Networks and interfaces as catalysts for polymer materials innovation

Michael E. Deagen,[1] Dylan J. Walsh,[1] Debra J. Audus,[2] Kenneth Kroenlein,[3] Juan J. de Pablo,[4] Kaoru Aou,[5] Kyle Chard,[6,7] Klavs F. Jensen,[1] and Bradley D. Olsen[1]*

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[2]Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

[3]Citrine Informatics, Redwood City, California 94063, USA

[4]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA

[5]The Dow Chemical Company, 2301 Brazosport Blvd, Freeport, Texas 77541, USA

[6]Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA

[7]Data Science and Learning Division, Argonne National Laboratory, Argonne, IL 60439, USA

*E-mail: bdolsen@mit.edu

ORCID iDs:

MED: 0000-0002-8034-0667

DJW: 0000-0001-7981-2770

DJA: 0000-0002-5937-7721

KK: 0000-0002-0210-0203

JJdP: 0000-0002-3526-516X

KA: 0000-0002-3481-8287

KC: 0000-0002-7370-4805

KFJ: 0000-0001-7192-580X

BDO: 0000-0002-7272-7140

Summary

Autonomous experimental systems offer a compelling glimpse into a future where closed-loop, iterative cycles—performed by machines and guided by artificial intelligence (AI) and machine learning (ML)—play a foundational role in materials research and development. To reach their full potential, these systems need to incorporate information from their environment and interact with human researchers. This perspective draws attention to the roles of *networks* and *interfaces*—of and between humans and machines—for the purpose of generating knowledge and accelerating innovation. Polymers, a class of materials with everyday impact and a massive global footprint, present a unique opportunity for the scalable application of informatics and automation to pressing societal challenges. To develop these networks and interfaces in polymer science, the Community Resource for Innovation in Polymer Technology (CRIPT)—a polymer data ecosystem based on novel polymer data model, representation, search, and visualization technologies—is introduced. The ongoing co-design efforts engage stakeholders in industry, academia, and government working in experimental and computational polymer science to uncover rapidly actionable, high-impact opportunities to build networks and bridge interfaces. Through its aspirational goal as an open-source platform of digital tools and services tailored around polymers and their data, CRIPT aims to provide a foundational technology for artificial intelligence, computation, and robotics and serve as a catalyst for innovation in polymers.

Introduction & Background

Polymers play an essential role in everyday life, appearing in food packaging, water purification, clothing, shelter, medical products, electronics, and transportation.[1, 2] Properties that make polymers ideal for many applications—low cost, high strength-to-mass ratio, high chemical resistance, and low embodied energy to synthesize and process—also yield hidden, delayed, or otherwise unquantifiable societal costs as these materials accumulate and incompletely decompose in natural ecosystems.[3] Concerns around global material consumption have led to calls for a more circular economy.[4] Addressing the complexity and scale of the full life cycle of polymeric materials requires human ingenuity combined with the best automation and informatics tools at our disposal.

The very nature of polymers—stochastic ensembles of large molecular chains—make them difficult to represent from an informatics perspective.[5] Polymer properties vary across multiple length and time scales, and these properties often depend on subtle changes in composition, structure, process history, and environmental exposure. The complex interplay between chemistry, composition, structure, and processing that impedes recycling and recovery efforts[6, 7] also complicates the aggregation, comparison, and remixing of data captured within different contexts (Figure 1).[8, 9] Existing materials databases and repositories provide meticulously curated data,[10, 11] but capturing sufficiently detailed metadata while remaining relevant to a wide swath of stakeholders remains an ongoing challenge. Part of the challenge involves technological considerations around data (volume, velocity, variety, etc.),[12] but social considerations—demonstration of value, alignment with motivations, proven reliability—remain paramount in convincing generally skeptical individuals to invest the requisite time and energy in a given resource (e.g., using the system, organizing data, adding metadata). Many of these challenges can be distilled and recast through the lens of networks and interfaces. In the context of polymer informatics, researchers can benefit greatly from a platform that captures the chemistry, process, and property metadata while providing seamless integration with both manual (human–machine) and automated workflows.

