NISTIR 8364

Materials Data: A Landscape Analysis and Potential Roadmap for the NIST Material Measurement Laboratory

Debra L. Kaiser
Robert J. Hanisch
James A. Warren
Zachary T. Trautt

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8364

National Institute of
Standards and Technology
U.S. Department of Commerce

NISTIR 8364

Materials Data: A Landscape Analysis and Potential Roadmap for the NIST Material Measurement Laboratory

Debra L. Kaiser
Robert J. Hanisch
Office of Data and Informatics
Material Measurement Laboratory

James A. Warren
Material Measurement Laboratory

Zachary T. Trautt
Materials Measurement Science Division
Material Measurement Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8364

April 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce
for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Internal Report 8364
Natl. Inst. Stand. Technol. Intern. Rep. 8364, 27 pages (April 2021)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8364

Foreword

The Material Measurement Laboratory (MML) supports the NIST mission by serving as the national reference laboratory for measurements in the chemical, biological, and material sciences. Its activities range from fundamental and applied research on the composition, structure, and properties of industrial, biological, and environmental materials and processes, to the development and dissemination of tools including reference measurement procedures, certified reference materials, critically evaluated data, and best practice guides that help assure measurement quality.

Abstract

In the past decade, numerous public and private sector documents have highlighted the need for materials data to facilitate advanced technologies in myriad industrial and economic sectors. These documents have been analyzed to identify prevalent gaps in the establishment of an interconnected materials data infrastructure akin to that envisioned in the federal agency-wide Materials Genome Initiative. This internal report uses a uniform schematic format to portray these gaps, illustrate progress in addressing the gaps, and propose an MML roadmap of action items to further address the gaps.

Key words

Data; materials data; materials data infrastructure; artificial intelligence; Materials Genome Initiative

Table of Contents

List of Figures

1. Executive Summary

From the advent of the Materials-by-Design approach [1] nearly two decades ago to the inception of the Materials Genome Initiative (MGI) in 2011 [2] and the proliferation of Artificial Intelligence (AI) over the past few years [3], the need for materials data has grown explosively. A 2018 economic analysis [4] involving more than 100 industry experts cited access to high-quality data—non-proprietary experimental data, computational data, metadata, and software code—as pivotal to advanced materials innovation. In the past decade, more than 20 public and private sector documents representing the collective views of hundreds of experts have clearly demonstrated a strong need for open, digital data on materials. Such data are essential to discover, develop, manufacture, and deploy advanced materials and next-generation components, devices, and systems constructed from these materials, thereby bolstering economic security and human well-being [2]. Whereas the MGI and commercial efforts have spurred the development of some tools for creating and managing data, an analysis of the aforementioned documents reveals technical and sociological gaps to the realization of a pervasive, interconnected materials data infrastructure. This document provides both a materials data landscape analysis and a potential roadmap based on the prevalent infrastructure gaps in a schematic format. The same format is used to illustrate progress by external organizations and the NIST Material Measurement Laboratory (MML) in addressing these gaps, and to propose an MML roadmap of action items, including efforts in collaboration with other organizations. This roadmap could be extended to include materials data activities in other NIST Laboratories.

2. Motivation

MML generates a large quantity of materials data, both experimental and computational, and, through the MGI, has developed some key tools for data management. Commensurate with the NIST mission, much of MML's data are of commercial relevance. A recent economic impact study of the MGI [4] cited high-quality[1] materials data as a primary bottleneck to industrial innovation and estimated an annual economic benefit of an improved materials data infrastructure at \$123B to \$270B. Further, the expansion of machine learning (ML) and AI in virtually all materials-enabled technology sectors is predicated on the availability of high-quality materials data. It behooves MML to develop a roadmap for prioritizing its materials data activities that espouses the FAIR (findable, accessible, interoperable, and reusable) principles for data sharing [5]. The primary audience for this report is MML management and researchers, although others across NIST and outside NIST may find the report to be of interest.

3. Methodology

From the onset, it was clear that the MGI, AI, and open data and science would be important and synergistic components of a materials data landscape. The analysis of such a landscape is an essential prerequisite to develop a materials data roadmap. These observations led to the identification of 31 documents concerning the above topics and published in the past decade (see Appendix A), arguably a rich source of information representing the views of hundreds of subject matter experts. More than 20 of these documents were focused entirely on materials data or made explicit reference to the criticality of materials data, for example, in advanced manufacturing. 375 salient excerpts were hand-culled from the 31 reports,

---

[1] high-quality is equated with great accuracy and reproducibility, and low, quantifiable uncertainty

compiled, and labeled as Goals and Recommendations, Gaps and Challenges, or Observations. These excerpts were analyzed by three methods:

(1) Manual review of the 375 excerpts[2] to generate a list of prevalent gaps and challenges;
(2) Training a latent Dirichlet allocation [6] (probabilistic) topic model (on-line tool [7] developed at Cornell University) using 323 excerpts from 24 reports as input to generate a list of the 15 most associated groupings of words[3] (see Appendix B); and
(3) Applying a root- and rules-based natural language processing approach using the full content of the 31 reports to generate topic models and taxonomic structures (only a limited survey was produced due to prohibitive cost).

The analysis revealed that gaps and challenges introduced in a 2008 report [8]—coordination among stakeholder groups, materials data interoperability, and high-throughput experimentation, to name a few—were highlighted in reports a decade later, e.g., Refs. [9-11]. A materials data lifecycle graphic was created to provide a snapshot of the results from method (1) which are corroborated by methods (2) and (3). This graphic, shown in Fig. 1, is organized by five broad areas entitled Creating Data, Managing Data, Applying Data, Data Sharing, and Building Communities. Sub-areas were identified for the first three areas. A decision was made to focus on two prevalent applications of data—Innovation and ML/AI. The areas and sub-areas in Fig. 1 were populated with the most common, high-priority gaps and challenges in the 31 reports. These gaps and challenges were stated as desired attributes and components of a pervasive, fully interconnected and integrated materials data infrastructure.

4. External Scan

To gauge progress towards addressing the gaps and challenges identified in this study, an external scan (outside of NIST) was conducted with input from subject matter experts. The results of this external scan—a subset of organizations and projects—is presented in Fig. 2 using the same graphical format as Fig. 1. (Acronyms and initialisms in Figs. 2–4 are defined in Appendix C and references and websites are provided in Appendix D.) While numerous e-collaboration platforms, data repositories, resource registries, and other tools for managing data have been created and are accessible, interoperability is generally lacking. One coordinated effort, OPTIMADE [12], is underway to make density functional theory databases interoperable through development of a common Application Programming Interface.

5. Internal Scan

NIST, notably MML, has made significant progress in addressing the gaps and challenges (stated as desired attributes and components) in Fig. 1. Examples of work completed and in progress are presented in Fig. 3. Regarding creating data, NIST Standard Reference Data [13] are critically evaluated (i.e., are the highest quality data), and there are many other high-quality, published datasets discoverable via the NIST Science Data Portal [14]. The Materials Data Facility [15], established by the NIST-supported Center for Hierarchical Materials Design (CHiMaD) [16], also provides a venue for data publication. In the area of managing data, the NIST Materials Data Repository [17] has more than 320 entries and is a substantial

---

[2] A dataset of these excerpts may be found at: https://doi.org/10.18434/mds2-2307.

