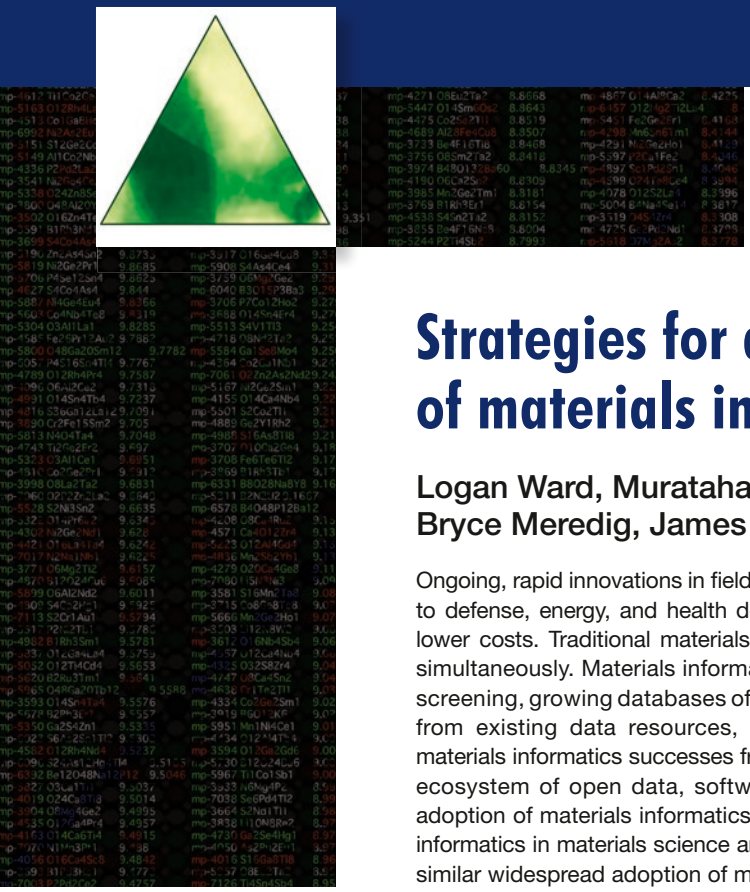


Strategies for accelerating the adoption of materials informatics

Logan Ward, Muratahan Aykol, Ben Blaiszik, Ian Foster, Bryce Meredig, James Saal, and Santosh Suram

Ongoing, rapid innovations in fields ranging from microelectronics, aerospace, and automotive to defense, energy, and health demand new advanced materials at even greater rates and lower costs. Traditional materials R&D methods offer few paths to achieve both outcomes simultaneously. Materials informatics, while a nascent field, offers such a promise through screening, growing databases of materials for new applications, learning new relationships from existing data resources, and building fast predictive models. We highlight key materials informatics successes from the atomic-scale modeling community, and discuss the ecosystem of open data, software, services, and infrastructure that have led to broad adoption of materials informatics approaches. We then examine emerging opportunities for informatics in materials science and describe an ideal data ecosystem capable of supporting similar widespread adoption of materials informatics, which we believe will enable the faster design of materials.

Introduction

The use of informatics techniques in materials science has increased dramatically in the past decade, and for good reason. Data-driven techniques, which focus on extracting predictive models or knowledge from data, are starting to yield revolutionary impacts in a multitude of fields such as transportation[1] and health care.[2] Informatics methods are also increasingly prevalent across diverse domains within materials science. For example, the use of data-driven methods has enabled the rapid discovery of new materials[3] and improved characterization methods,[4] while providing another path for understanding material behavior.[5] Overall, materials informatics has promise to create impactful advancements in many areas of materials science. But the question of how the adoption of these useful techniques can be speeded up remains.

The use of materials informatics has been especially prevalent in atomic-scale modeling. Materials discovery achieved by screening databases of material properties computed using quantum-mechanics-based calculations has become a routine tool in the development of materials,[6] as evidenced by the many publicly available databases of density functional theory (DFT) calculations.[7] These databases, in turn, have been used with automated analysis techniques to learn more about materials behavior. Furthermore, it is increasingly common to use machine learning to build fast, surrogate models based on DFT calculations. The broad adoption of informatics techniques in the atomic-scale modeling community suggests that it could serve as an example for how informatics could be used elsewhere.

In this article, we first discuss what has made informatics so successful in atomic-scale modeling. We start by identifying the major types of informatics studies used in atomic-scale modeling and assessing their impacts. We then focus on aspects of the data and the data infrastructure that have enabled the proliferation of informatics methods in atomic-scale modeling. Finally, we review the use of informatics by the broader materials research community and discuss the activities that may enable more frequent use of materials informatics.

