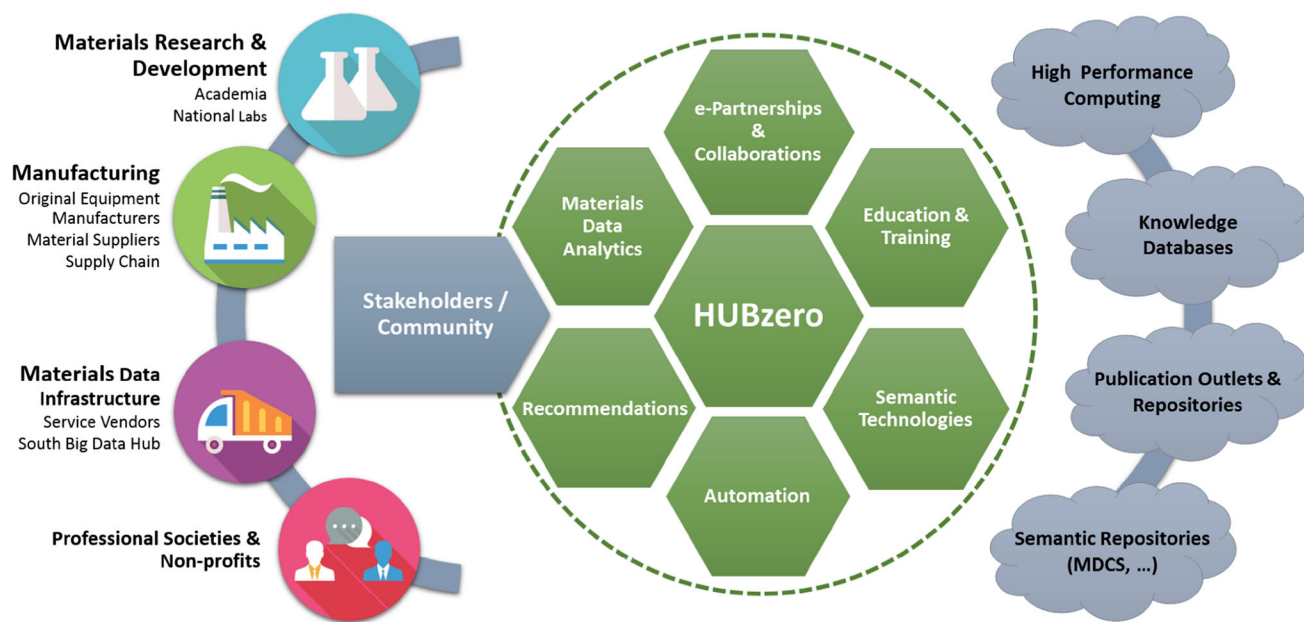


Fig. 5. Schematic of the MATIN e-collaboration platform being currently built and deployed to enhance collaborations among materials researchers at Georgia Tech's Institute for Materials. Reprinted with permission from S.R. Kalidindi et al.¹⁹

workflows to keep up with the fast pace of advances in this emerging field, it risks being left behind by competitors. However, if industry decides to invest to stay ahead of the impending transformation brought about by the data revolution, there may be a shortage of in-house expertise to lead this transformation. The explosion in the sheer numbers of the new online resources and services (including both open and for-free types) is a major challenge to digesting information. This sudden explosion in the MDSI resources, while exciting, also makes it very difficult for any industry to stay abreast and retrain their employees appropriately. Cybersecurity is another challenge. This is a particularly significant challenge for the small and medium-sized enterprises (SMEs) that make up a large fraction of the very extensive supply chain in the advanced materials-manufacturing ecosystem. In the opinion of the authors, this challenge presents a unique opportunity to establish a new kind of university-industry partnership where the university takes on a proactive leadership role in workforce development and training in the emerging MDSI fields, while the industry provides targeted guidance to direct the future development of MDSI. It is important to establish such win-win partnerships to ensure that the new capabilities generated by MDSI are sharply focused on addressing the primary gaps impeding practical accelerated materials innovation.

MATERIALS INNOVATION IN THE FUTURE

The materials innovation ecosystem illustrated in Fig. 2 offers a shared vision of coupling of experiments, computation, and data science via high-

throughput methods to accelerate the discovery and development of new and improved materials via appropriate multi-disciplinary interactions. Such an ecosystem will introduce and develop vital new technologies for materials development and certification. From a technical perspective, it is necessary to develop frameworks and protocols for automated data ingestion, structured data storage, high-throughput exploration (both experiments and models), and integrated data analytics. Furthermore, it will be essential to address the substantial cultural barriers to data sharing by developing novel e-collaboration tools that incentivize participation through increased productivity for all team members. Ultimately, it is envisioned that the confluence of an efficient and robust technical infrastructure with a diverse and committed set of stakeholders will create a vibrant data-driven materials innovation cyber-ecosystem capable of realizing the revolutionary impact of big data on the materials and manufacturing sectors.^{3,81,82} The main components of this materials innovation cyber-ecosystem are described below.