[3] One of the authors of this report (Z. Trautt) is developing Python code to train a similar topic model.

and diverse collection of data. The NIST Materials Resource Registry [18] also contains more than 320 entries and is interoperable with the Materials Data Facility registry. Among the other tools for managing data, the laboratory information management system (LIMS) for electron microscopy will have impact across the MML Divisions. NIST data are being applied for innovation through interactions with some of the Manufacturing USA Institutes [19]. There are dozens of ongoing projects in MML that are relevant to ML/AI; examples of recent publications can be found in Refs. [20–27]. Two notable community-building efforts led by NIST are the Phase Field Community Hub [28] and the High Throughput Experimental Materials Collaboratory [29].

6. Plan for Future Work

A proposed five-year roadmap of MML action items addressing the gaps and challenges in Fig. 1 is presented in Fig. 4, with year 1 (Y1) being FY2021 and year 5 (Y5) being FY2025. Action items to be conducted in collaboration with external partners are denoted in blue font. It is anticipated that these action items could be completed with the current level of STRS funding in the MML Divisions and in the MML Office of Data and Informatics (ODI). Several of these action items are highlighted below as demonstration pilots and projects that are aligned with the MML Focus Areas for FY2021.

- Within NIST
  - Continue support and development of a LIMS for electron microscopy for use across MML
  - Pilot at least one new LIMS effort in each of the six MML Divisions
  - Assess the state of data management in the MML Divisions through discussions with researchers, culminating in better-informed data management plans
  - Establish a NIST-wide additive manufacturing data repository
  - Develop benchmark datasets (experimental and computational) for training ML/AI models
- With external partners
  - Support the promulgation of the Materials Research Data Alliance[4] (MaRDA) in concert with other MGI federal agencies and with CHiMaD
  - Participate in MaRDA and other materials projects addressing high-priority gaps and challenges, such as the planned materials science pilot study for the Research Data Framework [30]
  - Develop a microstructures data toolkit to include a repository, descriptions of metadata for samples, processing, and imaging, a schema for data analytics, image analysis methods, ML models, and uncertainty characterization specifications

---

[4] https://web.archive.org/web/20210318184456/https://www.marda-alliance.org/

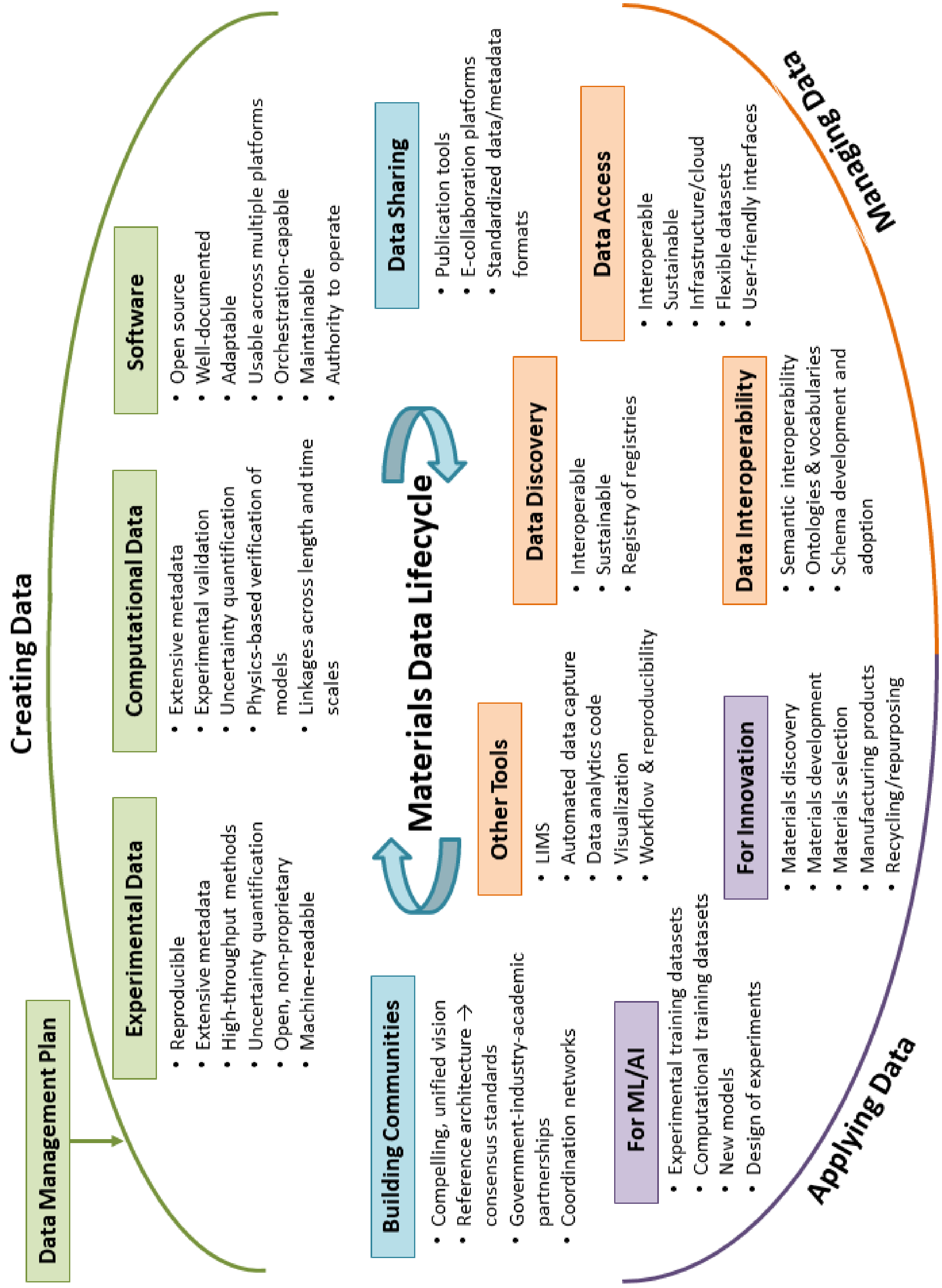
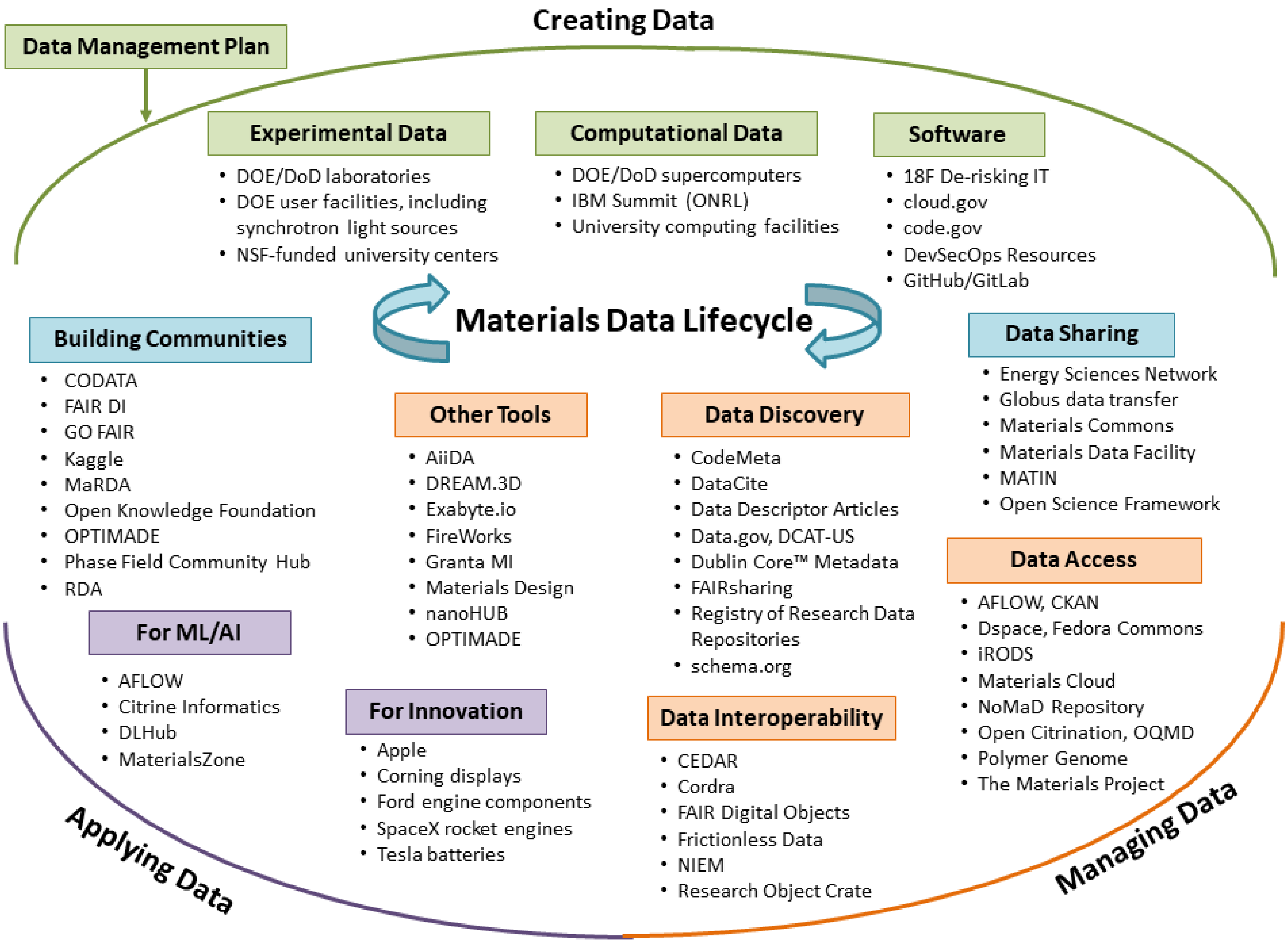