Materials informatics and atomic-scale modeling

Materials informatics in atomic-scale modeling has included a large variety of approaches and spanned diverse types

Logan Ward, The University of Chicago, and Data Science and Learning Division, Argonne National Laboratory, USA; loganw@uchicago.edu
Muratahan Aykol, Toyota Research Institute, USA; murat.aykol@tri.global
Ben Blaiszik, The University of Chicago, and Data Science and Learning Division, Argonne National Laboratory, USA; blaiszik@uchicago.edu
Ian Foster, Department of Computer Science, The University of Chicago, and Data Science and Learning Division, Argonne National Laboratory, USA; foster@anl.gov
Bryce Meredig, Citrine Informatics, USA; bryce@citrine.io
James Saal, QuesTek Innovations LLC, USA; jsaal@questek.com
Santosh Suram, Toyota Research Institute, USA; santosh.suram@tri.global
doi:10.1557/mrs.2018.204

© 2018 Materials Research Society

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press

of materials. We can identify three major uses of informatics: (1) screening for discovering new materials, (2) learning new materials science from large data sets, and (3) applying machine learning to enable faster materials development.

Materials discovery by screening data sets

A common type of study that uses informatics is the screening of databases to find existing materials that suit a new application. Screening databases is hardly a new concept in materials engineering,[8] but the first examples of creation of "large" data sets (with hundreds to thousands of entries) with high-throughput (HT) *ab initio* computations appeared only in the early 2000s.[9–12] The first applications of these HT searches were to identify new stable materials among a small range of chemistry and structure types or to predict crystal structures via composition. Greeley et al.[10] identified BiPt as a promising electrocatalyst following a screening strategy based on electrochemical stability and activity of a series of binary alloys in one of the first examples of an HT search. With the upsurge in computing power and collaborative development of robust automation codes supporting first-principles calculations, such HT studies paved the way for the creation of several large high-fidelity databases of material properties, such as the Materials Project (MP),[13] Open Quantum Materials Database (OQMD),[14,15] and Automatic Flow for Materials Discovery Library (AFLOWLIB).[16]

With the introduction of these databases to the community, the number, scale, and complexity of material screening studies increased significantly. In a typical screening study, domain experts first identify the material features that are (1) descriptive of the functionality requirements for the target application and (2) computable using the information exiting in the databases or within their supporting computational framework.[6] Many materials are then screened through a funnel of "filters" to identify the most promising ones to explore with more advanced calculations (e.g., diffusion, transport) or via experiment. For example, researchers sifted through ~130,000 materials in the OQMD to find 30 functional cathode coatings (e.g., barriers that neutralize acids from the electrolyte) for Li-ion batteries, using descriptors that included their calculated reactivities with acids in the electrolyte.[17] Given the openness of these databases, HT-DFT data have supported similar studies by researchers unaffiliated with the database creators.[18] These studies exemplify the acceleration in materials discovery enabled by HT-DFT databases, where the initial, costly screening phase of material selection is offloaded to computers, allowing the experiments to focus on the most viable materials.

Improved understanding of materials and methods

Some important trends are only visible in large data sets. Thus, another common type of informatics study is to extract design rules from large collections of materials data. Using techniques such as visualization and automated data mining algorithms, materials engineers have employed atomic-scale calculation data to further scientific knowledge and generate useful heuristics to employ during materials design.

HT computational databases also enable better understanding of the accuracy of their underlying methods. Kirklin et al. found that DFT errors in computed formation energies are at the level of 0.136 eV/atom, but can be brought down to 0.081 eV/atom with systematic corrections of energies of the reference state, as part of the OQMD effort.[15] De Jong et al. compiled a database of elastic properties of inorganic materials as part of the MP, which helped them demonstrate that the bulk and shear moduli calculated using DFT are mostly within ±15% of measured values.[19] Systematic analysis of the performance of a method is only possible when large resources of consistent data are available, and these studies illustrate how such databases lead to better understanding of the limitations of DFT.

Large databases of materials properties also allow for better understanding of materials, especially when combined with methods for automatically learning important descriptors. Meredig et al. used polynomial regression coupled with clustering techniques to learn that different factors affect the stability of zirconia for different groups of elements in the periodic table.[20] Emery et al. used a database of DFT-computed stabilities of 5000 perovskites to learn that conventional rules for predicting perovskite stability are not sufficient.[21] Ghiringhelli et al. and Kong et al. have used algorithms that produce simple, human-interpretable models (e.g., compressed sensing, decision trees) to understand which crystal structure forms at different compositions.[22,23] Compressed sensing has also proven useful for understanding how structure effects the dielectric properties of polymers.[24] In general, these studies demonstrate the potential scientific gains to be achieved by aggregating data.