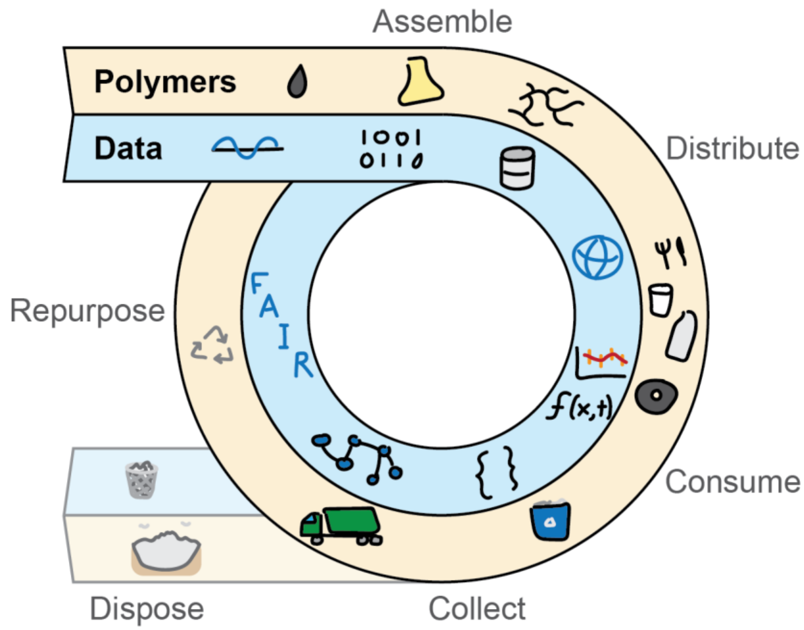

Figure 1: The life cycle of polymer materials and data. *Polymers form an essential part of our daily lives, yet they pose an increasing environmental threat as they steadily accumulate in landfills and other waste streams. From an informatics perspective, polymers also pose challenges in terms of relating their structure and processing. Maximizing the reusability of polymers and data requires non-trivial strategies, given that they are not always 'miscible.'*

An autonomous experimental system (AES), also known as a 'self-driving laboratory,' plans and carries out hypothesis testing in an iterative, closed-loop manner through novel use of robotics, computation, and artificial intelligence (AI) and machine learning (ML).[13-17] Demonstrations of autonomous experimentation in materials science include phase mapping; composition, process and property optimization; additive manufacturing design; and discovery of new materials systems.[18-23] Guided by AI algorithms, they often balance an "exploration–exploitation" tradeoff by weighing the benefits of exploration (maximizing the information gained by a given action) against exploitation (focusing on the highest-value regions) through approaches such as active learning.[24] The tireless nature of these systems combined with their robotic precision offers higher productivity and consistency compared to manual completion of similar tasks. By combining human ingenuity and the unique capabilities of machines into research networks, these closed-loop systems can leverage the best of both to form a collaborative intelligence.[25, 26]

As AES technologies advance, the need to train the next-generation workforce, facilitate collaboration across human stakeholders and machines, and encourage the sharing of data will only increase.[27-29] Sharing data across an organization can institutionalize expert knowledge for broader benefit, and inter-organizational sharing of data opens up opportunities for innovation through cross-pollinations of data, concepts, and ideas. The speed with which AESs incorporate and produce useful knowledge will impact the questions that researchers ask and how they go about answering those questions, providing an opportunity to unlock latent human potential. Reconfigurable implementations, such as flexible automation, show promise in this regard.[30]

This perspective highlights readily achievable, high-impact areas for accelerating the discovery and deployment of polymeric materials through the formation of information networks and bridging of interfaces (i.e., machine–machine, human–machine, human–human) to address global challenges (Figure 2). The early-stage development of a community-driven data ecosystem for capturing and sharing polymer knowledge is described. Such platforms are essential to build networks of information and people by providing streamlined interfaces for data and collaboration between humans and machines. Through this lens of networks and interfaces, the collaborative intelligence between humans and machines is explored, beginning with the transformation of data into knowledge.

Figure 2: Networks and interfaces as catalysts for polymer materials innovation. *Combining the mutual skills and capabilities of humans and machines requires attention to the networks formed between humans and machines, as well as improvements of their various interfaces (i.e., human–human, human–machine, and machine–machine), in order to translate research discoveries into lasting societal benefits.*

Transforming Data into Knowledge

The confluence of materials science and information science—expressed in frameworks such as the 'twin tetrahedra'[31]—can be described at a high level as a closed loop in which materials data (e.g., process, structure, properties) pass into information systems (e.g., through digital workflows and knowledge representations[32]) to generate actionable information to further improve those materials. The transformation of data into knowledge is an iterative process facilitated by the verification and distillation of information into robust models and expressive representations. Although digitalization has amplified the production and dissemination of information, humans still have limited time, attention, and capacity to accurately assess multi-dimensional problems. The volume of information generated by AESs will only exacerbate the need to filter and distill continuous streams of data and information into useful, actionable knowledge. Long before the Internet, the human desire to capture knowledge for posterity spurred the development of written language, the construction of libraries, and imaginings of mechanical contraptions that assimilate human thought.[33, 34] Today, AESs represent a novel thread in the age-old tapestry of knowledge generation, providing a powerful tool and new opportunities to free up human attention to pursue continued innovation.

