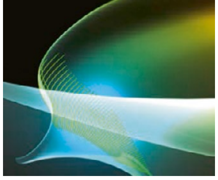
MRS Energy & Sustainability

doi:10.1557/s43581-024-00122-1



PERSPECTIVE

Building a roadmap for safer and sustainable material chemistries: Addressing the PFAS problem through informatics and data-driven chemistry

Arthur Fong, Environmental Technologies, Apple Inc., Cupertino, CA 95014, USA

Alexandra McPherson and **Mark Rossi**, Clean Production Action, Somerville, MA 02144, USA

Krishna Rajan, Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY 14260, USA

Address all correspondence to Mark Rossi at mark@cleanproduction.org

(Received: 27 February 2024; accepted: 29 October 2024)

ABSTRACT

The development of sustainable materials needs to take into account the chemicals that appear at different stages of material synthesis, processing, and manufacturing, including molecular precursors, solvents, and PFAS containing compounds. We describe an accelerated data-driven framework for designing safer material chemistries that also accounts for the impact of chemicals to ensure technical functionality and provide a holistic approach to sustainability.

Scientific research increasingly demonstrates that chemicals and materials essential for everyday products threaten natural systems and human health. Transitioning to sustainable, circular, and low-carbon economies depends critically on having safer chemicals available. We propose that materials scientists should also account the impact of the health hazards of chemicals associated with the synthesis, processing, and manufacturing of materials. These include molecular precursors for synthesis of new materials chemistries and chemicals used in various stages of materials processing and manufacturing such as solvents and persistent, accumulative, and highly hazardous. Per- and polyfluoroalkyl substances are examples of harmful chemicals that pose health and environmental risks. A major challenge is finding safer yet functional alternatives that also the meet necessary performance requirements in sustainable materials design and development. The exploration space to discover these is prohibitively large to explore. Hence, we are at a critical inflection point and a paradigm shift is needed to include the development of safer chemicals as part of the equation to accelerate the adoption of safer and more sustainable chemical materials. Using such chemicals as an example, we describe an accelerated data-driven framework for designing safer material chemistries that ensures technical functionality and provide a holistic approach to sustainability

Keywords artificial intelligence · chemical substitution · environmentally benign · lifecycle · sustainability · toxicity

Discussion

To create a safer and sustainable materials ecosystem rapidly, one must address many competing and conflicting environmental, economic, and social consequences, requiring new paradigm for materials research. We argue that materials informatics provides a framework to meet an expanded definition of materials performance that includes multiple metrics of functionality and the safety of chemicals used in materials synthesis, processing, and manufacturing. The power of this new research paradigm for materials innovation lies through discoveries that make it more feasible to address environmental and social impacts at the front end of material discovery, design, and deployment.

Chemical pollution reaches inflection point

Over the last five years, a series of groundbreaking reports have documented growing evidence that we have reached a tipping point with chemical pollution's impact on people and the environment. These reports reflect major trends that have emerged over the last few decades, including a 2022 study by Persson et al.[1] that chemical pollution has crossed a "planetary boundary." Chemical exposure and presence in the environment have significantly upset the balance required for the planet to maintain human life. This is the fifth of nine planetary boundaries that are in jeopardy of being crossed along with global heating, loss of biodiversity, the degradation of natural habitats, and the overabundance of nitrogen and phosphorus pollution. The study advocated for urgent measures to decrease chemical pollution by promoting circular economies. The authors recommend taking *"urgent action to reduce the harm associated with exceeding the boundary by reducing the production and releases of novel entities,"* noting that even so, the persistence of many novel entities and/or their associated effects will continue to pose a threat.[1]

At the root of the problem is the growing reliance on over 350,000 synthetic chemicals derived primarily from fossil fuels that are used ubiquitously in products, infrastructure, and industrial systems for modern-day society. Persson et.al[1] noted that since 1950, there has been a 50-fold increase in chemical production, and it is projected to triple again by 2050. Fossil fuels are the fundamental building blocks of over 96% of manufactured goods for our economy.[2] For example, 90% of downstream organic chemical production is made from seven essential petrochemicals including methanol; olefins like ethylene, propylene, and butadiene; and aromatic compounds such as benzene, toluene, and xylene–which entered the market in the 1940s and 1950s.[3]