Accelerating discovery with machine learning

A third common application of materials informatics is the use of machine learning to further accelerate materials design. In contrast to learning design rules, the emphasis in machine-learning-driven studies is less on advancing human understanding than building a predictive surrogate model. Consequently, many machine-learning studies use learning algorithms that create models that are exceptionally accurate, but too complex to be easily interpreted by humans. After careful validation, these models can perform materials screening at rates far faster than conventional computational tools.

Machine learning in atomic-scale modeling is often used to replace expensive *ab initio* methods with cheaper approximations. For example, machine-learning methods are often used to create force fields,[25] which predict the energy of and forces acting on an atom given its environment and can be used as in-place alternatives for *ab initio* calculations in many applications. Several groups have demonstrated the utility of directly integrating the construction of machine learning force fields into a calculation method (e.g., molecular dynamics).[26,27] Machine-learning models are also commonly used to estimate the aggregate properties of a material, such as formation enthalpies,[28,29] given its atomic structure, at rates several orders of magnitude faster than the methods used to create their

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press

training set. These fast surrogates can be used to guide HT searches for materials, which make it possible to explore spaces too large for conventional methods.[30–33]

Machine-learning methods also expand the space of properties accessible to atomic-scale modeling. Seko et al. combined experimental data and DFT calculations to quickly predict melting temperatures of inorganic binary octet materials,[34] which takes thousands of CPU hours to compute by conventional atomic-scale modeling methods.[35] Hutchinson et al. have also shown how to accurately estimate bandgap energies—a property known to be difficult to accurately estimate with DFT—by combining the Materials Project and experimental data with transfer learning.[36] Combining machine learning, atomic-scale modeling, and experimental data may soon become a common pattern for accessing material properties currently impossible to compute from the atomic scale. These studies, among many other examples,[37] illustrate how machine learning is becoming an integral part of performing atomic-scale calculations.

Why is materials informatics so prevalent in atomic-scale modeling?

Several factors have led to informatics in atomic-scale modeling being so successful. First, one key feature of atomic-scale modeling data is that they are available in sufficient amounts to do meaningful data analysis. While advances in computing power certainly play a role in the wide availability of DFT data, being able to automate the calculations has also played a transformative role. Reliable DFT codes coupled with workflow automation tools make it possible for small teams of researchers to generate useful data sets. The community has also made the automation toolkits open source,[38–40] which allows them to be repurposed easily by others. Widely available automation drives down the resource cost for creating data, which is a prerequisite for data-driven science.

The atomic-scale modeling community also benefits from the high quality of their data resources. To be useful for informatics studies, data must be machine accessible, well documented, in a uniform format, and accessible through web application programming interfaces (APIs).[7] Web APIs, in particular, increase the usability of a database by allowing access to the data without the need to maintain a local copy of the database, and have facilitated the reuse of atomic-scale modeling data.

Atomistic simulation data have another necessary trait for informatics studies: documented provenance. When identifying subtle trends in materials behavior, it is necessary to know whether a finding is not an artifact of, for example, differences in the calculation methods. In other words, data require metadata describing its provenance to be useful. Extracting metadata from computational data is easy—the output files of codes often contain a listing of the inputs. This combination of self-documenting data and the uniformity intrinsic to data sets produced with automated tools means that atomic-scale data have sufficient metadata and similar enough provenance for users to easily understand whether two data sets are compatible, which makes them reusable.

Beyond the availability of substantial amounts of usable data, open-source software has also enabled the rapid development of machine-learning techniques in atomic-scale modeling. There is a rich suite of open-source software libraries for manipulating atomic-scale simulation data, such as ase,[41] qmpy,[15] and pymatgen.[42] The open environment of sharing software makes it easy for researchers to write code for analyzing large resources of atomic-scale calculation data, which has accelerated the adoption of materials informatics in atomic-scale modeling.

Materials informatics in experimental materials science

Materials informatics is not limited to atomic-scale modeling and is starting to have a transformative effect in materials science in general. Databases are a particularly prominent in crystallography, which features the International Center for Diffraction Databases,[43] the Inorganic Crystal Structure Database,[44] and the Crystallography Open Database,[45] among many others. Many of these crystallographic databases provide detailed records of $>10^5$ crystal structures and the methods by which those structures were determined, all in a format—the Crystallographic Information File (CIF)[46]—that has been a community standard for more than 25 years. The ASM and ACerS Handbooks, which feature the properties of thousands of different metallic alloys and ceramics, originated in the 1890s and 1930s, respectively.