The coupling of expert knowledge and intuition with AI recommendations can form a foundational component of an ecosystem for autonomous research. Just like humans, any given AES operates within a particular context and will be subject to biases depending on its configuration, constraints, and environment. The potential for bias is exacerbated in the case of polymers, whose properties are highly sensitive to subtle differences in preparation and analysis. The fusion of theory into models provides one way to mitigate biased data and enhance the ability to navigate complex design spaces.[19, 35-37] Another way involves providing these disparate systems with a shared ecosystem for communicating and learning from experience in different contexts through collaborative networks.

Leveraging Networks to Achieve Scale

While autonomous research systems and human researchers may have the capacity to operate independently, a collaborative intelligence that combines the talents of humans and machines can unlock the value generated through network effects (e.g., Metcalfe's Law[38]). The incremental value added by each new node in a network exceeds the incremental cost of adding the node to the network, leading to super-linear scaling in innovation.[39] The World Wide Web, through its distributed network architecture and ability to reference resources through hyperlinks, has brought about significant areas for value creation. Networks of individuals and organizations have spatial components that impact the flows of tacit and codified knowledge among local clusters and through global pipelines.[40]

The development of open standards forms a foundation for unlocking the benefits of networks. Positive feedback loops for value generation are balanced by negative feedback loops when considering competing between networks—why participate in network B when network A is larger and offers more value today?—which can ultimately stifle innovation.[41] To counter such winner-take-all, lock-in scenarios, open standards provide a basis for healthy competition and yield net benefits for stakeholders.[42]

In the international scientific community, the notion of FAIR (findable, accessible, interoperable, reusable) guiding principles for data stewardship has gained steam and offers a shared understanding of the importance of data practices in materials research.[29, 43] These guiding principles argue for the use of applicable standards as well as annotation with machine-interpretable metadata. Essentially, a 'FAIR' approach to data management offers a way to extract value from research investment by extending the shelf-life of research outputs (i.e., data).

Efforts to facilitate the growth of a network should also consider ways to mitigate the negative aspects that arrive with scale. For example, the same super-linear scaling applies to negative socioeconomic aspects as well, including increasing the potential reach of bad actors, irrational behavior caused by groupthink, or a lack of perceived individual responsibility. While complex social dynamics plays a role, networks of machines and algorithms could potentially suffer from emergent biases, for example amplifying existing biases in their data.[44] Establishing proper checks and balances into the structure of these networks becomes important when considering their long-term sustainability. For example, human-centered AI emphasizes the role of human control in the development of reliable, safe, and trustworthy automation.[45] Finally, value generated within a network can form a divide between those who participate and those who do not. Acknowledging these potential disparities, and addressing them through interfaces, can strengthen the overall network through increased overall participation.

Bridging Interfaces of All Types

The addition of a node to a network results in the potential creation of multiple interfaces and their associated barriers to communication, so the engineering of interfaces becomes increasingly important as a network scales in size and complexity. Consideration of the various interfaces entailed in future of AES-facilitated research (i.e., machine–machine, human–machine, human–human) reveals numerous readily achievable, high-impact opportunities to lower these barriers, incorporate different perspectives, and serve as a catalyst for innovation.

Machine–machine interfaces involve careful architectural design and the application of appropriate protocols and data representations. Driven by application programming interfaces (APIs), these interfaces should consider performance (e.g., minimal latency and network bandwidth, portability, robustness) while maximizing the independence and scalability of components within the network. For example, Representational State Transfer (REST) describes constraints including uniform interface, stateless behavior, independent layers, and client-server communication that form essential aspects of the modern Web.[46] For platforms aiming to promote interoperability among autonomous systems, a REST API provides the necessary communication layer for these systems to exchange information, and developers can employ REST to create scalable applications through a microservices architecture. Beyond these considerations about the flow of data between machines, the data representations employed (e.g., schema-oriented, ontology-based) play an important role by implicitly determining the expressive capabilities of the interface.