High-Throughput Characterization

Most modern techniques in materials characterization focus on obtaining a few measurements of very high quality. Indeed, the majority of government agency-funded user facilities lie in this category; they are too expensive to replicate and properly support and sustain in university or industry laboratories. While valuable to basic science, these high-end, high-fidelity methods often fail to provide sufficiently rich datasets for maturation of multiscale materials models. Furthermore, the

small quantities of data captured in these protocols may not adequately represent the inherent microstructure variance that exists within the hierarchical material system being studied. It is imperative to develop and validate high-throughput measurement strategies capable of undertaking rapid but highly targeted explorations of PSP linkages of high value to manufacturing processes employed by the industry. Furthermore, it is necessary to deploy these novel strategies in shared user facilities to ensure broad access and accelerated learning of best practices. As a specific example, nanoindentation techniques have demonstrated the potential for such high-throughput explorations in scale-specific measurement of mechanical properties.^{83,84} The continued development of such approaches will drive the need for the expansion and the adoption of the data-driven materials development by providing large, rich, datasets for knowledge extraction.

Automated Ingestion

Currently, the multiscale characterization of a typical material can easily lead to the generation of terabytes of data from multimodal investigations that might include x-ray, microscopy, and spectroscopy techniques (see Fig. 6). However, the vast majority of this data are siloed on local storage disks making them difficult to share with team members. More importantly, the pertinent metadata necessary to place the data into a meaningful context is often not captured electronically or sometimes even lost. In order to promote and facilitate a data-driven materials innovation cyber-ecosystem, it will become increasingly important to establish standards and systems for the automated ingestion of data. Several projects have begun implementing strategies to automate the capture of electronic structure calculations (e.g., Materials Project,⁸⁵ AFLOW⁸⁶), yet the capture of most simulation results and nearly all experimental data will require the development of new approaches and tools. The concept of the “internet of things”⁸⁷ presents a novel paradigm for retrofitting and designing characterization facilities equipped with embedded electronics and network connectivity in order to facilitate the seamless ingestion of large quantities of high fidelity materials characterization data and meta-data. These techniques can be pioneered at academic and government materials characterization facilities, leading to open, high-value, materials datasets and protocols and best practices for implementation in industrial settings.

Storage and Curation

In order for automated data ingestion to be successful, it is critical that the ingestion systems are coupled with data storage structures capable of handling the big data that will be created. The inherently heterogeneous, diverse, and hierarchical

nature of materials characterization and simulation has led to a “variety” challenge (see Fig. 6),⁸⁸ where information must be fused across a multitude of length (angstroms to meters) and time (femtoseconds to years) scales. Furthermore, the available theoretical and experimental approaches in materials research are constantly evolving, creating the need for storage systems capable of adapting to fundamentally new data types while simultaneously supporting and curating legacy data. Designing data storage and archival systems that balance the flexibility needed to accommodate the variety of materials data with the structure and schemas needed for rapid retrieval and analysis of large datasets presents a significant challenge. Although in early stages, initiatives such as the NIST MDSC project⁵⁹ have demonstrated success in storing and curating complex materials datasets, and the continued development and implementation of such systems will be critical to the materials innovation ecosystem of the future.

Reproducible and Data-driven Analytics

As rich datasets are automatically generated, ingested, and stored, the need for advanced data analysis frameworks will become progressively more important. Tools for creating quantitative, templatable, and potentially invertible PSP linkages will be a cornerstone of the future materials data infrastructure. These tools must be able to fuse information across various length and time scales, as well as data from the diverse experimental and simulation techniques commonly employed in materials research. The use of statistical representations and frameworks present a promising route forward, as indicated by novel approaches such as n -point correlations,⁴⁸ materials knowledge systems,^{70,89} and Bayesian parameter estimation.⁹⁰ In addition to developing new analysis tools, it will also be necessary to capture and disseminate the workflows in which these tools are applied for data analysis. Workflow capture provides a transparent route to communicating materials data analysis, simultaneously improving reproducibility and transferability.⁹¹ Other fields have developed and adapted numerous tools to address workflow capture such as Galaxy (bioinformatics),⁹² Kepler (ecology/biology)⁹³ and KNIME (business/consulting);^{76,77} however, the materials community is currently lacking tools which are sufficiently customized to capture the diverse and evolving workflows associated with materials data analysis. Initial work at Georgia Tech has established “research blogs” as a platform for capturing workflows based on heterogeneous software and visualization tools. These blogs have been used to capture workflows for data-driven approaches to a wide variety of case studies in materials innovation.^{70–72} The use of well-defined workflows will also facilitate the quantification and propagation of uncertainty through complex models