A 2017 study estimated that the global economic costs of environmental chemical exposures could surpass 10% of the world's GDP, totaling approximately 11 trillion dollars.[4] Dr. Shanna Swan, a prominent expert in environmental and reproductive epidemiology, has researched and established a connection between falling sperm counts, exposure to endocrine disrupting chemicals (EDCs), and the fertility crisis."[5] Public awareness of the adverse impacts caused by materials and hazardous chemicals is rising rapidly with the increasing visibility of plastic pollution in oceans and PFAS in drinking water[6] As documented by Lane et.al,[7] hundreds of studies have been identified 'overburdened communities' that suffer from environmental injustice and serve as the frontline facing the burden of unsafe chemicals and industries. They argue "Green chemistry must aim to "simultaneously designs for functional performance and sustainability, including multifaceted environmental, economic and social considerations." In this article, we advocate that the design of new materials must take the same approach and include the hazard impact in all aspects of materials development in parallel to achieving its functionality.

Solutions to these problems will require a significant transformation in how we design and develop materials for manufacturing and products. Embedding the development of inherently safer and greener chemistries and materials as a critical component of our transition to sustainable, circular, and low-carbon economies will help scale solutions needed to address the problems of chemical and plastic pollution.

A paradigm shift in material innovation is needed

Materials and chemicals are central to every technology (e.g., photovoltaics, batteries), and demands for sustainable materials that are based on safer and benign-by-design chemistries are rising in response to the widespread impacts of chemical pollution. Companies, investors, governments, and environmental leaders have demonstrated that finding safer substitutes for chemicals of concern used in high volumes, such as problematic flame retardants, plasticizers, and solvents, is critical to economic and technological viability.[8] But the availability of safer materials for many applications is not at scale to meet market demand. Green chemistry holds great promise, but from an environmental lens, is often not prioritized at the materials design level unless there are regulatory and/or industry restrictions. On average, it takes years to transition to new high-performance materials, and in many cases, those new materials might be '*regrettable substitutes'* in that they pose other human health and environmental risks. New science-based solutions are needed to analyze society's complex reliance on hazardous chemicals and *rapidly* create innovative solutions.[9,10]

In the movement to create new sustainable materials for a low-carbon economy, often overlooked in material design are the hazards of the chemical building blocks of these materials. Addressing the inherent hazards of chemicals is core to the 12 Principles of Green Chemistry, leading figures, Paul Anastas and John Warner, emphasize in their foundational book, that Green chemistry is fundamentally about designing chemicals and processes that reduce or eliminate hazards at every stage of their development.[11] And addressing the inherent hazards of materials is core to the 12 Principles of Green Engineering: "Designers need to strive to ensure that all materials and energy inputs and outputs are as inherently nonhazardous as possible."[12] As material engineers design new materials for a low-carbon economy, it is also essential that these materials be safer and healthier for people and the planet.

Companies and environmental leaders have made progress in differentiating materials and chemicals in the marketplace by setting priorities around health and environmental impact data. For example, Apple partnered with Clean Production Action to implement a hazard assessment tool, the GreenScreen® for Safer Chemicals, which rates chemicals on a scale of 1 to 4, from most hazardous to least, according to 18 distinct human health and ecotoxicity criteria.[13] Clean Production Action developed the tool in recognition that guidance was needed to define the criteria for inherently safer chemistries. Apple collaborated with Clean Production Action to establish the GreenScreen Certified® standard, a publicly accessible tool designed to evaluate and encourage the use of safer chemicals in cleaners and degreasers. By applying both GreenScreen® and the U.S. Environmental Protection Agency's Safer Choice criteria for thorough assessment, Apple successfully transitioned all its final assembly sites to safer cleaning and degreasing alternatives.[14] In some cases, companies reformulated their products to remove chemicals of concern to meet GreenScreen Certified® requirements, including the removal of PFAS from product formulations. For the most part, however, this process identified better materials available today but did not redefine innovation and discovery for new materials to enter the marketplace.