Human–machine interfaces remain a critical aspect of autonomous systems and the focus of a large area of research in human–computer interaction. One must recognize the role of humans not only during the course of their operation, but in their iterative development and implementation as well.[45, 47] For example, mixed-initiative user interface design considers the coupling of human control through direct manipulation with systems that provide automated services.[48] Representations for communicating and accurately portraying research findings to human researchers play a major role in facilitating this interface, including data visualization and interaction design.[49, 50]

Human–human interfaces (i.e., technically rich communication) form the basis of collaborative research and development, bringing together contributors with diverse perspectives to drive innovation and push the limits of scientific research. Enabling people to contextualize and apply knowledge involves making it easier to quantify observations, their context, and their reliability; create distilled and expressive representations of findings; and get those findings in front of others who can augment or refine them. For example, an effective figure or elegant mathematical expression can impart understanding much more rapidly than a verbose paragraph describing the same concept. Computational notebooks show promise as a medium for communicating interactively through the integration of code and prose.[51] Other emerging technologies such as augmented reality (AR) offer new ways create shared representations and reason over virtual models within physical spaces.[52] New methods and media may emerge, but the essential goals of these tools to augment human thought processes and communication will continue.

For the scientific community to address the pressing needs of today, it should consider how these various interfaces impact the networked system as a whole. These opportunities are readily actionable in the sense that many improvements can be made by appropriately configuring existing technologies, and they are high-impact because they affect the growth and long-term sustainability of these systems. Sustainability of these networks and their interfaces becomes critical in order to address the persistent and global scale of polymers, a multi-billion-dollar industry that produces hundreds of millions of metric tons per year, the majority of which currently ends up in landfills.[53]

Community Resource for Innovation in Polymer Technology

The Community Resource for Innovation in Polymer Technology (CRIPT)—a collaboration between stakeholders in industry, academia, and government—aims to provide a platform for researchers in polymer science and engineering to capture and share knowledge. An ecosystem such as CRIPT provides a digital infrastructure to capture the complexities of polymer data and metadata (including necessary chemical detail) in a structured and searchable way (Figure 3). One key philosophy of CRIPT is to 'meet researchers where they are at,' allowing them focus on the most relevant data for their use case and providing them with tools to augment their data with appropriate metadata. Through user interfaces and data standards, this collaborative platform enables the networking and interfacing of humans and machines for the broader purpose of accelerating materials discovery to solve key societal challenges around the design and repurposing of polymers.

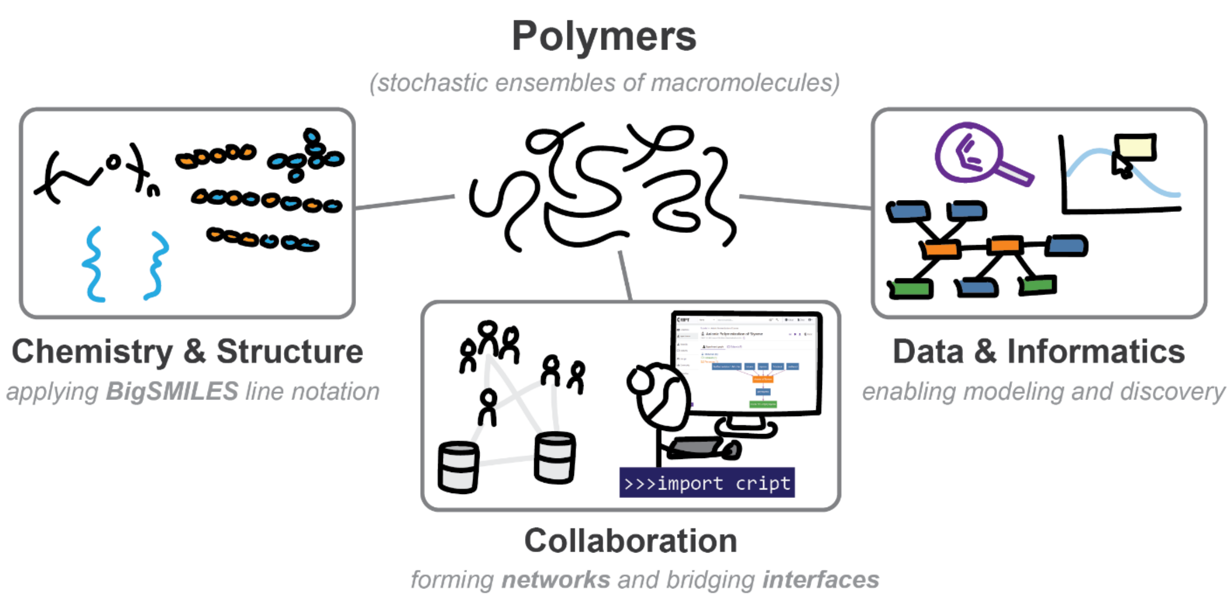

Figure 3 CRIPT as a facilitator of networks and interfaces for polymer informatics. *The vision for CRIPT is an open-source ecosystem of data and software tailored for polymer materials. These objectives will be realized by enabling collaboration among stakeholders, linking polymer representations to detailed metadata, and offering multiple interfaces and on-ramps for stakeholders to harness the power of polymer informatics.*