Fig. 1. Gaps and challenges stated as desired attributes and components



,

Fig. 2. Results of external scan. (Phase Field Community Hub was started by NIST.)

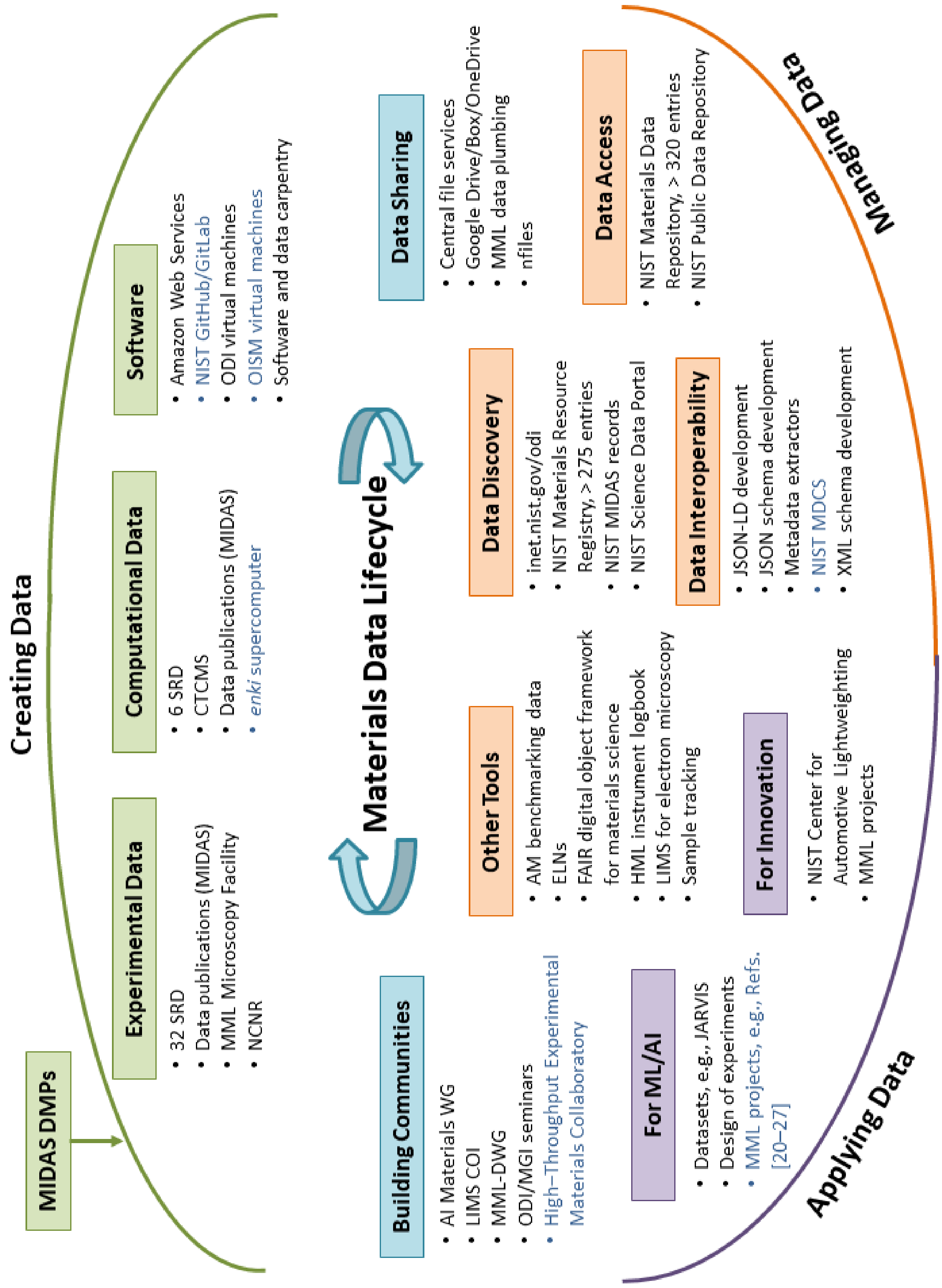


Fig. 3. Results of internal scan. Entries in blue text are collaborations with other NIST laboratories.