Differentiating chemicals and materials in the marketplace based on defined environmental and human health metrics is a critical step forward in addressing the chemical pollution problem and creating sustainable materials. But this approach alone will not create the change needed to scale solutions that reduce chemical pollution and restore planetary boundaries. Bringing tools like the GreenScreen® that organize and aggregate complex environmental and human data into the front end of material design and discovery will help material scientists make better and more sustainable design choices. But this cannot be done in isolation; the environmental data need to be relational to and integrated with the technical and economic data that drive material design and discovery.

A new approach is essential–one that leverages artificial intelligence (AI) and big data to expedite the design and creation of materials and chemicals that not only fulfill engineering and technical requirements but are also cleaner, safer, and more sustainable. This can be achieved by integrating both technical and environmental considerations at every stage of designing, developing, and utilizing materials and chemicals. In essence, we need to innovate with more intelligent chemical solutions. This means materials and chemicals that are *benign-by-design*, which entails designing materials and chemicals with inherent safety, tailored for specific functionalities, by proactively considering hazard impacts across all stages from, synthesis to processing and ultimately recycling. It also involves understanding how material and chemical choices at various production stages interact. To foster safer chemistries, we need a comprehensive understanding of the molecular mechanisms driving environmental and human toxicity, aiming to redefine performance standards. This broader approach integrates material functionality with the potential hazards embedded in synthesis and fabrication processes Fig. 1).

One of the key challenge in embracing this holistic method is the diverse and disconnected nature of data-intense research in chemistry, environmental science, toxicology, and materials science that is siloed across scientific communities, industry, and other stakeholders. We need new and innovative approaches that take advantage of advances in AI and data sciences to *coalesce* information from decades of research on chemistries, materials, and their environmental and human health impact.[14–17] Aside from organized databases such as PubMed, the harvesting of information has been revolutionized in the last few years with the advent of techniques such as Large Language Models that can help to gather large volumes of information from both text and graphics.[18] Scientists will use these aggregated data sets and tools to guide and inform the development of smarter chemistries for new materials and chemicals. This approach, grounded in rational design and data-driven tools, holds promise for expediting the discovery of materials that achieve both technical functionality and minimal hazard impact. We define 'rational design' as a method that avoids trial-and-error, instead developing materials through predictive insights into the fundamental science that dictates material performance.

To establish a rational design paradigm, we need to bring information on why a chemistry is hazardous together with a robust information infrastructure to look for other potentially safer chemistries. We need to deploy a new set of multifaceted data analytical tools that can unravel the multiscale relationships across the entire system (from molecular structure to engineering behavior), coupled with an understanding of the mechanisms that control behavior, and achieve this within a reasonable timeframe. This requires embedding computational and experimental tools that are capable of building and interrogating a robust information infrastructure to provide guidelines for materials/chemistry selection, design, and discovery. Hence, we need to integrate a benign-by-design search process for developing new chemicals and materials chemistries that also avoid 'regrettable substitutes.'[19–23]

This approach represents a paradigm shift in material design and discovery that can accelerate the development of inherently safer and better-performance materials. It is based on new multifaceted, data analytical tools that fuse knowledge and information that is currently disaggregated and not readily accessible. It utilizes data-driven screening methods capable of estimating potential chemical toxicity and guiding the design of safer alternatives, all while preserving the material's chemistry and functionality for its intended engineering application..

Building the toolkit for a rational design

To understand the need for a toolkit, consider for instance the problem that we face with PFAS. Polyfluoroalkyl substances (PFAS) are chemicals characterized by having at least one carbon atom fully fluorinated. They possess important properties such as chemical and thermal stability, and the ability to repel water and oil, PFAS compounds are used in many industrial and commercial products. However, due to their high chemical resiliency, they do not break down in the environment; thus, the phrase 'forever chemical' is attributed to PFAS chemistry. PFAS also bioaccumulate in the environment and are chemicals that raise significant concerns because of their environmental and human health hazards.