Polymeric materials pose unique challenges around their specification and representation that necessitate the development of a dedicated data ecosystem. Informatics approaches for small molecules have limited applicability to polymeric materials because polymers are stochastic, form complex topologies, and have properties that depend strongly on processing. Such challenges present a need to innovate and develop tools that extend the capabilities of cheminformatics. These tools must capture the variety of experimental and computational data with chemical fidelity within a flexible data model such that modern cheminformatics methods can be brought to bear on polymer problems.

Key technological precursors to CRIPT include a novel way to represent polymers in a database, as well as a structured approach to linking polymeric structure to characterization data. First, the BigSMILES line notation captures a compact representation of a polymer through the notion of stochastic objects as well as specification of the bonding constraints for these repeating units.[54] The PolyDAT data schema provides a file format for capturing reaction and characterization data as structured JSON,[55] and this notion is currently being extended to capture relationships between materials, processing, and data within a graph-based data model. Additional efforts in canonicalization provide a way to efficiently reference polymers in a database, whose multiplicity otherwise impedes such efforts.[56]

At the time of writing (July 2022), the CRIPT platform exists in a pre-release development phase with a limited group of early adopters in academia, industry, and government settings. The involvement of multiple stakeholders has been an essential aspect of the co-design and co-creation of the platform.[57] Through interviews and focus groups, these stakeholders have provided valuable input and insight through their unique perspectives and variety of data-related requirements. The resulting CRIPT structure uses a REST API with a Python software development kit (SDK) to enable programmatic access to data along with a web-based graphical user interface (GUI) for organizing and interacting with data visually, two key human–machine interfaces within the platform. The programmatic interface enables developers to augment existing interfaces (e.g., data ingestion, workflow automation), while the GUI provides an important endpoint for visual validation of uploaded data and easy browsing. To engender trust in data uploaded to CRIPT, organization- and institution-independent ORCID identifiers are used as the basis for user authentication.[58] The system integrates with Globus[i] to manage authentication and file storage, with the eventual goal of direct integration with characterization instruments across user facilities.[59, 60] Current and future efforts aim to deliver additional value through AI-based data validation, more automated data ingestion workflows, integration with robotic research platforms, and the ability to configure private instances of CRIPT for stakeholders in industry with strict data sharing policies.

The long-term vision of CRIPT as a community-driven digital ecosystem for polymer innovation hinges on open-source code contributions, robust data pipelines, and clear articulation of value to researchers. First, the Python SDK enables programmers outside the core development team to develop application-specific software and share with others in the CRIPT community who may adapt and improve upon this software to further enrich the data resource. Second, the data model strikes a balance between configurability and rigidity that accommodates a variety of use cases while still promoting unified annotation with metadata. Finally, the web application aims to make it easier for a given researcher to accomplish their work, providing incentive to invest the necessary time and attention to data management beyond the minimum requirements of a funding mandate. By empowering individual polymer researchers to manage their data, CRIPT will play an important role in the propagation of polymer knowledge.

Measurable value generation of a platform such as CRIPT can be defined through the time and cost savings from using the platform compared to existing alternatives. For example, a tool that can parse and suggest a template for visualizing a dataset saves minutes of time otherwise spent configuring a plotting tool. Hours otherwise spent combing the literature for the properties of a polymer—a task that may be avoided entirely due to the associated cost—can be greatly reduced through structure-based searches linked to verifiable property metadata. Identifying useful targets for synthesis alone could save months of research effort. Other, less easily measurable forms of value may include the spark of insight that results in a fruitful new research direction, a new perspective that deepens current understanding of physics, or a new process that slows down or reverses the current problems of plastic waste. The fog at the current frontier of polymer informatics will lift as new discoveries and innovations are made and shared—one network and interface at a time—ultimately paving the way for the next frontier.