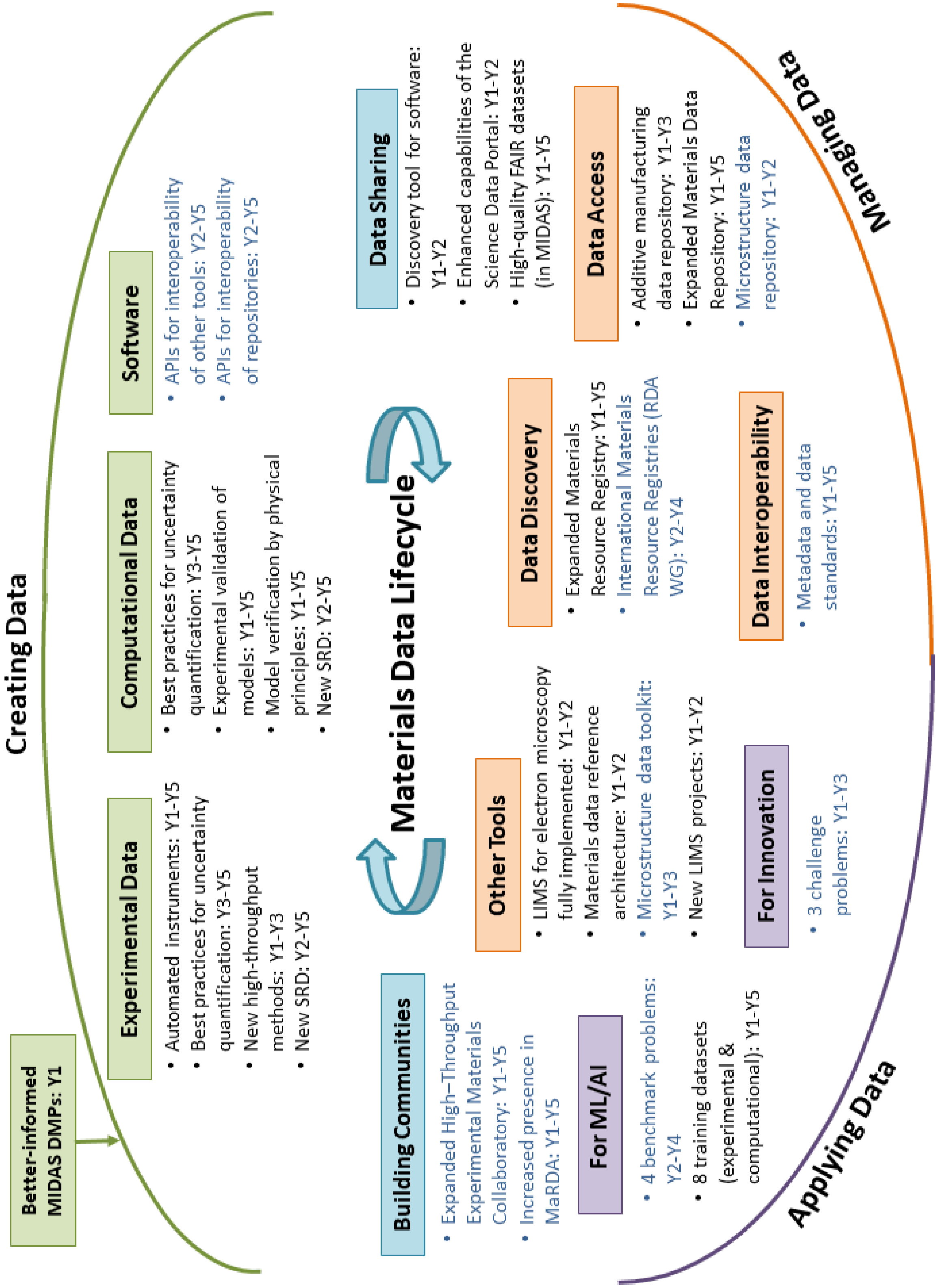


Fig. 4. Potential MML five-year roadmap of data activities. Entries in blue text are collaborations with external partners. Y1 is FY2021 through Y5 is FY2025.

References

[1] Olson GB (2000) Designing a new material world. *Science* 288:993-998. https://doi.org/10.1126/science.288.5468.993

[2] National Science and Technology Council, Committee on Technology (2011) Materials Genome Initiative for global competitiveness. Available at https://web.archive.org/web/20201110143912/https://mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf

[3] Executive Office of the President, National Science and Technology Council, Committee on Technology (2016) Preparing for the future of artificial intelligence. Available at https://web.archive.org/web/20210106122555/https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

[4] Scott T, Walsh A, Anderson B, O'Connor A, Tassey G (2018) Economic analysis of national needs for technology infrastructure to support the Materials Genome Initiative. Available at https://web.archive.org/web/20201213171449/https://www.nist.gov/system/files/documents/2020/02/06/MGI%20Final%20Report.pdf

[5] Wilkinson MD et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1):160018. https://doi.org/10.1038/sdata.2016.18

[6] Blei DM, Ng AY, Jordan M (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3:993-1022. Available at https://web.archive.org/web/20210111212854/https://dl.acm.org/doi/10.5555/944919.944937

[7] Mimno D (2019) jsLDA: in browser topic modeling. Available at https://web.archive.org/20200616203616/https://mimno.infosci.cornell.edu/jsLDA/index.html

[8] The National Research Council (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/12199

[9] Liu X, Furrer D, Kosters J, Holmes J (2018) Vision 2040: a roadmap for integrated, multiscale modeling and simulation of materials and systems. Available at https://web.archive.org/web/20210113020922/https://ntrs.nasa.gov/api/citations/20180002010/downloads/20180002010.pdf

[10] The Minerals, Metals & Materials Society (2018) Advanced computation and data in materials and manufacturing: core knowledge gaps and opportunities (The Minerals, Metals & Materials Society, Pittsburgh, PA). https://doi.org/10.7449/coreknowledge_1

[11] U.S. Department of Energy (2018) Workshop on artificial intelligence applied to materials discovery and design. Available at https://web.archive.org/web/20210112232020/https://www.energy.gov/sites/prod/files/2018/03/f49/AI%20Applied%20to%20Materials%20Discovery%20and%20Design_Workshop%20Summary%20Report.pdf

[12] Open Databases Integration for Materials Design Consortium (2020) OPTIMADE. Available at https://web.archive.org/web/20210112232248/https://www.optimade.org/index

[13] National Institute of Standards and Technology (2021) Standard Reference Data. Available at https://web.archive.org/web/20210112232429/https://www.nist.gov/srd

[14] National Institute of Standards and Technology (2021) NIST Science Data Portal, version 1.4.7. Available at https://data.nist.gov/sdp/#/

[15] Northwestern University (2020) The Materials Data Facility. Available at https://web.archive.org/web/20210112233013/https://materialsdatafacility.org/

[16] Northwestern University (2020) CHiMaD: Center for Hierarchical Materials Design. Available at https://web.archive.org/web/20201208145732/https://chimad.northwestern.edu/

[17] National Institute of Standards and Technology (2020) NIST Materials Data Repository. Available at https://web.archive.org/web/20210322155259/https://materialsdata.nist.gov/

[18] National Institute of Standards and Technology (2020) NIST Materials Resource Registry. Available at https://web.archive.org/web/20210113000011/https://materials.registry.nist.gov/

[19] Manufacturing USA (2021) Join the innovation revolution. Available at https://web.archive.org/web/20210113000212/https://www.manufacturingusa.com/

[20] Vasudevan RK et al. (2019) Materials science in the artificial intelligence age: high-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics. *MRS Communications* 9(3):821-838. https://doi.org/10.1557/mrc.2019.95

[21] Joress H et al. (2020) A high-throughput structural and electrochemical study of metallic glass formation in Ni-Ti-Al. *ACS Combinatorial Science* 22(7):330-338. https://doi.org/10.1021/acscombsci.9b00215

[22] Yang Y, Zheng M, Jagota A (2020) Learning to predict single-wall carbon nanotube-recognition DNA sequences. *npj Computational Materials* 5:3. https://doi.org/10.1038/s41524-018-0142-3.