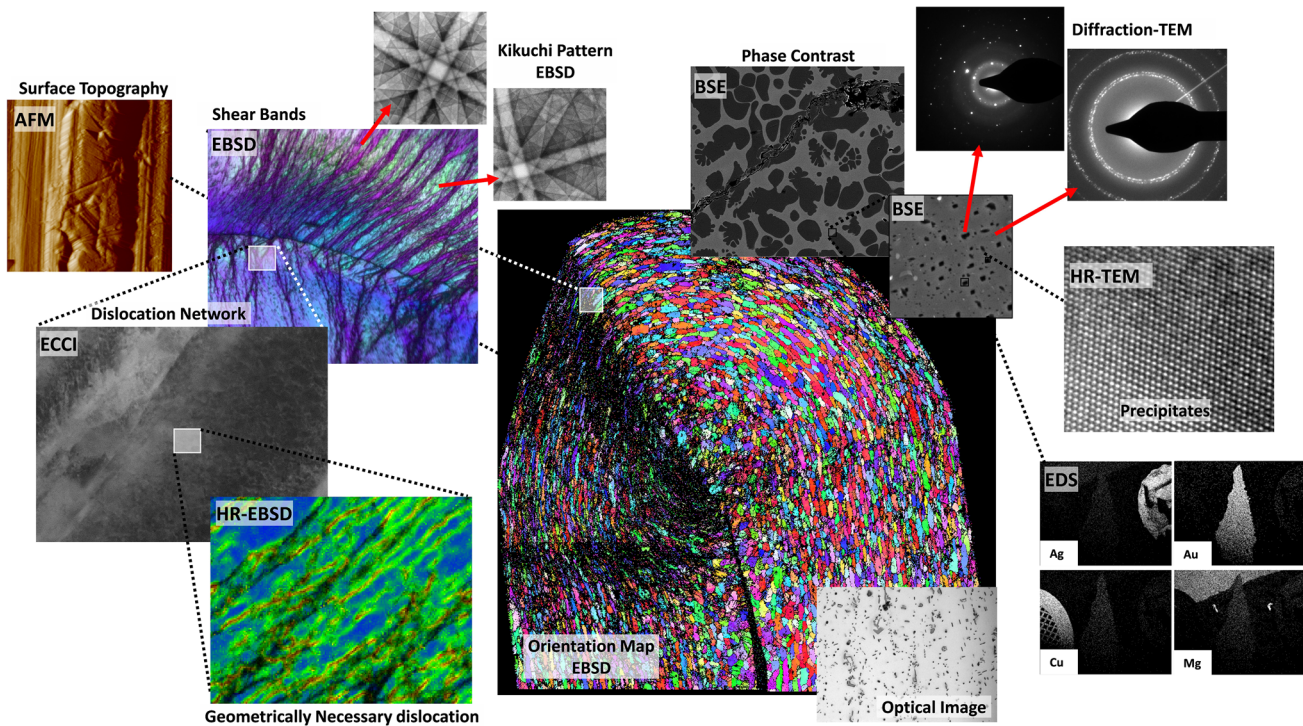


Fig. 6. Illustration of the diversity of data collected to characterize internal structure of dual-phase steels. The techniques include electron backscatter diffraction (EBSD), atomic force microscopy (AFM), electron channeling contrast imaging, backscattered electrons (BSE), transmission electron microscopy (TEM), and energy dispersive x-ray spectroscopy (EDS). This is only a representative sampling of the numerous techniques used in current materials characterization protocols. Reprinted with permission of Ali Khosravani.

to provide decision support in materials design and development. Such systems will include verification and validation as part of a feedback loop, providing a natural route to include more information in order to improve the confidence of the decision support system. This data-enabled coupling of diverse simulation and experimental results through statistical models is expected to have a profound impact on materials design and development processes, and their coupling with manufacturing.