Databases of experimentally measured properties of materials are also growing significantly. For example, NIMS in Japan has many databases of material properties, including polymer properties[47] and fatigue properties of metal alloys.[48] NIST in the United States hosts several databases of properties for material classes such as structural ceramics, oxide glasses, and superconductors.[49] Many existing databases are the result of curation by centralized authorities (e.g., national laboratories), but individual laboratories are beginning to curate data and make it available via online databases.[50,51] However, few of these databases are ready to use with informatics techniques because they lack the uniform data formats or APIs required for informatics software.

As in atomic-scale modeling, using data to drive materials design also has a long-standing tradition. For example, Zunger, Villars, and others have assembled databases of material properties to predict properties ranging from crystal structure and type of phase diagram to superconducting transition temperature—leading to a better understanding of materials and new design tools.[52–54] The CALculation of PHAse Diagrams (CALPHAD) approach—an underpinning of many successful computational design methods—relies on databases that are the product of extensive curation of experimental and computational data.[55]

There have been several important successes in the use of experimental data and machine learning to accelerate the discovery of new materials. For example, the thermoelectric property database developed by Gaultois et al.[51] was employed

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press

to discover Mn-Ru-Ge thermoelectrics.[56] A new class of oxygen evolution reaction catalysts was identified by employing machine learning to downselect from HT electrochemical screening data sets.[57] Active machine-learning techniques[58,59] coupled with experimental data have been used to discover better shape-memory alloys and new metallic glasses.[3,60] An array of machine-learning algorithms has been developed to automate phase mapping from x-ray diffraction data, which enables faster and automated extraction of composition–structure–property relationships.[61] With the continued growth of digital materials databases, the prevalence of materials screening and machine learning should only increase.

Expanding the impact of materials informatics

As described, materials informatics techniques combined with computational and experimental data can be impactful for materials science. How can we make such studies more prevalent, simple to conduct, and effective? The single greatest step that would expand adoption of materials informatics would be to make more data easily accessible to informatics. However, assembling well-described data sets of materials data is challenging, and especially so when using experimental data. Unlike atomic-scale modeling data, experimental data do not often start in digital formats that document provenance, and they often require even more metadata (e.g., processing conditions) to understand fully. Obtaining experimental data are expensive and time-consuming, making it difficult for individual researchers to create the volume of data needed to leverage informatics techniques. Furthermore, many existing collections of materials data are not digital and, therefore, are not readily usable. In this section, we discuss several short- and long-term efforts that will provide the ecosystem of tools necessary to facilitate broader use of materials informatics (as illustrated in **Figure 1**).

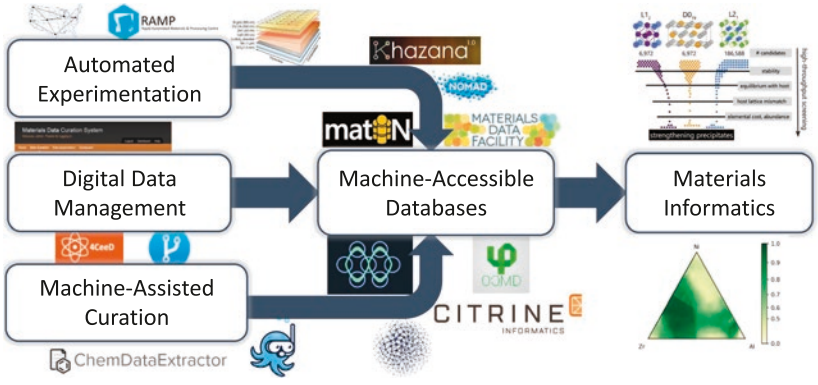


Figure 1. Our vision for the data infrastructure that will allow materials informatics to become more prevalent. The key requirement for materials informatics is a database that is machine-accessible (i.e., easy to use from software). We propose that the development and widespread use of automated experimentation tools, digital data management software (e.g., electronic lab notebooks), and automated methods for curating data will accelerate the growth and improve the availability of the well-curated data necessary for materials informatics.

Automating experiments

Exponential increase in computing power combined with open, community-developed software enabled the automation of computational workflows to produce data sets amenable to informatics. Physical experiments are, of course, significantly more challenging to automate than calculations. There is, however, substantial progress in high-throughput experimentation (HTE), which focuses on evaluating many materials in rapid succession and is uniquely suited to creating informatics-ready data—large resources of consistently measured and consistently processed samples. Beyond the technical limitations of automating certain types of experiments, adoption of HTE is slowed by limited access to HTE equipment, such as combinatorial synthesis machines (e.g., cosputtering chambers) or automated characterization equipment (e.g., robotic x-ray diffractometers). Consequently, collaborations that allow scientists to access a distributed ecosystem of HT synthesis, characterization, and data infrastructure are necessary[62] along with development of new HT experimental facilities.[63,64] Beyond HTE, the development of autonomous materials experimentation techniques capable of designing, executing, and analyzing their own experiments could accelerate the development of materials and the creation of useful data sets even further.[65] Increased use of HT and autonomous methods will produce richer materials data, which could lead to the same proliferation of informatics techniques in many fields as it has in atomic-scale modeling.