[23] Mahynski NA, Mao R, Pretti E, Shen VK, Mittal J (2020) Grand canonical inverse design of multicomponent colloidal crystals. *Soft Matter* 16(13):3187-3194. https://doi.org/10.1039/C9SM02426C

[24] DeCost BL et al. (2020) Scientific AI in materials science: a path to a sustainable and scalable paradigm. *Machine Learning: Science and Technology* 1:033001. Available at https://iopscience.iop.org/article/10.1088/2632-2153/ab9a20

[25] Choudhary K et al. (2020) High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *npj Computational Materials* 6(1):1-13. https://doi.org/10.1038/s41524-020-0337-2

[26] Mahynski NA et al. (2020) Flat-histogram extrapolation as a useful tool in the age of big data. *Molecular Simulation*. https://doi.org/10.1080/08927022.2020.1747617

[27] Choudhary K et al. (2021) Computational scanning tunneling microscope image database. *Scientific Data* 8(1):1-9. https://doi.org/10.1038/s41597-021-00824-y

[28] National Institute of Standards and Technology (2020) PFHub: The Phase Field Community Hub. Available at https://web.archive.org/web/20201208050334/https://pages.nist.gov/pfhub/

[29] National Institute of Standards and Technology (2020) High–Throughput Experimental Materials Collaboratory. Available at: https://web.archive.org/web/20210113001519/https://www.nist.gov/programs-projects/high-throughput-experimental-materials-collaboratory

[30] Hanisch RJ, Kaiser DL, Carroll BC (2021) Research Data Framework (RDaF): motivation, development, and a preliminary framework core. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 1500-18. https://doi.org/10.6028/NIST.SP.1500-18

Appendix A: Reports

A1. The National Research Council (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and National security (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/12199

A2. Atkins D, Hey T, Hedstrom M (2011) National Science Foundation advisory committee for cyberinfrastructure task force on data and visualization. Available at https://web.archive.org/web/20170801194131/https://www.nsf.gov/cise/oac/taskforces/TaskForceReport_Data.pdf

A3. National Science and Technology Council, Committee on Technology (2011) Materials Genome Initiative for global competitiveness. Available at https://web.archive.org/web/20201110143912/https:/mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf

A4. The Minerals, Metals & Materials Society (2013) Integrated Computational Materials Engineering (ICME): implementing ICME in the aerospace, automotive, and maritime industries. (The Minerals, Metals & Materials Society, Warrendale, PA). Available at https://web.archive.org/web/20210113003549/http://d3em.tamu.edu/wp-content/uploads/2016/04/Report-TMS_icme_study_2013PDF.pdf

A5. National Science and Technology Council, Committee on Technology (2014) Materials Genome Initiative (MGI) strategic plan. Available at https://web.archive.org/web/20210101132245/https://www.mgi.gov/sites/default/files/documents/mgi_strategic_plan_-_dec_2014.pdf

A6. Bethel EW (2016) Report of the DOE workshop on management, visualization, and analysis of experimental and observational data – the convergence of data and computing. (Lawrence Berkeley National Laboratory, United States). Available at https://web.archive.org/web/20210322174231/https://escholarship.org/uc/item/3vf1w91z

A7. Wilkinson MD, Dumontier M, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. *Nature Scientific Data* 3(1):160018. https://doi.org/10.1038/sdata.2016.18

A8. National Science and Technology Council, Committee on Technology (2016) Preparing for the future of artificial intelligence. Available at https://web.archive.org/web/20210111211258/https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

A9. National Science and Technology Council, Committee on Technology (2016) The National artificial intelligence research and development strategic plan. Available at https://web.archive.org/web/20201227162352/https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

A10. U.S. Department of Energy (2016) Advanced manufacturing office multi-year program plan for fiscal years 2017 through 2021 (draft). Available at https://web.archive.org/web/20201116211034/https://www.energy.gov/sites/prod/files/2017/01/f34/Draft%20Advanced%20Manufacturing%20Office%20MYPP_1.pdf

A11. National Academies of Sciences, Engineering, and Medicine (2017) Fostering integrity in research (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/21896

A12. National Academies of Sciences, Engineering, and Medicine (2017) An assessment of the National Institute of Standards and Technology Material Measurement Laboratory: fiscal year 2017 (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/24975

A13. The Minerals, Metals & Materials Society (2017) Harnessing materials innovations to support next-generation manufacturing technologies. Available at https://web.archive.org/web/20210113013911/http://mforesight.org/projects-events/harnessing-materials/

A14. The Minerals, Metals & Materials Society (2017) Building a materials data infrastructure: opening new pathways to discovery and innovation in science and engineering (The Minerals, Metals & Materials Society, Pittsburgh, PA). Available at https://web.archive.org/web/20210322180143/https://www.nsf.gov/attachments/297691/public/Building-a-Materials-Data-Infrastructure-Study.pdf

A15. U.S. Department of Energy (2018) Workshop on artificial intelligence applied to materials discovery and design. Available at https://web.archive.org/web/20210112232020/https://www.energy.gov/sites/prod/files/2018/03/f49/AI%20Applied%20to%20Materials%20Discovery%20and%20Design_Workshop%20Summary%20Report.pdf

A16. U.S. Department of Commerce (2018) U.S. Department of Commerce strategic plan 2018-2022. Available at https://web.archive.org/web/20201230161937/https://www.commerce.gov/sites/default/files/2020-08/us_department_of_commerce_2018-2022_strategic_plan.pdf

A17. MForesight: Alliance for Manufacturing Foresight (2018) Manufacturing high entropy alloys: pathway to industrial competitiveness. Available at https://web.archive.org/web/20210113020227/http://mforesight.org/projects-events/hea-mfg/

A18. U.S. Department of Energy, Secretaría de Energía, Mexico, Canadian Institute for Advanced Research (2018) Materials acceleration platform: accelerating advanced energy materials discovery by integrating high-throughput methods with artificial intelligence. Available at https://web.archive.org/web/20201112015548/http://mission-innovation.net/wp-content/uploads/2018/01/Mission-Innovation-IC6-Report-Materials-Acceleration-Platform-Jan-2018.pdf

A19. National Academies of Sciences, Engineering, and Medicine (2018) Open science by design: realizing a vision for 21st century research (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/25116