Industry Outreach and Consortia

The participation of industry will be critical to the long-term success and sustainability of a data-centric materials innovation ecosystem. From the materials research perspective, industry buy-in constitutes verification of the utility and economic advantages of data-driven approaches to materials research and development. Conversely, from the materials data perspective, the emergence of commercial ventures seeking to deliver materials data services will establish self-sustaining components of the materials data infrastructure. In order to facilitate these impacts of materials data on industry, it will be necessary for academic leaders to promote both adoption and commercialization of data science tools, as well as acting as honest brokers in identifying mutually beneficial partnerships between various stakeholders in the materials innovation

ecosystem shown in Fig. 2. Companies with products in computational materials simulation, data science, and life cycle engineering are included in this broader ecosystem. In 2016, Georgia Tech has established a new center called IDEAS: Materials Design, Development and Deployment (IDEAS:MD3) specifically to promote such university–industry–national laboratory partnerships. Briefly, this new center aims to provide:

- Low entry cost opportunities to familiarize and train current industry workforce in the emerging concepts and toolsets of MDSI.
- Increased awareness in the industry of the potential benefits and pitfalls in the adoption of modern data science tools and e-collaboration platforms for materials discovery and development.
- Increased level of cross-disciplinary collaborations between university and industry experts at the nexus of materials science, manufacturing, data science, and high-throughput methods.

Community building is an important cornerstone for this activity. Consequently, IDEAS:MD3 aims to execute a number of community-building activities that include workshops and tutorials in the emerging field of MDSI. As a specific example, a materials “data challenge” was conducted as a two-day competition⁹⁴ employing ASM international’s SMDD (structural materials data demonstration) project.⁹⁵

E-Collaboration Platforms

In order to address both technical and cultural barriers to data-centric materials design and discovery, it will be important to develop e-collaboration platforms to encourage the formation of diverse, multi-disciplinary teams and facilitate the sharing of data, intermediate results, and workflows amongst team members.⁹⁶ Such platforms⁹⁷ are fundamentally different from the many existing data and code repositories in that they specifically seek to address cultural challenges to adoption of materials informatics approaches by creating communities of like-minded materials researchers, data scientists, and industry leaders who will act as early adopters of materials data approaches. As such, the success of these platforms will depend on effective social networking strategies for recruiting and retaining key stakeholders as well as organically identifying win-win partnerships between members. In addition, e-collaboration platforms will act as a front-facing portal to the technical infrastructure of the data-centric materials innovation ecosystem, facilitating easy access to computing cyberinfrastructure and sharing of materials datasets and workflows (cf. Fig. 5). The previously discussed MATIN effort at Georgia Tech⁷⁸ is an example of an e-collaboration platform in the materials data sciences sector.

Education and Training Programs

As data-driven methods begin to impact the materials research and development sector, the skills expected of a materials researcher will evolve. In order to address this critical need, it will be necessary to modify and design academic curricula that address and incorporate relevant data science approaches to produce a new cadre of materials engineers with skills at the intersection of data and materials sciences. Recently, innovative educational programs such as the FLAMEL IGERT at Georgia Tech,⁷³ the D³EM traineeship at Texas A&M,⁹⁸ ICME courses and Master's programs/certificates at Ruhr-University Bochum,⁹⁹ Mississippi State¹⁰⁰ and Northwestern University,¹⁰¹ and summer institutes for computational and data-driven materials science at the Technical University of Denmark,¹⁰² University of Michigan,¹⁰³ Texas A&M,¹⁰⁴ Lawrence Livermore National Laboratory,¹⁰⁵ and University of Florida¹⁰⁶ have begun to address these educational challenges at the graduate level. However, it will be critical to explore strategies of expanding these educational paradigms to various levels of education (e.g., undergraduate, high school) and to promote their adoption into more diverse educational sectors (e.g., undergraduate and minority-serving institutions). Furthermore, the envisioned rapid and transformational effects of big data on materials innovation will require re-training of much of the current workforce. This too will necessitate the development of novel

educational programs aimed at full-time employees, such as distance learning and massive open online courses (MOOCs) in materials data science tools and techniques.¹⁰⁷ The envisioned data-centric materials innovation ecosystem will incorporate these education and training approaches in order to nurture, grow, and sustain the data-aware materials workforce of the future.

SUMMARY

In summary, the instantiation of a data-centric materials innovation ecosystem presents a significant challenge that will require a concerted effort from numerous and diverse stakeholders. Significant progress has already been made towards achieving many of the necessary technical constituents of this synergistic network of disciplines and sub-fields. Hence, the most important, and perhaps most challenging, step in the realization of the materials data revolution will be the successful integration of these pieces into a coherent framework, and the proactive adoption of this framework by the academic, industrial, and governmental stakeholders of the materials innovation ecosystem. The intention of this vision for the role of materials data sciences in the future innovation framework is to spur discussion within the community and foster adoption of these or similar approaches among and between various stakeholders.