Digitizing materials data

Another issue limiting the use of materials informatics is that the metadata for most materials data is never recorded in a digital form. Unlike atomic-scale modeling data, metadata for experiments is rarely contained within the data files themselves, but instead is often placed in separate, nondigital resources (e.g., lab notebooks). A suite of materials-science-centric electronic lab notebook software and Laboratory Information Management Systems (LIMSs) that assist in the describing, managing, and sharing of data have been developed to address this problem.[66–68] Widespread use of these tools in materials science will reduce the amount of human effort required to make data useful for informatics.

Structuring materials data

Significant amounts of useful materials data are and will continue to be embedded in the text of research articles. Natural Language Processing (NLP) will play a key role in extracting and structuring the data to make them usable by materials informatics. Olivetti et al. recently aggregated synthesis parameters from more than 640,000 journal articles using NLP and demonstrated the use of these data in machine-learning-based synthesis planning.[69]

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press

Others have demonstrated automated extraction of small molecule and polymer properties, respectively.[70,71] The continued development of science-focused NLP tools (e.g., ChemDataExtractor[70]) and analogous systems for retrieving metadata from computer file systems[72] offer routes for drastically increasing the amount of machine-accessible materials data for informatics.

Improving data accessibility

Beyond being digital and structured, data must be easily sharable and accessible to software to enable informatics. As outlined by the FAIR Principles,[73] research requires data to be findable, accessible, interoperable, and reusable. Several efforts in the materials community have emerged to simplify the process of helping data meet the FAIR principles. For example, the Materials Data Facility provides a route for publishing, describing, and indexing materials data of any form or size on distributed data storage.[74] Citrination[75] allows scientists to publish data in a flexible, well-documented format, and provides easily usable machine-learning tools. Both services provide software interfaces to allow for automated data querying, processing, and access. The widespread adoption of materials informatics will require more databases to deploy machine-accessible interfaces and increased adoption of data publication practices. Even if not immediately useful, open data and machine-accessible databases lay the groundwork for future materials innovation via informatics.

Summary

Materials informatics is well on its way toward becoming a common and critically important means for materials scientists to increase the speed of materials discovery, enhance their characterization and analysis methods, and derive new understanding of material behavior. As highlighted in this article, informatics techniques such as screening databases and machine learning have made impacts across many disciplines of materials science, and these advancements are made possible by well-curated and machine-accessible databases. The key enabler for the continued expansion of materials informatics applications will be more structured, machine-accessible data sets. As the process for creating these databases is resource-intensive, we propose that software will play a key role in generating the data sets needed for informatics, whether by extracting data from research articles, aiding scientists in curating data as they are being created, or by performing experiments autonomously. The growing availability of materials data that such automation would spur and the continual advancements in informatics will continue to create revolutionary changes in the way materials are discovered and designed.

Acknowledgments

B.B., I.F., and L.W. were supported by financial assistance Award 70NANB14H012 from the US Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Material Design (CHiMaD), by the National Science Foundation as part of the Midwest Big Data Hub under NSF Award No. 1636950 "BD Spokes: SPOKE: MIDWEST: Collaborative: Integrative Materials Design (IMaD): Leverage, Innovate, and Disseminate," and by the US Department of Energy Contract DE-AC02-06CH11357.