A20. Liu X, Furrer D, Kosters J, Holmes J (2018) Vision 2040: a roadmap for integrated, multiscale modeling and simulation of materials and systems. Available at https://web.archive.org/web/20210113020922/https://ntrs.nasa.gov/api/citations/20180002010/downloads/20180002010.pdf

A21. Scott T, Walsh A, Anderson B, O’Connor A, Tassey G (2018) Economic analysis of national needs for technology infrastructure to support the Materials Genome Initiative. Available at https://web.archive.org/web/20201213171449/https://www.nist.gov/system/files/documents/2020/02/06/MGI%20Final%20Report.pdf

A22. National Science and Technology Council, Committee on Technology (2018) Strategy for American leadership in advanced manufacturing. Available at https://web.archive.org/web/20201119185350/https://www.whitehouse.gov/wp-content/uploads/2018/10/Advanced-Manufacturing-Strategic-Plan-2018.pdf

A23. The Minerals, Metals & Materials Society (2018) Advanced computation and data in materials and manufacturing: core knowledge gaps and opportunities (The Minerals, Metals & Materials Society, Pittsburgh, PA). https://doi.org/10.7449/coreknowledge_1

A24. H.R. 5759 (115th): 21st Century Integrated Digital Experience Act. Public Law 115-336; 2018 Dec. 20. Available at https://web.archive.org/web/20210113021428/https://www.congress.gov/bill/115th-congress/house-bill/5759

A25. H.R. 4174 (115th): Foundations for Evidence-Based Policymaking Act of 2018. Public Law 115-435; 2019 Jan 14. Available at https://web.archive.org/web/20210104103234/https://www.congress.gov/bill/115th-congress/house-bill/4174/text

A26. Baker N et al. (2019) Workshop report on basic research needs for scientific machine learning: core technologies for artificial intelligence (Department of Energy, United States). https://doi.org/10.2172/1478744

A27. National Academies of Sciences, Engineering, and Medicine (2019) Frontiers of materials research: a decadal survey (The National Academies Press, Washington, D.C.). https://doi.org/10.17226/25244

A28. National Science and Technology Council, Committee on Technology (2019) The National artificial intelligence research and development strategic plan: 2019 update. Available at https://web.archive.org/web/20210106122534/https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

A29. United States, Executive Office of the President Donald Trump (11 February 2019) Executive Order 13859: *Maintaining American Leadership in Artificial Intelligence. Federal Register*, Vol. 84, 14 Feb. 2019, pp. 3967-3972. Available at https://web.archive.org/web/20210322183203/https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

A30. The Minerals, Metals & Materials Society (2019) Verification & validation of computational models associated with the mechanics of materials. (The Minerals, Metals & Materials Society, Pittsburgh, PA). Available at https://web.archive.org/web/20210322184111/https://www.tms.org/portal/PUBLICATIONS/Studies/Verification___Validation/portal/Publications/Studies/VV/vv.aspx?hkey=bfc29011-1a4e-4534-8b89-08a8cc44becc

A31. The Computing Community Consortium, The Association for the Advancement of Artificial Intelligence (2019) A 20-year community roadmap for artificial intelligence research in the US. Available at https://web.archive.org/web/20201129014036/https://arxiv.org/ftp/arxiv/papers/1908/1908.02624.pdf

Appendix B: Results of Latent Dirichlet Allocation Topic Model

Output from the model in terms of groupings of associated words:

- standards government nist[i] agencies interoperability identify needed federal community technologies
- materials methods characterization machine design learning high-throughput discovery synthesis development
- models needed algorithms properties machine learning material physical integrate science
- data research tools access automated results model code platform applications
- industry tactic develop establish mdi[ii] academia programs federal mechanisms value
- existing experiments challenge simulations use key interfaces specific platforms robust
- community mdi mse[iii] sharing critical types resources among required work
- data digital infrastructure materials quality discovery sources information analysis integration
- manufacturing materials data tools support advanced technologies analytics infrastructure systems
- data datasets metadata principles challenge published important enable making variety
- data between knowledge open available including use sharing repositories useful
- computational models experimental materials modeling validation uncertainty tools process testing
- lack across conditions platforms both scale able types however challenges
- materials development science engineering icme[iv] research tools integrated software length
- data materials storage sharing lack long-term funding best practices limited

---

[i] NIST

[ii] materials data infrastructure

[iii] materials science and engineering

[iv] integrated computational materials engineering

Appendix C: Acronyms and Initialisms

Used in the Body of the Report (in alphabetical order)

AI	Artificial Intelligence
CHiMaD	Center for Hierarchical Materials Design
e-collaboration	Electronic collaboration
FAIR	Findable, Accessible, Interoperable, and Reusable
FY	Fiscal year
LIMS	Laboratory Information Management System
MaRDA	Materials Research Data Alliance
MGI	Materials Genome Initiative
ML	Machine Learning
MML	Material Measurement Laboratory
NIST	National Institute of Standards and Technology
ODI	Office of Data and Informatics
OPTIMADE	Open Databases Integration for Materials Design
STRS	Scientific and Technical Research Services
Y	Year

Used in Figs. 1-4 (in alphabetical order)

AFLOW	Automatic FLOW for Materials Discovery
AI	Artificial Intelligence
AiiDA	Automated Interactive Infrastructure and Database for Computational Science
AM	Additive Manufacturing
API	Application Programming Interface
CEDAR	Center of Excellence for Document Analysis and Recognition
CKAN	Comprehensive Knowledge Archive Network
CODATA	Committee on Data of the International Science Council
COI	Community of Interest
CTCMS	Center for Theoretical and Computational Materials Science (NIST)
DCAT-US	Data Catalog Vocabulary–United States
DevSecOps	Development, Security, and Operations
DLHub	Data and Learning Hub for Science
DMP	Data Management Plan
DoD	Department of Defense
DOE	Department of Energy
DWG	Data Working Group (MML-NIST)
ELN	Electronic Laboratory Notebook
FAIR	Findable, Accessible, Interoperable, and Reusable
FAIR DI	FAIR Data Infrastructure
GO FAIR	Global Open FAIR
Granta MI	Granta Material Intelligence
HML	Hollings Marine Laboratory (NIST)
iRODS	Integrated Rule-Oriented Data System
IT	Information Technology
JARVIS	Joint Automated Repository for Various Integrated Simulations

JSON	JavaScript Object Notation
JSON-LD	JSON for Linking Data
LIMS	Laboratory Information Management System
MaRDA	Materials Research Data Alliance
MATIN	Materials Innovation Network
MDCS	Materials Data Curation System (NIST)
MGI	Materials Genome Initiative
MIDAS	Management of Institutional Data Assets (NIST)
ML	Machine Learning
MML	Material Measurement Laboratory (NIST)
NCNR	NIST Center for Neutron Research
NIEM	National Information Exchange Model
NIST	National Institute of Standards and Technology
NoMaD	Novel Materials Discovery (Repository)
NSF	National Science Foundation
ODI	Office of Data and Informatics (NIST)
OISM	Office of Information Systems Management (NIST)
OPTIMADE	Open Databases Integration for Materials Design
OQMD	Open Quantum Materials Database
ORNL	Oak Ridge National Laboratory
RDA	Research Data Alliance
SRD	Standard Reference Data(bases) (NIST)
WG	Working Group
XML	Extensible Markup Language

Appendix D: References and Websites for Entries in the Figures

The following list of references and websites (for organizations and projects) are organized by **Fig. number** and ***areas*** (i.e., colored boxes). The entries in each area are in alphabetical order. Acronym and initialism definitions used in the figures are provided in Appendix C.