Conclusions and Perspectives

Digital ecosystems such as CRIPT will play a critical role in the future development of AESs. First, an open, standard communication protocol is a foundational technology for autonomous systems to send and receive information. Second, the perspective of CRIPT to 'meet researchers where they are at' in terms of data practices and requirements instills an important human element to the platform. Networks and interfaces, such as those enabled by CRIPT, form a critical part of the broader fundamental shift in division of labor suggested by the current trajectory of autonomous research. The notion of human-centered AI for reliable, safe, and trustworthy autonomous systems comes into play here,[45] where thoughtful design provides humans with requisite control mechanisms to effectively guide the autonomous systems appropriately and offers checks and balances as these networks of humans and machines grow in scale. Iterative development of such systems becomes possible by building stakeholder relations, involving consumers in the co-design of the platform and its features, and recognizing the role of humans in the networks and interfaces of these systems. A well-designed artificial intelligence should feel natural to interact with while simultaneously supporting the development of human skill and mastery.

Advancements in robotics, algorithms, and computation will continue to accelerate materials discovery, while networks and interfaces (human and machine) facilitated by open informatics platforms such as CRIPT will act as catalysts to fully unlock this potential. Integration of autonomous experimentation into the typical materials research workflow involves paradigmatic shifts in the way knowledge is generated, validated, and disseminated. By focusing on readily actionable, high-impact opportunities in forming networks and bridging their interfaces, polymer informatics will produce transformative and lasting societal benefit.

Acknowledgments

This work is primarily supported by the National Science Foundation Innovation and Technology Ecosystems (NSF 2040636) Convergence Accelerator program.

Author Contributions

Conceptualization, M.E.D, D.J.W., B.D.O.; Writing – Original Draft, M.E.D; Writing – Review & Editing, M.E.D., D.J.W, D.J.A., K.K., J.J.dP., K.A., K.C., K.F.J., B.D.O.

Declaration of Interests

The authors declare no competing interests.

References

1. Geyer R, Jambeck JR, Law KL. Production, use, and fate of all plastics ever made. Sci Adv. 2017;3(7): e1700782. doi: 10.1126/sciadv.1700782.
2. Jin FL, Li X, Park SJ. Synthesis and application of epoxy resins: A review. J Ind Eng Chem. 2015;29: 1-11. doi: 10.1016/j.jiec.2015.03.026.
3. Zalasiewicz J, Waters CN, do Sul JAI, Corcoran PL, Barnosky AD, Cearreta A, et al. The geological cycle of plastics and their use as a stratigraphic indicator of the Anthropocene. Anthropocene. 2016;13: 4-17. doi: 10.1016/j.ancene.2016.01.002.
4. Kirchherr J, Reike D, Hekkert M. Conceptualizing the circular economy: An analysis of 114 definitions. Resour Conserv Recy. 2017;127: 221-32. doi: 10.1016/j.resconrec.2017.09.005.