ACKNOWLEDGEMENTS

SRK and AM acknowledge support from NIST 70NANB14H191 and internal funding from Georgia Tech's IDEAS grant. DLM is grateful for the support of the Georgia Tech Institute for Materials, as well as the Carter N. Paden, Jr. Distinguished Chair in Metals Processing.

REFERENCES

1. T.M. Pollock, J.E. Allison, D.G. Backman, M.C. Boyce, M. Gersh, E.A. Holm, R. LeSar, M. Long, A.C. Powell, J.J. Schirra, D.D. Whitis, and C. Woodward, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security* (Washington, DC: The National Academies Press, 2008).
2. G.J. Schmitz and U. Prael, *Integr. Mater. Manuf. Innov.* (2014). doi:10.1186/2193-9772-3-2.
3. Materials Genome Initiative for Global Competitiveness (National Science and Technology Council, 2011), http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf. Accessed 29 Mar 2016.
4. M. Drosback, *JOM* 66, 334 (2014).
5. The Materials Genome Initiative Strategic Plan (Materials Genome Initiative National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative, 2014). https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf. Accessed 29 Mar 2016.
6. J. Warren, The materials genome initiative, data, open science, and NIST (Research Data Alliance, 2014). <https://rd-alliance.org/sites/default/files/RDA-MGI-ODIpptx.pptx.pdf>. Accessed 29 Mar 2016.
7. T.W. Eager and M. King, *Technol. Rev.* 98, 42 (1995).