PFAS take innumerable forms and each compound has a complex network of relationships within the PFAS family. Hence, as the list of new substances in the PFAS family grows rapidly, the features that distinguish the structure-functional relationships of different chemistries are not easy to identify. Establishing clear criteria for classifying PFAS compounds is

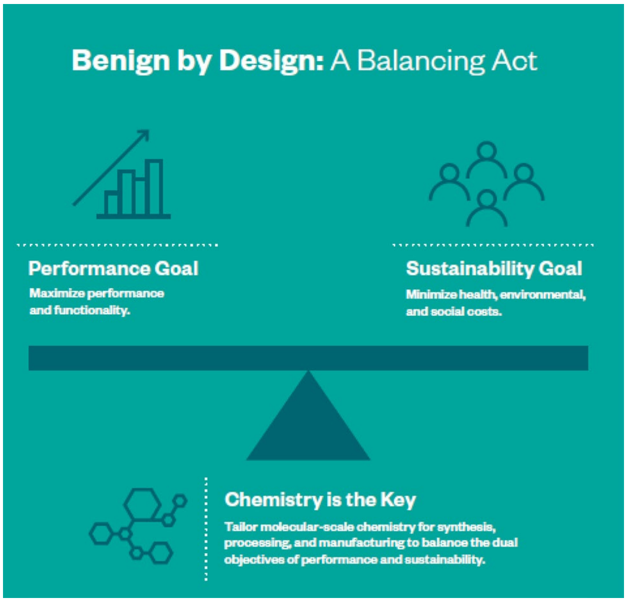

Benign by Design: A Balancing Act
Performance Goal
Maximize performance and functionality.
Sustainability Goal
Minimize health, environmental, and social costs.
Chemistry is the Key
Tailor molecular-scale chemistry for synthesis, processing, and manufacturing to balance the dual objectives of performance and sustainability.

Figure 1. Meeting both sustainability and performance goals at the outset of chemical and materials discovery requires a comprehensive understanding of the multiscale impact of molecular scale chemistry on both targeted functionality as well as its health, environmental, and social footprint.

critical for guiding the search for new and safer chemistries or the modifications of existing chemistries. The work in applying machine learning involves results with respect to available data.[24,25] The interpretation of such models are in fact based on understanding the chemistry and that helps to understand the level of confidence of a machine learning base result. This provides a foundation for the next generation of models, and identifies where data needs to be collected.

In the search for safer substitutes, we need to address the following questions:

1. Can we discover fundamental chemical/structural criteria at the molecular scale that we could use to find alternatives?
2. From the information in #1 above, can we identify hidden or difficult to identify chemical characteristics at the molecular level that would suggest what features we should avoid, or which fundamental chemistry should be changed?
3. Are there completely different classes of chemicals that can serve as substitutes for existing PFAS applications? If there are, then we need to expand an already massive chemical search space to an almost infinite one.

Although there is extensive work on the toxicological assessment of chemicals,[26] the question that we seek to address is how do we use that information to find safer alternatives in a timely manner without compromising technical functionality, cost, availability, and other key factors? Every decision that we make may have consequences. For example, a less hazardous material might be extremely rare. A chemical that degrades more slowly may not be as efficient. A less hazardous solvent used in the manufacturing process could affect the final product's longevity.

Informatics-based materials discovery paradigm has the potential to significantly advance knowledge of materials and chemical safety. For example, by analyzing key molecular features, we can

- Select safer alternatives to hazardous chemicals currently in use,
- Achieve highly precise knowledge of the combinations of chemistries, molecular design characteristics, processing strategies, and
- Control variables that will most efficiently and rapidly provide the desired results.

In addition to transforming hazard assessment approaches, the new paradigm will foster stronger convergence between data science and environmental science, toxicology, chemistry, and materials science, leveraging this knowledge to speed up the shift toward safer materials and clean production processes.

It is important to distinguish between materials chemistry and the chemicals that are used to make materials (Fig. 2). *Materials chemistry* addresses the final product, such as a battery, including how the product is used, how it degrades over time, and ultimately how it is decommissioned. For example, if hazardous materials are in the product, are they encapsulated? How much hazardous material persists after the product reaches the end of its expected lifespan? We also need to be concerned about the *chemicals used to make materials* and address a priori the hazards of the precursor chemicals involved in materials synthesis.[27] For example, are there hazardous solvents required for manufacturing, even if these solvents never end up in the final product? Is there a hazardous process involved in synthesizing any of the needed materials?

