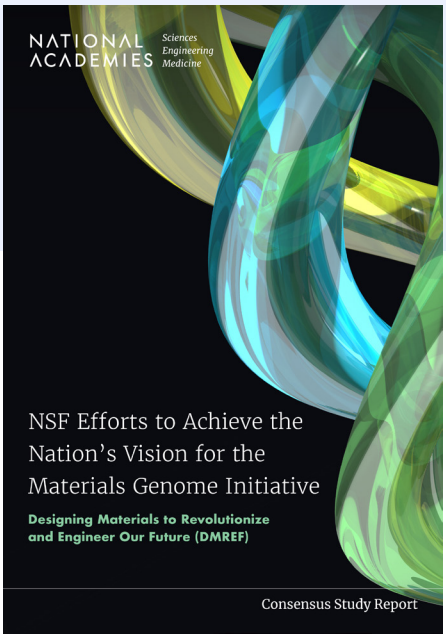
NATIONAL ACADEMIES Sciences Engineering Medicine

Consensus Study Report
Highlights
Data and Mathematics

NSF Efforts to Achieve the Nation's Vision for the Materials Genome Initiative



Designing Materials to Revolutionize and Engineer Our Future (DMREF)

Since the launch of the Materials Genome Initiative (MGI) in 2011, significant advances have been realized in computational materials science, data science, and materials informatics. The development of data-driven methods such as machine learning and artificial intelligence approaches, and their direct integration with computational or experimental discovery platforms, create an urgency for more centralized gathering of the data produced during research.

At the request of the National Science Foundation (NSF), the National Academies of Sciences, Engineering, and Medicine convened an expert committee to examine how NSF's Designing Materials to Revolutionize and Engineer Our Future (DMREF) program is helping achieve the nation's vision for the MGI. *NSF Efforts to Achieve the Nation's Vision for the Materials Genome Initiative* evaluates the DMREF program's goals, progress, and scientific accomplishments and highlights future opportunities to accelerate materials research from fundamentals to deployment. This document highlights some of the committee's recommendations focused on enabling the DMREF program to take full advantage of the latest capabilities in data and mathematics. The complete report and recommendations can be read or downloaded at https://nap.nationalacademies.org/catalog/26723.

THE RAPID PACE OF DATA GENERATION

Advances in basic materials science and computational modeling of materials behavior at all scales continue to be needed to improve the accuracy of predictions and the breadth of their applicability. **DMREF should continue to welcome projects that identify and**

fill gaps in computational approaches to achieve critically needed improvements in breadth, accuracy, and efficiency (Recommendation 6.1). Furthermore, **The MGI and DMREF initiative should continue to promote and accelerate a culture of reproducibility of computational procedures, sharable and reusable workflows, and interoperability among software tools (Recommendation 6.1).**

The research supported by DMREF is heavily reliant on a cutting-edge, high-performance computing infrastructure. Access to appropriate high-performance computing resources for DMREF research is becoming increasingly challenging as the materials research community expands. Thus**, for the United States to maintain its competitiveness and leadership in materials research, NSF should continue to innovate and expand U.S. high-performance computing infrastructures, including those based on next-generation computing platforms and those designed for material simulations (Recommendation 6.3).**

Certain advances in the materials discovery process are ripe for development, especially on the experimental side of research. Computational materials science, data science, and materials informatics typically predict a "window of materials space" where excellent or optimal functional properties could be obtained, but they are often not as well suited for detailed optimization of a material, which requires a fine-grained exploration of composition and processing space. Advanced tools using rapid and automated synthesis of advanced complex materials coupled with high-throughput characterization techniques are needed, as is collaboration with the private sector. To that end, **DMREF should enhance material discovery by revolutionizing experimental approaches to efficiently explore synthesis and processing parameter space through automated and autonomous processes that integrate synthesis, processing, characterization, analysis, and simulation (Recommendation 6.7).**

DATA STORAGE AND INFRASTRUCTURE NEEDS

At present, materials data are distributed among a number of servers and repositories. Although the National Institute of Standards and Technology maintains a Materials Data Repository, DMREF is uniquely situated to promote the integration of theory, modeling, experimentation, data science, robotics, and other novel areas in materials research.