8. Implementing ICME in the Aerospace, Automotive, and Maritime Industries (The Minerals, Metals and Materials Society, 2013), <http://www.tms.org/icmestudy/>. Accessed 29 Mar 2016.
9. Modeling Across Scales: A Roadmapping Study for Connecting Materials Models and Simulations Across Length and Time Scales (The Minerals, Metals and Materials Society, 2015). <http://www.tms.org/multiscalestudy/>. Accessed 29 Mar 2016.
10. J.H. Panchal, S.R. Kalidindi, and D.L. McDowell, *Comput. Aided Des.* 45, 4 (2013).
11. S.R. Kalidindi, *Int. Mater. Rev.* 60, 150 (2015).
12. D.L. McDowell, *Int. J. Plast* 26, 1280 (2010).
13. A.A. Salem, S.R. Kalidindi, and S.L. Semiatin, *Acta Mater.* 53, 3495 (2005).
14. F. Roters, P. Eisenlohr, L. Hantcherli, D.D. Tjahjanto, T.R. Bieler, and D. Raabe, *Acta Mater.* 58, 1152 (2010).
15. X.P. Wu, S.R. Kalidindi, C. Necker, and A.A. Salem, *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* 39A, 3046 (2008).
16. D.L. McDowell, J.H. Panchal, H.-J. Choi, C.C. Seepersad, J.K. Allen, and F. Mistree, *Integrated Design of Multiscale, Multifunctional Materials and Products* (New York: Elsevier Inc., 2010).
17. H.J. Choi, D.L. McDowell, J.K. Allen, D. Rosen, and F. Mistree, *J. Mech. Des.* 130, 031402 (2008).
18. D.L. McDowell and S.R. Kalidindi, *MRS Bull.* 41, 326 (2016). doi:10.1557/mrs.2016.61.
19. S.R. Kalidindi, D.B. Brough, S. Li, A. Cecen, A.L. Blekh, F. Yannick, P. Congo, and C. Campbell, *MRS Bull.* (2016). doi:10.1557/mrs.2016.164.
20. Digital Thread for Smart Manufacturing (National Institute of Standards and Technology, 2014). <http://www.nist.gov/el/msid/syseng/dtsm.cfm>. Accessed 29 Mar 2016.
21. "DMDII" (Digital Manufacturing and Design Innovation Institute, 2016). <http://dmdii.uilabs.org/>. Accessed 29 Mar 2016.
22. The Digital Twin (General Electric, 2015). <http://gelooka.head.economist.com/digital-twin/>. Accessed 29 Mar 2016.
23. "Digital Tapestry" (Lockheed Martin, 2016). <http://www.lockheedmartin.com/us/ssc/digital-tapestry.html>. Accessed 29 Mar 2016.
24. A National Strategic Plan for Advanced Manufacturing (National Science and Technology Council Executive Office of the President, 2012). https://www.whitehouse.gov/sites/default/files/microsites/ostp/iam_advancedmanufacturing_strategicplan_2012.pdf. Accessed 29 Mar 2016.
25. National Network for Manufacturing Innovation (Advanced Manufacturing National Program Office, 2015), <https://www.manufacturing.gov/nmii-institutes/>. Accessed 29 Mar 2016.
26. Materials Data Analytics: A Path-Finding Workshop (ASM International, 2015). <http://www.asminternational.org/documents/10192/25925847/1-MDA-Henry+Intro+2015-10-08.pdf/a67c84f5-4f44-48e3-b096-5a70352b338a>. Accessed 29 Mar 2016.
27. Building an Integrated MGI Accelerator Network (Materials Accelerator Network, 2014). <http://acceleratonet.org/events/past-events/building-an-integrated-mgi-accelerator-network/>. Accessed 29 Mar 2016.
28. J.H. Forsmark, J.W. Zindel, L. Godlewski, J. Zheng, J.E. Allison, and M. Li, *Integr. Mater. Manuf. Innov.* (2015). doi:10.1186/s40192-015-0033-0.
29. W.F. Maier, K. Stowe, and S. Sieg, *Angewandte Chem. Int. Edition* 46, 6016 (2007).
30. R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, and H. Lam, *ACS Comb. Sci.* 13, 579 (2011).
31. C.G. Simon and S. Lin-Gibson, *Adv. Mater.* 23, 369 (2011).
32. J.C. Zhao, *Chin. Sci. Bull.* 59, 1652 (2014).
33. M.L. Green, I. Takeuchi, and J.R. Hatrick-Simpers, *J. Appl. Phys.* 113, 231101 (2013).
34. H. Springer and D. Raabe, *Acta Mater.* 60, 4950 (2012).