Our goal: find better starting points

In the following section, we discuss an approach for information/data gathering, harvesting, and sharing that will provide the foundation for identifying pathways for choosing options that fulfill both performance needs and safety criteria. This approach will also pinpoint key gaps in models and experiments that must be addressed to move beyond trial-and-error methods, paving the way for designing safer and more sustainable materials and chemicals. Our goal is to find better, 'smarter' starting points that can most quickly identify the most promising pathways for improvement. By improving our ability to predict chemical performance, we will quickly discover *which* pathways we should pursue first in order to have the greatest chance at success.

What we need to do are the following:

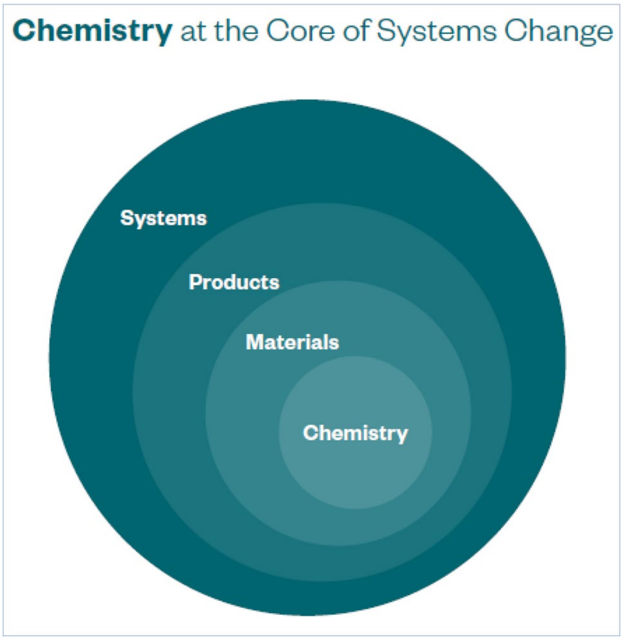


Figure 2. It is crucial to comprehend the effects of chemistry at the most fundamental level and trace its impact through all aspects of the materials design from synthesis to manufacturing to performance and finally, its impact on the environmental and health (adapted from M. Rossi, A. Blake, The Plastics Scorecard, 2014, p. 10[18]).

- ***Identify*** foundational 'chemical design rules' that can guide the selection, search, and screening of new chemistries that meet the multiscale benign-by-design objectives.[28]
- ***Aggregate*** information from varied sources, including quantitative physico-chemical models, environmental, toxicology, and clinical data, as along with descriptive and heuristic insights to build a broad, multidimensional evidentiary portfolio on the chemicals of interest, and simultaneously identify gaps in knowledge.[29]
- ***Extract*** features or 'signatures' that capture the complexity of data correlations gathered from the above-mentioned steps to serve as guideposts for navigating the information landscape to search for alternatives and/or discover pathways for new chemistries.[30]
- ***Embed*** this entire knowledge base into a shared information system that can serve both as a repository and as an 'atlas' for identifying new chemistries and pathways for discovering new and 'smarter' chemistries.[31,32]

Identify: discovering multiscale structure-function relationships

The high degree of complexity in structure-function relations makes it difficult to extract 'design rules' for selection and substitution. Hence, we need a strong foundational understanding of multiscale structure-function relationships on toxicity, and we need to identify those relationships that track hazard impacts concurrently to chemistry and their engineering functionality.

Establishing criteria for selecting alternatives is difficult because the underlying chemistry (at the molecular level) that explains a chemical's behavior (functionality, toxicity, etc.) is very complex and involves many factors. There are multiple factors at the molecular level that define the 'signature' of a chemical compound. Some of these factors contribute to and determine the level of the chemical's toxicity. Hence, we need this molecular-level information to understand why a chemical is hazardous so that we can find alternatives. Hence, the accelerated design of materials is driven by the rational selection of molecules based on an understanding of structure and properties and the mechanisms that control behavior (properties).

An important step in capturing this complexity is to take advantage of machine-readable representations of molecular structure that can encapsulate the numerous relationships that link molecular chemistry to the multitude of measures representing its functionality and toxicity. This enables us to develop an AI approach to detect complex patterns in the multidimensional data space that links structure to function. We can detect patterns that provide critical clues on what aspects of structure govern properties; information that otherwise would be missed from empirical observations.