To that end, the report also recommends **NSF and DMREF should take a leadership role to support the development of a national plan for a platform that creates interoperable systems that allow for comprehensive collection, dissemination, and use of computational and experimental materials data (Recommendation 6.4).**

Experimental materials science has some unique data challenges associated with its highly specialized instrumentation for characterization. Managing the sheer volume of data can constrain the important effort that should be devoted to curation, deep analysis, and scientific discovery from these large data sets, with the result that the data are not exploited as much as they could be. For the MGI-DMREF program to be truly successful, a conspicuous emphasis is required in DMREF projects to identify and report on any and all experimental factors, even if many of these factors are not variables for that particular work or publication. This will eventually allow meaningful assimilation of experimental data into larger databases that will be of greater use for theory, simulations, and materials informatics studies.

To empower DMREF researchers and the broader materials science community to adapt this and other best practices, it is critical to lower the barriers to data capture and curation. Thus, the report recommends that **DMREF should incentivize the deposition and curation of experimental data on synthesis, processing, and characterization of materials, and consider addressing scientific issues limiting the integration of heterogeneous data from diverse experimental and modeling sources, including uncertainty quantification, error propagation, verification and validation, and methods to create trust in the data set for subsequent engineering decisions (Recommendation 6.8).**

In addition to establishing and sustaining data repositories, there is a known need to facilitate the integration of distributed tools and platforms that generate data in federated systems. The policies implemented by the federal agencies through their investments may become catalysts to the community to cost-effectively implement findable, accessible, interoperable, and reusable (FAIR) data practices, adopt community-developed metadata standards, and establish a means to assess data quality. Looking to this future, this report also recommends that **DMREF should coordinate with other NSF programs and U.S. funding agencies to develop policies for handling data and a long-term plan for materials data infrastructure that will underpin the next generation of DMREF research (Recommendation 6.14).**

EARLY STEPS

Certain steps are suggested by the report for near-term developments that would be early measures in the long-term health of these data repositories. For example, the report recommends that for current and ongoing work, **the DMREF website should provide web links to all of the disparate databases where DMREF principal investigators store data and include descriptions of the database content (Recommendation 6.5).** The report also suggests that **NSF should consider supporting the development of infrastructure for effective collaboration (e.g., an electronic laboratory notebook platform that also serves as a collaboration platform, as well as data sharing), if a lack of infrastructure is identified. Such technology should not only enable effective communication between team members but also reduce the barrier in practicing the FAIR principles in data curation to optimize the output of collaborative science (Recommendation 6.21).**

COMMITTEE ON ADVISING NSF ON ITS EFFORTS TO ACHIEVE THE NATION'S VISION FOR THE MATERIALS GENOME INITIATIVE **Ronald Latanision (NAE),** Exponent, Inc., *Co-Chair*; **Karin M. Rabe (NAS),** Rutgers, The State University of New Jersey, *Co-Chair*; **Raymundo Arroyave,** Texas A&M University; **Gerbrand Ceder (NAE),** University of California, Berkeley; **Feliciano Giustino,** The University of Texas at Austin; **Amit Goyal (NAE),** State University of New York at Buffalo; **Olivia Graeve,** University of California, San Diego; **David Green,** University of Virginia; **Preeti Kamakoti,** ExxonMobil; **Robert V. Kohn,** New York University; **John Mauro (NAE),** The Pennsylvania State University; **Risto Nieminen (NAS),** Aalto University (emeritus professor); **Sherine Obare,** Western Michigan University; **Tresa Pollock (NAE),** University of California, Santa Barbara; **Aarti Singh,** Carnegie Mellon University; **Deidre Strand,** Wildcat Discovery Technologies; **Katsuyo Thornton,** University of Michigan; **Richard Vaia (NAE),** Air Force Research Laboratory; **Karen L. Wooley (NAS),** Texas A&M University

STAFF **Erik B. Svedberg,** Scholar and Study Director; **Michelle Schwalbe,** Director, National Materials and Manufacturing Board, and Director, Board on Mathematical Sciences and Analytics; **Neeraj Gorkhaly,** Associate Program Officer; **Amisha Jinandra,** Associate Program Officer; **Joseph Palmer,** Senior Project Assistant

FOR MORE INFORMATION

This Consensus Study Report Highlights was prepared by the National Academies' National Materials and Manufacturing Board based on the report *NSF Efforts to Achieve the Nation's Vision for the Materials Genome Initiative: Designing Materials to Revolutionize and Engineer Our Future* (DMREF) (2022).

The study was sponsored by the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

This Consensus Study Report is available from the National Academies Press (800) 624-6242 | http://www.nap.edu | http://www.nationalacademies.org

Division on Engineering and Physical Sciences

NATIONAL ACADEMIES *Sciences Engineering Medicine*