References

1. J.M. Lutin, A.L. Kornhauser, E. Lerner-Lam, *ITE J.* **83**, 28 (2013).
2. D.C. Angus, *JAMA* **314**, 767 (2015).
3. F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, *Sci. Adv.* **4**, eaaq1566 (2018).
4. S.V. Kalinin, B.G. Sumpter, R.K. Archibald, *Nat. Mater.* **14**, 973 (2015).
5. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *NPJ Comput. Mater.* **3**, 54 (2017).
6. S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, *Nat. Mater.* **12**, 191 (2013).
7. L. Lin, *Mater. Perform. Charact.* **4**, 148 (2015).
8. M.F. Ashby, *Mater. Sci. Technol.* **5**, 517 (1989).
9. H. Smithson, C.A. Marianetti, D. Morgan, A. Van der Ven, A. Predith, G. Ceder, *Phys. Rev. B Condens. Matter* **66**, 144107 (2002).
10. J. Greeley, T.F. Jaramillo, J. Bonde, I.B. Chorkendorff, J.K. Nørskov, *Nat. Mater.* **5**, 909 (2006).
11. S. Curtarolo, D. Morgan, G. Ceder, *Calphad* **29**, 163 (2005).
12. G.K.H. Madsen, *J. Am. Chem. Soc.* **128**, 12140 (2006).
13. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, *APL Mater.* **1**, 11002 (2013).
14. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **65**, 1501 (2013).
15. S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, *NPJ Comput. Mater.* **1**, 15010 (2015).
16. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, *Comput. Mater. Sci.* **58**, 227 (2012).
17. M. Aykol, S. Kim, V.I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, *Nat. Commun.* **7**, 13779 (2016).
18. A. Jain, K.A. Persson, G. Ceder, *APL Mater.* **4**, 53102 (2016).
19. M de Jong, W. Chen, H. Geerlings, M. Asta, K.A. Persson, *Sci. Data* **2**, 1 (2015).
20. B. Meredig, C. Wolverton, *Chem. Mater.* **26**, 1985 (2014).
21. A.A. Emery, J.E. Saal, S. Kirklin, V.I. Hegde, C. Wolverton, *Chem. Mater.* **28**, 5621 (2016).
22. L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, *Phys. Rev. Lett.* **114**, 105503 (2015).
23. C.S. Kong, W. Luo, S. Arapan, P. Villars, S. Iwata, R. Ahuja, K. Rajan, *J. Chem. Inf. Model.* **52**, 1812 (2012).
24. A. Mannodi-Kanakkithodi, T.D. Huan, R. Ramprasad, *Chem. Mater.* **29**, 9001 (2017).
25. J. Behler, *J. Chem. Phys.* **145**, 170901 (2016).
26. Z. Li, J.R. Kermode, A. De Vita, *Phys. Rev. Lett.* **114**, 96405 (2015).
27. M.S. Jørgensen, U.F. Larsen, K.W. Jacobsen, B. Hammer, *J. Phys. Chem. A* **122**, 1504 (2018).
28. F. Faber, A. Lindmaa, O.A. von Lilienfeld, R. Armiento, *Int. J. Quantum Chem.* **115**, 1094 (2015).
29. L. Ward, R. Liu, A. Krishna, V.I. Hegde, A. Agrawal, A. Choudhary, C. Wolverton, *Phys. Rev. B Condens. Matter* **96**, 24104 (2017).
30. G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, *Chem. Mater.* **22**, 3762 (2010).
31. B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B Condens. Matter* **89**, 94104 (2014).
32. F.A. Faber, A. Lindmaa, O.A. von Lilienfeld, R. Armiento, *Phys. Rev. Lett.* **117**, 135502 (2016).
33. A. Seko, H. Hayashi, H. Kashima, I. Tanaka, *Phys. Rev. Mater.* **2**, 13805 (2018).
34. A. Seko, T. Maekawa, K. Tsuda, I. Tanaka, *Phys. Rev. B Condens. Matter* **89**, 54303 (2014).
35. Q.-J. Hong, A. van de Walle, *Calphad* **52**, 88 (2016).
36. M.L. Hutchinson, E. Antono, B.M. Gibbons, S. Paradiso, J. Ling, B. Meredig, *Comput. Sci. Learning* (2017), https://arxiv.org/abs/1711.05099.
37. L. Ward, C. Wolverton, *Curr. Opin. Solid State Mater. Sci.* **21**, 167 (2017).
38. K. Mathew, J.H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I. Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S.P. Ong, K. Persson, A. Jain, *Comput. Mater. Sci.* **139**, 140 (2017).
39. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, *Comput. Mater. Sci.* **111**, 218 (2016).