Figure 1

Software

Orchestration – National Institute of Standards and Technology (2021) Computer security resource center glossary, orchestration. Available at https://web.archive.org/web/20210319075840/https://csrc.nist.gov/glossary/term/Orchestration

Authorization to operate – National Institute of Standards and Technology (2021) Computer security resource center glossary, authorization to operate. Available at https://web.archive.org/web/20210319131116/https://csrc.nist.gov/glossary/term/authorization_to_operate

Figure 2

Computational Data

IBM Summit (ORNL) – Saltmarch A (2020 April 16) IBM's Summit—the supercomputer fighting coronavirus. *Medical Expo E-Mag*. Available at https://web.archive.org/web/20210117162921/http://emag.medicalexpo.com/summit-the-supercomputer-fighting-coronavirus/

Software

18F De-Risking IT – U.S. General Services Administration and 18F (2020) De-risking guide. Available at https://web.archive.org/web/20210219210936/https://derisking-guide.18f.gov/

cloud.gov – U.S. General Services Administration and 18F (2014) What is cloud.gov? Available at https://web.archive.org/web/20210317001224/https://cloud.gov/docs/overview/what-is-cloudgov/

code.gov – U.S. General Services Administration and 18F (2016) Sharing America's code. Available at https://code.gov/

DevSecOps – Forcepoint (2021) DevSecOps. Available at https://web.archive.org/web/20210317001915/https://www.forcepoint.com/cyber-edu/devsecops

GitHub – https://github.com/

GitLab – https://web.archive.org/web/20210316025104/https://about.gitlab.com/

Data Sharing

Energy Sciences Network – https://web.archive.org/web/20210303214713/https://www.es.net/

Globus data transfer – https://web.archive.org/web/20210317003958/https://www.globus.org/

Materials Commons – Center for PRedictive Integrated Structural Materials Science (2021) Materials Commons 2.0. Available at https://web.archive.org/web/20210319140142/https://materialscommons.org/

Materials Data Facility – Center for Hierarchical Materials Design (2019) Materials Data Facility. Available at https://web.archive.org/web/20210317005733/https://materialsdatafacility.org/

MATIN – Georgia Tech Institute for Materials (2021) Materials Innovation Network. Available at https://web.archive.org/web/20210317010728/https://matin.gatech.edu/

Open Science Framework – https://osf.io/

Data Access

AFLOW – http://aflowlib.org/

CKAN – https://web.archive.org/web/20210317125333/https://ckan.org/

DSpace – https://web.archive.org/web/20210317125618/https://www.dspace.com/en/inc/home.cfm

Fedora Commons – https://web.archive.org/web/20210304032409/https://duraspace.org/fedora/

iRODS – https://web.archive.org/web/20210317130503/https://docs.irods.org/4.1.5/

Materials Cloud – https://www.materialscloud.org/

NoMaD Repository – https://www.nomad-coe.eu/the-project/nomad-repository

Open Citrination – https://web.archive.org/web/20210317132450/https://citrine.io/research/open-citrination-platform/

OQMD – http://oqmd.org/

Polymer Genome – https://web.archive.org/web/20201017022516/https://www.polymergenome.org/

The Materials Project – https://web.archive.org/web/20210317170657/https://materialsproject.org/

Data Discovery

CodeMeta – https://web.archive.org/web/20210317171106/https://codemeta.github.io/

DataCite – https://web.archive.org/web/20210317172855/https://datacite.org/

Data Descriptor articles – Data Descriptor (2021) articles in *Scientific Data* (Springer Nature, New York, NY). Available at https://web.archive.org/web/20210317172108/https://www.nature.com/sdata/articles?type=data-descriptor

Data.gov – U.S. General Services Administration and 18F (2009) Data.gov. Available at https://web.archive.org/web/20210316044758/https://www.data.gov/

DCAT-US – Federal Enterprise Data Resources (2020) DCAT-US schema version 1.1 (Project Open Data Metadata Schema). Available at https://web.archive.org/web/20210319001117/https://resources.data.gov/resources/dcat-us/

Dublin Core™ Metadata (Initiative) – https://web.archive.org/web/20210318200824/https://dublincore.org/

FAIRsharing – https://web.archive.org/web/20210318131417/https://fairsharing.org/

Registry of Research Data Repositories – https://web.archive.org/web/20210318132130/https://www.re3data.org/

schema.org – https://web.archive.org/web/20210318132507/https://schema.org/

Data Interoperability

CEDAR – https://web.archive.org/web/20210311040518/https://cedar.buffalo.edu/

Cordra – https://web.archive.org/web/20210318133253/https://www.cordra.org/

FAIR Digital Objects – Schwardmann U (2020) Digital objects – FAIR digital objects: which services are required? *Data Science Journal* 19(15):1-6. http://doi.org/10.5334/dsj-2020-015

Frictionless Data – https://web.archive.org/web/20210318134057/https://frictionlessdata.io/

NIEM – https://web.archive.org/web/20210318134739/https://www.niem.gov/

Research Object Crate – https://web.archive.org/web/20210126204858/https://www.researchobject.org/ro-crate/

Other Tools

AiiDA – https://web.archive.org/web/20210209113340/https://www.aiida.net/

DREAM.3D – http://dream3d.bluequartz.net/

Exabyte.io – https://exabyte.io/

FireWorks – https://web.archive.org/web/20210318163108/https://materialsproject.github.io/fireworks/

Granta MI – Ansys (2021) Ansys Granta MI Enterprise. Available at https://web.archive.org/web/20210318164853/https://www.ansys.com/products/materials/granta-mi

Materials Design – https://www.materialsdesign.com/

nanoHUB – https://web.archive.org/web/20210318171049/https://nanohub.org/

OPTIMADE – https://web.archive.org/web/20210112232248/https://www.optimade.org/index

For Innovation

Apple – ARC Tehnica (2020) Here's why Apple believes it's an AI leader—and why it says critics have it all wrong. Available at https://web.archive.org/web/20210318172939/https://arstechnica.com/gadgets/2020/08/apple-explains-how-it-uses-machine-learning-across-ios-and-soon-macos/

Corning displays – Penaranda J, Fontaine R (2019) Artificial intelligence and machine learning are at the edge of what? Available at https://web.archive.org/web/20210318173335/https://www.corning.com/worldwide/en/products/display-glass.html

Ford engine components – Marr B (2019) The amazing ways the Ford Motor Company uses artificial intelligence and machine learning. Available at https://www.forbes.com/sites/bernardmarr/2019/05/17/the-amazing-ways-the-ford-motor-company-uses-artificial-intelligence-and-machine-learning/?sh=6e04b508e49a