Machine learning has been highly effective in analyzing existing data to uncover significant patterns in existing data on the properties of the family of PFAS molecules, (including bioactivity, bond strength, and sources) and used to make predictions.[33]

Many studies use supervised machine learning methods, where molecular structural information serves as input features and known properties act as labels. However, a major challenge is that there are far more PFAS compounds with unknown chemical structures than those with known properties. Additionally, the number of PFAS with known properties is significantly less than those with identified structures. Conversely, unsupervised learning–an exploratory machine learning approach capable of uncovering hidden patterns or groupings in data without needing labels–has not been fully utilized in PFAS research.

We have used visualization methods to overcome this problem by presenting complex information in a 3D format to uncover underlying patterns or groupings in data without the use of any labels.[34] For example, we have compiled extensive information on chemical compounds, including PFAS, by leveraging AI, chemical modeling, and data analysis to develop a 'data atlas' that maps the intricate chemical landscape, enabling exploration and insights into property relationships. This PFAS-Atlas can be used for predicting and estimating fundamental physical properties of PFAS chemicals that have not yet been measured; uncovering hierarchical patterns in existing classification systems; and integrating data from diverse sources. The Atlas also provides a common platform to compare different chemicals with respect to their toxicity.

Developed using open-source data, the PFAS-Atlas is adaptable to updates as PFAS classification methods advance and evolve. Changes can be seamlessly integrated into the classification program's source code, ensuring that new classification patterns from scientific research are quickly incorporated.

Additionally, it functions as an inference tool, allowing for rapid assessment of the potential functionalities of new PFAS molecules by comparing them with existing property data.[35] Most recently, for example, Patlewicz et.al[36] shown how the classification models can serve as a practical means of guiding toxicological testing.

Aggregate: harvesting and gathering diffuse information

Information on toxicity and hazard impact is diverse, sparse, and unstructured, and covers all types of data sources; for example, toxicology models, clinical data, archived laboratory and field data in databases, interpretation of data embedded in text and diagrams, to mention a few. The first challenge that we face is gathering all this information. The second problem is that there are far more chemicals than the known information on the toxicity of chemicals. Hence, it is a daunting task to build empirical, statistical, and/or physico-chemical models to establish a predictive framework to rapidly find safer alternatives.

While there have been, and continue to be, significant efforts to build repositories and databases on chemical safety, the enormous volume of chemicals in use far exceeds the knowledge and data that we have on the complex environmental and health hazards associated with these chemicals. Further, a large amount of information exists in the form of heuristic information, embedded in texts, diagrams, scientific articles, and reports. It is critical that we harvest this knowledge, format it as structured information, and add it to databases on materials and chemical safety. This combined information, together with the tools and techniques of statistics and machine learning, then becomes a rich resource for physico-chemical modeling and offers a powerful framework for the accelerated discovery and property predictions of new chemistries.

Advancements in Natural Language Processing, particularly with the advent of Large Language Models, have and Graphics Recognition techniques can aid in converting unstructured textual data to encoded variables. Transforming unstructured text data into encoded variables enables its use in machine learning applications. Natural Language Processing (NLP) and Large Language Models (LLMs) offer advanced tools to extract targeted information from vast textual corpora, making it suitable for input into machine learning paradigms like classification or regression. These models can function in both context-independent and context-sensitive modes. When context is considered, NLP and LLMs are useful for tasks such as sequence classification, question answering, language modeling, and translation. LLM-based NLP methods make it possible to annotate entire articles and develop classification tools to extract relevant keywords. By combining text mining of published literature–especially for information not found in curated databases–with chemistry-guided machine learning, we can expedite the development of safer and cleaner technologies by predicting the chemical impact of material synthesis and manufacturing processes.[37–39]

Extract: accelerating the screening and finding of alternative chemistries

The major challenge in screening known chemistries with unknown properties is that we are searching for large amounts of chemical space with limited experimental data. The question is whether we can predict with reasonable confidence the likelihood that a molecule meets multiple criteria that are critical for identifying hazard impacts. By combining molecular chemistry fundamentals with AI methods, it is possible to create a *digital signature* to identify and search, in an automated fashion, the essential molecular structures within a molecule that greatly contribute to a chemical's properties.