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press

40. T. Mayeshiba, H. Wu, T. Angsten, A. Kaczmarowski, Z. Song, G. Jenness, W. Xie, D. Morgan, *Comput. Mater. Sci.* **126**, 90 (2017).
41. A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, E.D. Hermes, P.C. Jennings, P. Bjerre Jensen, J. Kermode, J.R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K.S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K.W. Jacobsen, *J. Phys. Condens. Matter* **29**, 273002 (2017).
42. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013).
43. http://www.icdd.com/products/pdf4.htm.
44. A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch, *Acta Crystallogr. B Struct. Sci.* **58**, 364 (2002).
45. S. Grazulis, D. Chateigner, R.T. Downs, A.T. Yokochi, M. Quiros, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, A. Le Bail, *J. Appl. Crystallogr.* **42**, 726 (2009).
46. S.R. Hall, F.H. Allen, I.D. Brown, *Acta Crystallogr. A* **47**, 655 (1991).
47. S. Otsuka, I. Kuwajima, J. Hosoya, Y. Xu, M. Yamazaki, *2011 Int. Conf. Emerging Intell. Data. Web Technol.* (IEEE, 2011), pp. 22–29.
48. M. Yamazaki, Y. Xu, in *Volume 6: Materials and Fabrication, Parts A and B*, ASME 2009 Pressure Vessels and Piping Conference (ASME, 2009), pp. 1561–1568.
49. https://srdata.nist.gov/gateway/gateway.
50. B.L. DeCost, M.D. Hecht, T. Francis, B.A. Webler, Y.N. Picard, E.A. Holm, *Integr. Mater. Manuf. Innov.* **6**, 197 (2017).
51. M.W. Gaultois, T.D. Sparks, C.K.H. Borg, R. Seshadri, W.D. Bonificio, D.R. Clarke, *Chem. Mater.* **25**, 2911 (2013).
52. A. Zunger, *Phys. Rev. B Condens. Matter* **22**, 5839 (1980).
53. D.G. Pettifor, *Mater. Sci. Technol.* **4**, 675 (1988).
54. P. Villars, J. Phillips, *Phys. Rev. B Condens. Matter* **37**, 2345 (1988).
55. H. Lukas, S.G. Fries, B. Sundman, *Computational Thermodynamics* (Cambridge University Press, Cambridge, UK, 2007).
56. T.D. Sparks, M.W. Gaultois, A. Oliynyk, J. Brgoch, B. Meredig, *Scr. Mater.* **111**, 10 (2015).
57. S.K. Suram, J.A. Haber, J. Jin, J.M. Gregoire, *ACS Comb. Sci.* **17**, 224 (2015).
58. P.V. Balachandran, D. Xue, J. Theiler, J. Hogden, T. Lookman, *Sci. Rep.* **6**, 19660 (2016).
59. J. Ling, M. Hutchinson, E. Antono, S. Paradiso, B. Meredig, *Integr. Mater. Manuf. Innov.* **6**, 207 (2017).
60. D. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, *Nat. Commun.* **7**, 11241 (2016).
61. S.K. Suram, Y. Xue, J. Bai, R. Le Bras, B. Rappazzo, R. Bernstein, J. Bjorck, L. Zhou, R.B. van Dover, C.P. Gomes, J.M. Gregoire, *ACS Comb. Sci.* **19**, 37 (2017).
62. 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, *Appl. Phys. Rev.* **4**, 11105 (2017).
63. https://www.csiro.au/en/Research/MF/Areas/Chemicals-and-fibres/RAMP.
64. X.-G. Lu, *Sci. Bull.* **60**, 1966 (2015).
65. P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, B. Maruyama, *NPJ Comput. Mater.* **2**, 16031 (2016).
66. A. Dima, S. Bhaskarla, C. Becker, M. Brady, C. Campbell, P. Dessauw, R. Hanisch, U. Kattner, K. Kroenlein, M. Newrock, A. Peskin, R. Plante, S.-Y. Li, P.-F. Rigodiat, G.S. Amaral, Z. Trautt, X. Schmitt, J. Warren, S. Youssef, *JOM* **68**, 2053 (2016).
67. B. Puchala, G. Tarcea, E.A. Marquis, M. Hedstrom, H.V. Jagadish, J.E. Allison, *JOM* **68**, 2035 (2016).
68. P. Nguyen, S. Konstanty, T. Nicholson, T. O'Brien, A. Schwartz-Duval, T. Spila, K. Nahrstedt, R.H. Campbell, I. Gupta, M. Chan, K. McHenry, N. Paquin, *2017 17th IEEE/ACM Int. Symp. Cluster, Cloud and Grid Comput. (CCGRID)* (IEEE, 2017), pp. 11–20.
69. E. Kim, K. Huang, A. Tomala, S. Matthews, E. Strubell, A. Saunders, A. McCallum, E. Olivetti, *Sci. Data* **4**, 170127 (2017).
70. M.C. Swain, J.M. Cole, *J. Chem. Inf. Model.* **56**, 1894 (2016).
71. R.B. Tchoua, K. Chard, D.J. Audus, L.T. Ward, J. Lequieu, J.J. De Pablo, I.T. Foster, *2017 IEEE 13th Int. Conf. e-Science (e-Science)* (IEEE, 2017), pp. 109–118.
72. P. Beckman, T.J. Skluzacek, K. Chard, I. Foster, *Proc. 29th Int. Conf. Scientific Statistical Database Mgmt.—SSDBM '17* (ACM Press, New York, 2017), https://dl.acm.org/citation.cfm?doid=3085504.3091116.
73. M.D. Wilkinson, M. Dumontier, Ij.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.G. 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, *Sci. Data* **3**, 160018 (2016).
74. B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, I. Foster, *JOM* **68**, 2045 (2016).
75. J. O'Mara, B. Meredig, K. Michel, *JOM* **68**, 2031 (2016). □



Logan Ward is a postdoctoral researcher in the Computation Institute at The University of Chicago. He earned his PhD degree in 2017 from Northwestern University in materials science and engineering, where he studied the application of machine learning to the design of materials. His current research interests include the development of tools that simplify the use of machine learning and other data-driven techniques in materials science. Ward can be reached by email at loganw@uchicago.edu.