SpaceX rocket engines – Dexlab (2020) How AI powers space missions like those of SpaceX's – a study. Available at https://web.archive.org/web/20210318174144/https://www.dexlabanalytics.com/blog/how-ai-powers-space-missions-like-those-of-spacexs-a-study

Tesla batteries – Oliver S (2020) Electric-car batteries get a boost from artificial intelligence. *The Wall Street Journal* (Dow Jones & Company, New York, NY). Available at https://web.archive.org/web/20210324115513/https://www.wsj.com/articles/electric-car-batteries-get-a-boost-from-artificial-intelligence-11604422792

For ML/AL

AFLOW – http://aflowlib.org/
Citrine Informatics – https://citrine.io/
DLHub – Li Z et al. (2021) DLHub: Simplifying publication, discovery, and use of machine learning models in science. *J. Parallel Distributed Computing* 147: 64-76. https://doi.org/10.1016/j.jpdc.2020.08.006
MaterialsZone – https://web.archive.org/web/20210117054044/https://www.materials.zone/

Building Communities

CODATA – https://web.archive.org/web/20210318182544/https://codata.org/
FAIR DI – https://web.archive.org/web/20210318183100/https://www.fair-di.eu/
GO FAIR – https://web.archive.org/web/20210318183244/https://www.go-fair.org/
Kaggle – https://www.kaggle.com/
MaRDA – https://web.archive.org/web/20210318184456/https://www.marda-alliance.org/
Open Knowledge Foundation – https://web.archive.org/web/20210318184747/https://okfn.org/
OPTIMADE – https://web.archive.org/web/20210112232248/https://www.optimade.org/index
Phase Field Community Hub – National Institute of Standards and Technology (2020) PFHub: The Phase Field Community Hub. Available at https://web.archive.org/web/20201208050334/https://pages.nist.gov/pfhub/
RDA – https://web.archive.org/web/20210318185907/https://www.rd-alliance.org/

Figure 3

Experimental Data

MIDAS – National Institute of Standards and Technology (2021) NIST Management of Institutional Data Assets. Available at https://midas.nist.gov/
MML Microscopy Facility – National Institute of Standards and Technology (2019) MML Microscopy Facility. Available at https://web.archive.org/web/20201214072609/https://www.nist.gov/mml/materials-science-and-engineering-division/mml-microscopy-facility
SRD – National Institute of Standards and Technology (2020) Standard Reference Data. Available at https://web.archive.org/web/20210112232429/https://www.nist.gov/srd

Computational Data

CTCMS – National Institute of Standards and Technology (2019) NIST/MML Center for Theoretical and Computational Materials Science. Available at https://web.archive.org/web/20210318192912/https://www.nist.gov/programs-projects/nistmml-center-theoretical-and-computational-materials-science
enki – National Institute of Standards and Technology (2019) Computation platform for AI/ML. Available at https://web.archive.org/web/20210318193057/https://www.nist.gov/programs-projects/computation-platform-aiml

Software

Amazon Web Services – https://web.archive.org/web/20210318193241/https://aws.amazon.com/compliance/nist/

NIST GitHub/GitLab – https://web.archive.org/web/20210318193526/https://github.com/usnistgov

ODI – National Institute of Standards and Technology (2021) Material Measurement Laboratory Office of Data and Informatics. Available at https://web.archive.org/web/20210318193756/https://www.nist.gov/mml/odi

OISM – National Institute of Standards and Technology (2021) Office of Information Systems Management. Available at https://web.archive.org/web/20210318193921/https://www.nist.gov/oism

Software and Data Carpentry – National Institute of Standards and Technology (2019) Software and Data Carpentry. Available at https://web.archive.org/web/20210318194126/https://www.nist.gov/programs-projects/software-and-data-carpentry

Data Access

NIST Materials Data Repository – National Institute of Standards and Technology (2020) NIST Materials Data Repository. Available at https://web.archive.org/web/20201206141006/https://materialsdata.nist.gov/

NIST Public Data Repository – National Institute of Standards and Technology (2021) About Public Data Repository. Available at https://data.nist.gov/pdr/about

Data Discovery

NIST Materials Resource Registry – National Institute of Standards and Technology (2021) NIST Materials Resource Registry, version 2.12.0. Available at https://web.archive.org/web/20210113000011/https://materials.registry.nist.gov/

NIST Science Data Portal – National Institute of Standards and Technology (2021) NIST Science Data Portal, version 1.4.7. Available at https://data.nist.gov/sdp/#/

Data Interoperability

JSON-LD development – https://web.archive.org/web/20210315025221/https://json-ld.org/

JSON schema development – https://web.archive.org/web/20210318194419/https://github.com/usnistgov/mgi-json-schema

Metadata extractors – https://web.archive.org/web/20210318194847/https://github.com/drewnoakes/metadata-extractor

NIST MDCS – National Institute of Standards and Technology (2019) NIST Materials Data Curation System. Available at https://web.archive.org/web/20210318195042/https://www.nist.gov/programs-projects/materials-data-curation-system

XML schema development – https://web.archive.org/web/20210318195243/https://github.com/usnistgov/hl7-schemas

Other Tools

AM benchmarking data – National Institute of Standards and Technology (2020) Additive manufacturing benchmark test series (AM–Bench). Available at https://web.archive.org/web/20210318195423/https://www.nist.gov/ambench/benchmark-test-data

FAIR digital object framework for materials – National Institute of Standards and Technology (2020) Facilitating the adoption of the FAIR digital object framework in materials science. Available at https://web.archive.org/web/20210318195542/https://www.nist.gov/programs-projects/facilitating-adoption-fair-digital-object-framework-material-science

HML – National Institute of Standards and Technology (2021) Hollings Marine Laboratory. Available at https://web.archive.org/web/20210318195756/https://www.nist.gov/mml/hollings-marine-laboratory

LIMS for electron microscopy – Lau JW, Devers RF, Newrock M, Greene G (2019) Laboratory information management systems for electron microscopy: evaluation of the 4ceed data curation platform. *NIST Journal of Research* 124:124034. https://doi.org/10.6028/jres.124.034; Taillon JA, Plante RL, Newrock MW, Lau JW, Greene GR (2020) NexusLIMS: Leveraging shared microscopy resources for data analysis with a configurable laboratory information management system. *Microscopy and Microanalysis* 26(S2):2950-2952. https://doi.org/10.1017/S1431927620023314

For Innovation

NIST Center for Automotive Lightweighting – National Institute of Standards and Technology (2020) NIST Center for Automotive Lightweighting. Available at https://web.archive.org/web/20210318195928/https://www.nist.gov/lightweighting

For ML/AI

JARVIS – Choudhary K et al. (2020) The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Computational Materials* 6:173. https://doi.org/10.1038/s41524-020-00440-1

Building Communities

High–Throughput Experimental Materials Collaboratory – National Institute of Standards and Technology (2020) High–Throughput Experimental Materials Collaboratory. Available at: https://web.archive.org/web/20210113001519/https://www.nist.gov/programs-projects/high-throughput-experimental-materials-collaboratory

Figure 4

Data Discovery

International Materials Resource Registries (RDA WG) – https://web.archive.org/web/20210318200140/https://www.rd-alliance.org/groups/working-group-international-materials-resource-registries.html