35. F. Warchomicka, C. Poletti, M. Stockinger, and T. Henke, *Int. J. Mater. Form.* 3, 215 (2010).
36. D.B. Miracle, J.D. Miller, O.N. Senkov, C. Woodward, M.D. Uchic, and J. Tiley, *Entropy* 16, 494 (2014).
37. J.C. Zhao, M.R. Jackson, L.A. Peluso, and L.N. Brewer, *JOM* 54, 42 (2002).
38. J.C. Zhao, *J. Mater. Res.* 16, 1565 (2001).
39. O.L. Warren and T.J. Wyrobek, *Meas. Sci. Technol.* 16, 100 (2005).
40. V.V. Shastry, V.D. Divya, M.A. Azeem, A. Paul, D. Dye, and U. Ramamurty, *Acta Mater.* 61, 5735 (2013).
41. S.M. Han, R. Shah, R. Banerjee, G.B. Viswanathan, B.M. Clemens, and W.D. Nix, *Acta Mater.* 53, 2059 (2005).
42. E. Menendez, C. Templier, G. Abrasonis, J.F. Lopez-Barbera, J. Noguez, K. Temst, and J. Sort, *Cryst. Eng. Commun.* 16, 3515 (2014).
43. C.A. Tweedie, D.G. Anderson, R. Langer, and K.J. Van Vliet, *Adv. Mater.* 17, 2599 (2005).
44. G.W. Milton, *The Theory of Composites* (Cambridge: Cambridge University Press, 2002).
45. W.F. Brown, *J. Chem. Phys.* 23, 1514 (1955).
46. S. Torquato, *Random Heterogeneous Materials* (New York: Springer-Verlag, 2002).
47. S. Torquato and G. Stell, *J. Chem. Phys.* 82, 980 (1985).
48. S. Torquato and G. Stell, *J. Chem. Phys.* 77, 2071 (1982).
49. B.L. Adams, S.R. Kalidindi, and D. Fullwood, *Microstructure Sensitive Design for Performance Optimization* (New York: Elsevier Inc., 2013).
50. D.T. Fullwood, S.R. Niezgodza, B.L. Adams, and S.R. Kalidindi, *Progress Mater. Sci.* 55, 477 (2010).
51. E. Kroner, *Statistical modelling. Modelling Small Deformations of Polycrystals*, ed. J. Gittus and J. Zarka (London: Elsevier Science Publishers, 1986), p. 229.
52. E. Kroner, *J. Mech. Phys. Solids* 25, 137 (1977).
53. S.R. Kalidindi, *Hierarchical Materials Informatics* (Oxford: Elsevier Inc., 2015).
54. S.R. Kalidindi, S.R. Niezgodza, and A.A. Salem, *JOM* 63, 34 (2011).
55. S.R. Niezgodza, Y.C. Yabansu, and S.R. Kalidindi, *Acta Mater.* 59, 6387 (2011).
56. A. Cecen, T. Fast, E.C. Kumbur, and S.R. Kalidindi, *J. Power Sources* 245, 144 (2014).
57. S.R. Niezgodza, A.K. Kanjarla, and S.R. Kalidindi, *Integr. Mater. Manuf. Innov.* (2013). doi:10.1186/2193-9772-2-3.
58. What is Datafication?—Definition and Examples (Upfront Analytics, 2015). <http://upfrontanalytics.com/datafication-definition-examples/>. Accessed 29 Mar 2016.
59. Materials Data Curation System (Github, 2014). <https://github.com/usnistgov/MDCS>. Accessed 29 Mar 2016.
60. J.G. Kaufman and E.F. Begley, *Adv. Mater. Process.* 161, 35 (2003).
61. A.S. Varde, E.F. Begley, and S. Fahrenholz-Mann, MatML: XML for information exchange with materials property data, in *Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms* (2006), p. 47.
62. R.D. Chirico, M. Frenkel, V.V. Diky, K.N. Marsh, and R.C. Wilhoit, *J. Chem. Eng. Data* 48, 1344 (2003).
63. The R Project for Statistical Computing (The R Project, 2016). <https://www.r-project.org/>. Accessed 29 Mar 2016.
64. E. Jones, T. Oliphant, and P. Peterson, SciPy: Open Source Scientific Tools for Python (Enthought, 2016). <http://www.scipy.org>. Accessed 29 Mar 2016.
65. S.V.D. Walt, S.C. Colbert, and G. Varoquaux, *Comput. Sci. Eng.* 13, 22 (2011).
66. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesney, *J. Mach. Learn. Res.* 12, 2825 (2011).
67. S. Seabold, J. Perktold, in *Proceedings of the 9th Python in Science Conference* (2010), p. 57.
68. W. McKinney, in *Proceedings of the 9th Python in Science Conference* (2010), p. 51.
69. Matlab: The language of Technical Computing (Mathworks, 2016). <http://www.mathworks.com/products/matlab/index.html>. Accessed 29 Mar 2016.