Muratahan Aykol is a research scientist in the Accelerated Materials Design and Discovery Program at the Toyota Research Institute (TRI). He received his PhD degree in materials science and engineering from Northwestern University, where he worked on high-throughput computational discovery of battery materials. Prior to TRI, he was the Materials Informatics and Infrastructure Postdoctoral Fellow at Lawrence Berkeley National Laboratory on the Materials Project team. He has authored more than 30 publications and his research interests include using artificial intelligence for materials discovery and addressing complex issues from synthesizability to degradation. Aykol can be reached by email at murat.aykol@tri.global.



Ben Blaiszik is a research scientist at the University of Chicago with an appointment in the Data Science and Learning Division at Argonne National Laboratory. He received his PhD degree in theoretical and applied mechanics from the University of Illinois at Urbana-Champaign studying self-healing polymers. He co-leads the Materials Data Facility effort, a National Institute of Standards and Technology Materials Genome Initiative project, to build and deploy data services to make materials data more easily publishable, discoverable, and reusable, regardless of data size and location. Blaiszik can be reached by email at blaiszik@uchicago.edu.



Ian Foster is a professor in the Department of Computer Science at The University of Chicago, where he teaches and supervises graduate students in the areas of distributed computing and data-driven discovery. Until 2016, he directed The University of Chicago/Argonne Computation Institute, which provided an intellectual home for large-scale interdisciplinary projects involving computation. He also holds a fellow appointment in the University's Institute for Molecular Engineering and is director of the Data Science and Learning Division at Argonne National Laboratory. Foster can be reached by email at foster@anl.gov.

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press



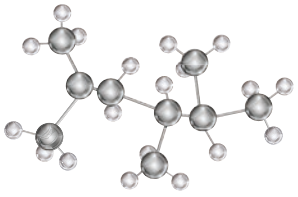
Bryce Meredig is cofounder and Chief Science Officer of Citrine Informatics, and leads many of the company's research and technology development activities. He earned his PhD degree in materials science from Northwestern University and his BAS and MBA degrees at Stanford University, where he is also on the faculty of the Department of Materials Science and Engineering. His research interest focuses on the application of machine learning to materials science. Meredig can be reached by email at bryce@citrine.io.



James Saal is a manager of technology at QuesTek Innovations LLC, where he is responsible for management and technical progress on all materials and model development activities. During his PhD degree at The Pennsylvania State University, he constructed CALPHAD databases for a variety of metal alloy, refractory, and ceramic oxide systems. Before joining QuesTek in 2014, he was a postdoctoral researcher at Northwestern University, aiding development of the Open Quantum Materials Database. He specializes in thermodynamic CALPHAD modeling and *ab initio* prediction of thermodynamic and kinetic properties with density functional theory. Saal can be reached by email at jsaal@questek.com.



Santosh Suram is a research scientist in the Accelerated Materials Design and Discovery Program at the Toyota Research Institute (TRI). He received his PhD degree in materials science and engineering from Iowa State University, with a focus on the application of materials informatics to atom probe tomography data sets. Prior to TRI, he was a research engineer in the High-Throughput Experimentation Group at the Joint Center for Artificial Photosynthesis, California Institute of Technology, where he worked on rapid discovery of materials for solar-fuels applications. He has authored more than 30 publications and his research interests focus on artificial intelligence for materials discovery, including automating materials characterization and theory-experiment integration. Suram can be reached by email at santosh.suram@tri.global.



SciVid Science in Video COMPETITION

We are seeking videos that can teach, explain, amaze, amuse or fascinate!

Generate a broader interest in materials science while competing for recognition and prizes. Submit a two-minute high-impact video about your research or a concept in materials science to SciVid. Viewing/voting will take place online and during the 2018 MRS Fall Meeting.

For additional information or to view last year's winning videos, visit
www.mrs.org/sci-vid

SUBMISSION OPENS:
September 1, 2018

DEADLINE:
October 15, 2018

CASH PRIZES:
1st place: $1000
2nd place: $500
3rd place: $300
People's Choice: $700

Co-sponsors:
MATERIALS RESEARCH SOCIETY FOUNDATION
Goodfellow
DREXEL UNIVERSITY A.J. Drexel Nanomaterials Institute

https://doi.org/10.1557/mrs.2018.204 Published online by Cambridge University Press