5. Audus DJ, de Pablo JJ. Polymer Informatics: Opportunities and Challenges. ACS Macro Lett. 2017;6(10): 1078-82. doi: 10.1021/acsmacrolett.7b00228.
6. Ragaert K, Delva L, Van Geem K. Mechanical and chemical recycling of solid plastic waste. Waste Manage. 2017;69: 24-58. doi: 10.1016/j.wasman.2017.07.044.
7. Coates GW, Getzler YDYL. Chemical recycling to monomer for an ideal, circular polymer economy. Nat Rev Mater. 2020;5(7): 501-16. doi: 10.1038/s41578-020-0190-4.
8. Patra TK. Data-Driven Methods for Accelerating Polymer Design. ACS Polym Au. 2021;2(1): 8-26. doi: 10.1021/acspolymersau.1c00035.
9. Sha WX, Li Y, Tang S, Tian J, Zhao YM, Guo YQ, et al. Machine learning in polymer informatics. InfoMat. 2021;3(4): 353-61. doi: 10.1002/inf2.12167.
10. Chen LH, Pilania G, Batra R, Huan TD, Kim C, Kuenneth C, et al. Polymer informatics: Current status and critical next steps. Materials Science & Engineering R-Reports. 2021;144. doi: 10.1016/j.mser.2020.100595.
11. Himanen L, Geurts A, Foster AS, Rinke P. Data-Driven Materials Science: Status, Challenges, and Perspectives. Adv Sci. 2019;6(21): 1900808. doi: 10.1002/advs.201900808.
12. Kaisler S, Armour F, Espinosa JA, Money W, editors. Big data: Issues and challenges moving forward. 2013 46th Hawaii international conference on system sciences; 2013: IEEE. doi: 10.1109/HICSS.2013.645.
13. Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, et al. Autonomous experimentation systems for materials development: A community perspective. Matter. 2021;4(9): 2702-26. doi: 10.1016/j.matt.2021.06.036.
14. Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Sheberla D, Montoya JH, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. Nat Rev Mater. 2018;3(5): 5-20. doi: 10.1038/s41578-018-0005-z.
15. MacLeod BP, Parlane FGL, Morrissey TD, Hase F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20): eaaz8867. doi: 10.1126/sciadv.aaz8867.
16. Montoya JH, Aykol M, Anapolsky A, Gopal CB, Herring PK, Hummelshoj JS, et al. Toward autonomous materials research: Recent progress and future challenges. Appl Phys Rev. 2022;9(1): 011405. doi: 10.1063/5.0076324.
17. Coley CW, Eyke NS, Jensen KF. Autonomous Discovery in the Chemical Sciences Part I: Progress. Angew Chem Int Ed. 2020;59(51): 22858-93. doi: 10.1002/anie.201909987.
18. Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11(1): 5966. doi: 10.1038/s41467-020-19597-w.
19. Sun SJ, Tiihonen A, Oviedo F, Liu Z, Thapa J, Zhao YC, et al. A data fusion approach to optimize compositional stability of halide perovskites. Matter. 2021;4(4): 1305-22. doi: 10.1016/j.matt.2021.01.008.
20. Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, et al. Autonomy in materials research: a case study in carbon nanotube growth. npj Comput Mater. 2016;2(1): 1-6. doi: 10.1038/npjcompumats.2016.31.
21. Gongora AE, Xu B, Perry W, Okoye C, Riley P, Reyes KG, et al. A Bayesian experimental autonomous researcher for mechanical design. Sci Adv. 2020;6(15): eaaz1708. doi: 10.1126/sciadv.aaz1708.
22. Ren F, Ward L, Williams T, Laws KJ, Wolverton C, Hattrick-Simpers J, et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. Sci Adv. 2018;4(4): eaaq1566. doi: 10.1126/sciadv.aaq1566.
23. Pollice R, Dos Passos Gomes G, Aldeghi M, Hickman RJ, Krenn M, Lavigne C, et al. Data-Driven Strategies for Accelerated Materials Design. Acc Chem Res. 2021;54(4): 849-60. doi: 10.1021/acs.accounts.0c00785.

24. Ferguson AL, Brown KA. Data-Driven Design and Autonomous Experimentation in Soft and Biological Materials Engineering. Annu Rev Chem Biomol Eng. 2022;13: 25-44. doi: 10.1146/annurev-chembioeng-092120-020803.
25. Epstein SL. Wanted: Collaborative intelligence. Artif Intell. 2015;221: 36-45. doi: 10.1016/j.artint.2014.12.006.
26. Duros V, Grizou J, Sharma A, Mehr SHM, Bubliauskas A, Frei P, et al. Intuition-Enabled Machine Learning Beats the Competition When Joint Human-Robot Teams Perform Inorganic Chemical Experiments. J Chem Inf Model. 2019;59(6): 2664-71. doi: 10.1021/acs.jcim.9b00304.
27. Warren JA. The Materials Genome Initiative and artificial intelligence. MRS Bull. 2018;43(6): 452-7. doi: 10.1557/mrs.2018.122.
28. de Pablo JJ, Jackson NE, Webb MA, Chen L-Q, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. npj Comput Mater. 2019;5(1): 1-23. doi: 10.1038/s41524-019-0173-4.
29. Scheffler M, Aeschlimann M, Albrecht M, Bereau T, Bungartz HJ, Felser C, et al. FAIR data enabling new horizons for materials research. Nature. 2022;604(7907): 635-42. doi: 10.1038/s41586-022-04501-x.
30. MacLeod BP, Parlane FGL, Brown AK, Hein JE, Berlinguette CP. Flexible automation accelerates materials discovery. Nat Mater. 2022;21(7): 722-6. doi: 10.1038/s41563-021-01156-3.
31. Deagen ME, Brinson LC, Vaia RA, Schadler LS. The materials tetrahedron has a "digital twin". MRS Bull. 2022. doi: 10.1557/s43577-021-00214-0.
32. Davis R, Shrobe H, Szolovits P. What is a knowledge representation? AI Magazine. 1993;14(1): 17-33.
33. Stockwell F. A History of Information Storage and Retrieval. Jefferson, North Carolina: McFarland & Company, Inc.; 2000.
34. Bush V. As we may think. The Atlantic Monthly. 1945;176(1): 101-8.
35. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O. The high-throughput highway to computational materials design. Nat Mater. 2013;12(3): 191-201. doi: 10.1038/nmat3568.
36. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys. 2019;378: 686-707. doi: 10.1016/j.jcp.2018.10.045.
37. McDannald A, Frontzek M, Savici AT, Doucet M, Rodriguez EE, Meuse K, et al. On-the-fly autonomous control of neutron diffraction via physics-informed Bayesian active learning. Appl Phys Rev. 2022;9(2): 021408. doi: 10.1063/5.0082956.
38. Metcalfe B. Metcalfe's law after 40 years of ethernet. Computer. 2013;46(12): 26-31.
39. West G. Scale: The universal laws of life, growth, and death in organisms, cities, and companies: Penguin; 2018.
40. Bathelt H, Malmberg A, Maskell P. Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geog. 2004;28(1): 31-56. doi: 10.1191/0309132504ph469oa.
41. Farrell J, Klemperer P. Coordination and lock-in: Competition with switching costs and network effects. Handbook of Industrial Organization. 2007;3: 1967-2072. doi: 10.1016/S1573-448X(06)03031-7.
42. Shapiro C, Varian HR, Carl S. Information rules: A strategic guide to the network economy: Harvard Business Press; 1998.
43. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1): 160018. doi: 10.1038/sdata.2016.18.

44. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. ACM Comput Surv. 2021;54(6): 1-35. doi: 10.1145/3457607.
45. Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. Int J Hum-Comput Interact. 2020;36(6): 495-504. doi: 10.1080/10447318.2020.1741118.
46. Fielding RT, Taylor RN. Principled design of the modern web architecture. ACM Trans Internet Technol. 2002;2(2): 115-50.
47. Christensen M, Yunker LPE, Shiri P, Zepel T, Prieto PL, Grunert S, et al. Automation isn't automatic. Chem Sci. 2021;12(47): 15473-90. doi: 10.1039/d1sc04588a.
48. Horvitz E, editor Principles of mixed-initiative user interfaces. Proceedings of the SIGCHI conference on Human Factors in Computing Systems; 1999. doi.
49. Ward MO, Grinstein G, Keim D. Interactive data visualization: foundations, techniques, and applications: AK Peters/CRC Press; 2010.
50. Munzner T. Visualization analysis and design: CRC press; 2014.
51. Rule A, Tabard A, Hollan JD, editors. Exploration and explanation in computational notebooks. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018. doi: 10.1145/3173574.3173606.
52. Fombona-Pascual A, Fombona J, Vicente R. Augmented Reality, a Review of a Way to Represent and Manipulate 3D Chemical Structures. J Chem Inf Model. 2022;62(8): 1863-72. doi: 10.1021/acs.jcim.1c01255.
53. Heller MC, Mazor MH, Keoleian GA. Plastics in the US: toward a material flow characterization of production, markets and end of life. Environ Res Lett. 2020;15(9): 094034. doi: 10.1088/1748-9326/ab9e1e.
54. Lin TS, Coley CW, Mochigase H, Beech HK, Wang W, Wang Z, et al. BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules. ACS Cent Sci. 2019;5(9): 1523-31. doi: 10.1021/acscentsci.9b00476.
55. Lin TS, Rebello NJ, Beech HK, Wang Z, El-Zaatari B, Lundberg DJ, et al. PolyDAT: A Generic Data Schema for Polymer Characterization. J Chem Inf Model. 2021;61(3): 1150-63. doi: 10.1021/acs.jcim.1c00028.
56. Lin T-S, Rebello NJ, Lee G-H, Morris MA, Olsen BD. Canonicalizing BigSMILES for Polymers with Defined Backbones. Manuscript in review. 2022.
57. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. Co-Design. 2008;4(1): 5-18. doi: 10.1080/15710880701875068.
58. Haak LL, Fenner M, Paglione L, Pentz E, Ratner H. ORCID: a system to uniquely identify researchers. Learned Publ. 2012;25(4): 259-64. doi: 10.1087/20120404.
59. Chard K, Tuecke S, Foster I. Efficient and Secure Transfer, Synchronization, and Sharing of Big Data. IEEE Cloud Comput. 2014;1(3): 46-55. doi: 10.1109/mcc.2014.52.
60. Tuecke S, Ananthakrishnan R, Chard K, Lidman M, McCollam B, Rosen S, et al., editors. Globus Auth: A research identity and access management platform. 2016 IEEE 12th International Conference on e-Science (e-Science); 2016: IEEE. doi: 10.1109/eScience.2016.7870901.

---

[i] Certain commercial products are identified in this paper in order to specify the procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the product identified is necessarily the best available for the purpose.