70. D. Wheeler, D. Brough, T. Fast, S. Kalidindi, A. Reid, PyMKS: Materials Knowledge System in Python (Figshare, 2014). <http://dx.doi.org/10.6084/m9.figshare.1015761>. Accessed 29 Mar 2016.
71. Materials Informatics Class Fall 2014 (Georgia Institute of Technology, 2014). <http://materials-informatics-class-fall2014.github.io/>. Accessed 29 Mar 2016.
72. Materials Informatics Class Fall 2015 (Georgia Institute of Technology, 2015). <http://materials-informatics-class-fall2015.github.io/>. Accessed 29 Mar 2016.
73. From Learning, Analytics, and Materials to Entrepreneurship and Leadership Doctoral Traineeship Program (Georgia Institute of Technology, 2016). <http://www.flamel.gatech.edu/>. Accessed 29 Mar 2016.
74. K. Börner, N. Contractor, H.J. Falk-Krzesinski, S.M. Fiore, K.L. Hall, J. Keyton, B. Spring, D. Stokols, W. Trochim, and B. Uzzi, *Sci. Transl. Med.* 2, 49 (2010).
75. D. Stokols, S. Misra, R.P. Moser, K.L. Hall, and B.K. Taylor, *Am. J. Prev. Med.* 35, S96 (2008).
76. KNIME: Open for Innovation (KNIME, 2016). <https://www.knime.org/>. Accessed 29 Mar 2016.
77. M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel, KNIME: the Konstanz information miner. *Studies in Classification, Data Analysis, and Knowledge Organization*, ed. H. Bock, W. Gaul, M. Vichi, and C. Weihs (New York, NY: Springer, 2008), p. 319.
78. MATIN: An e-Collaboration Platform for Materials Informatics (Georgia Institute of Technology, 2016). <https://www.matin.gatech.edu>. Accessed 5 July 2016.
79. Georgia Tech Institute for Materials (Georgia Institute of Technology, 2016). <http://www.materials.gatech.edu/>. Accessed 29 March 2016.
80. HUBzero: Platform for Scientific Collaboration (HUBzero Foundation, 2016). <https://hubzero.org/>. Accessed 29 Mar 2016.
81. Integrated Computational Materials Engineering (National Academies of Engineering, 2008). <http://www.nap.edu/catalog/12199/integrated-computational-materials-engineering-a-transformational-discipline-for-improved-competitiveness>. Accessed 29 Mar 2016.
82. A National Advanced Manufacturing Portal (Advanced Manufacturing National Program Office, 2016). www.manufacturing.gov. Accessed 5 July 2016.
83. M. Göken, M. Kempf, M. Bordenet, and H. Vehoff, *Surf. Interface Anal.* 27, 302 (1999).
84. S. Pathak and S.R. Kalidindi, *Mater. Sci. Eng.: R: Rep.* 91, 1 (2015).
85. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, *APL Mater.* 1, 011002 (2013).
86. R.H. Taylor, F. Rose, C. Toher, O. Levy, K. Yang, M. Buongiorno, M.B. Buongiorno, and S. Curtarolo, *Comput. Mater. Sci.* 93, 178 (2014).
87. L. Atzori, A. Iera, and G. Morabito, *Comput. Netw.* 54, 2787 (2010).
88. Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, Addressing big data issues in scientific data infrastructure, in *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems* (2013), p. 48.
89. T. Fast, S.R. Niezgodna, and S.R. Kalidindi, *Acta Mater.* 59, 699 (2011).
90. J. Wellendorff, K.T. Lundgaard, A. Møgelhøj, V. Petzold, D.D. Landis, J.K. Nørskov, T. Bligaard, and K.W. Jacobsen, *Phys. Rev. B Condens. Matter Mater. Phys.* 85, 235149 (2012).
91. S. M. Arnold, B. A. Bednarzyk, N. Austin, I. Terentjev, D. Cebon, and W. Marsden, Information management workflow and tools enabling multiscale modeling within ICME Paradigm, in *57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference* (2016), p. 1174.
92. B. Giardine, C. Riemer, R.C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, and J. Taylor, *Genome Res.* 15, 1451 (2005).
93. B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E.A. Lee, J. Tao, and Y. Zhao, *Concurr. Comput. Pract. Exp.* 18, 1039 (2006).
94. Materials Data Challenge 2016 (Georgia Institute of Technology, 2016). <http://www.flamel.gatech.edu/data-challenge-2016>. Accessed 5 July 2016.
95. Structural Materials Data Demonstration Project (ASM International, 2015). <http://www.asminternational.org/web/cmdnetwork/projects/structural-materials>. Accessed 29 Mar 2016.
96. S.R. Kalidindi and M.D. Graef, *Annu. Rev. Mater. Res.* 45, 171 (2015).
97. A.F. Rutkowski, D.R. Vogel, M. Van Genuchten, T.M.A. Bemelmans, and M. Favier, *IEEE Trans. Prof. Commun.* 45, 219 (2002).
98. NRT: Data-Enabled Discovery and Design of Energy Materials (D3EM) (Texas A&M University, 2015). <https://engineering.tamu.edu/news/2015/08/19/texas-am-receives-3-million-nsf-grant-to-train-graduate-students>. Accessed 29 Mar 2016.
99. Master's Program in Materials Science and Simulation at Ruhr-University Bochum (Ruhr-University Bochum, 2016). <http://www.icams.de/content/master-course-mss>. Accessed 29 Mar 2016.
100. ICME courses at Mississippi State University (Mississippi State University, 2015). https://icme.hpc.msstate.edu/mediawiki/index.php/Mississippi_State_University. Accessed 29 Mar 2016.
101. ICME Masters certificate at Northwestern University (Northwestern University, 2016). <http://www.mccormick.northwestern.edu/materials-science/documents/graduate/icme-brochure.pdf>. Accessed 29 Mar 2016.
102. CAMD Summer School on Electronic Structure Theory and Materials Design (Technical University of Denmark, 2016). <http://www.fysik.dtu.dk/english/Research/CAMD/Events/Summer-school-2016>. Accessed 29 Mar 2016.
103. The University of Michigan Summer School on Integrated Computational Materials Education (University of Michigan, 2016). <http://icmed.engin.umich.edu>. Accessed 29 Mar 2016.
104. IIMEC Summer School on Computational Materials Science Across Scales (Texas A&M University, 2015). <http://engineering.tamu.edu/news/2015/08/03/texas-am-hosts-fourth-iimec-summer-school-on-computational-materials-science-across-scales>. Accessed 29 Mar 2016.
105. Lawrence Livermore National Laboratory Computational Chemistry and Materials Science Summer Institute (Lawrence Livermore National Laboratory, 2015). https://www-pls.llnl.gov/?url=jobs_and_internships-internships-ccms. Accessed 29 Mar 2016.
106. Cyberinfrastructure for Atomistic Materials Science Center Summer Workshop (University of Florida, 2016). <http://cams.mse.ufl.edu>. Accessed 29 Mar 2016.
107. Materials Data Science and Informatics (Coursera). <https://www.coursera.org/learn/material-informatics>. Accessed 20 July 2016.