This significantly accelerates a rational design approach to screening large amounts of molecules where the property information space is sparsely populated. As an example, we have demonstrated methods for identifying critical molecular structures that significantly influence a chemical's ability to function as an endocrine disruptor.[40] This now allows us to accelerate the identification of physically meaningful 'digital' signatures to screen and find alternative chemistries.

Embed: enabling an information network for all stakeholders

The vision for the next generation of information infrastructure must support the FAIR principles for building and establishing databases–Findability, Accessibility, Interoperability, and Reusability.[41] Our vision for using informatics to design smart chemistries is to establish a trustworthy machine-enabled system that allows easy access, provides search capabilities, and operates across communities of stakeholders. To achieve this goal, our information infrastructure will have the following capabilities:

- Create a data-driven approach to gather and harvest complex information from diverse fields (environmental, health, social, economic, chemical, and materials sciences) to guide the design of safer materials and chemicals that meet targeted engineering functionality.
- Unify multiscale approaches, experimental methods, and computational techniques to engineer innovative molecular chemistries for smarter chemistries.
- Identify approaches for advancing our chemical knowledge to achieve high chemical precision derived from the combination of chemistries; molecular design characteristics; materials processing strategies; and control of those experimental variables that will most efficiently and rapidly provide the desired result.
- Create accessible information for diverse stakeholders, including academia, industry, policymakers, and advocacy groups.

The FAIR principles are even more relevant and challenging for our benign-by-design paradigm as it is imperative that new discoveries, methods, and processes are accessible to diverse stakeholders and benefit all communities.

Alongside the FAIR principles, it is equally essential to incorporate considerations of data justice. Not all stakeholder communities have access to the information or are part of the innovation process. Hence, it is important that our information infrastructure can break down current barriers to access so that all stakeholders can participate in and contribute to discussions on the benign-by-design paradigm for smarter chemistries. We also need a methodology that is capable of including economic and social factors in the use of materials and chemicals so that smarter chemistries can transform the unequal impact that hazardous materials and chemicals currently have on different communities. We require a broader definition of performance that encompasses engineering functionality, hazard impacts, and social and economic considerations. Any effort to develop smarter chemistries must incorporate all these factors at the outset, so that we embed social justice metrics into our informatics-driven framework for smarter chemistry.[42–44] By embedding the principles of "FAIR *and* just" at the outset, we would be able to achieve this goal of accelerated discovery and design of new materials that meet targeted functionalities and minimize or eliminate hazard impacts, for the betterment of all communities. We also recognize that holistic solutions must be underpinned by end-use-focused material science and chemistry, all framed by considerations of environmental impact, and social and economic justice.

Conclusion

Given that chemical pollution has surpassed safe levels for both the planet and humanity, this innovative approach aims to accelerate our search for safer and smarter chemistry solutions for materials essential to society. Time is of the essence. Corporations, investors, nonprofits, and governments worldwide are advancing programs to drive safer chemicals and more sustainable materials in their supply chains, production, and products, but the quest for these innovations has exposed the gaps in our understanding of what makes a material perform or function the way it does. Informatics-based material discovery approaches will advance knowledge of materials and chemistry that was previously unknown or uncertain. Expanding this toolbox will facilitate solutions to complex issues like PFAS by offering deeper insights into the molecular properties and mechanisms responsible for environmental and human toxicity, as well as the performance characteristics essential to specific chemistries. The resulting aggregated and organized data will guide decision-making that leads to the development at the scale of sustainable materials that are benign-by-design for targeted functionalities. An expanded definition of materials performance and sustainability that includes functionality and hazard impacts inherent in the synthesis and fabrication of materials will be embedded into the front end of research and development for material innovation programs. This is a grand challenge, one that will only be met by a broad set of stakeholders collaborating to bring disparate pieces of information together to fill data gaps, avoid the development of regrettable substitutions for materials of concern, and accelerate the pace of change toward solutions that contribute to net zero circular economies.

Author contributions

Not applicable.

Funding

This research was supported by funding from the JPB Foundation Grant #2020-1949 and the NSF Engines Development Award grant #2315307. It has also been supported by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program. This publication, Building a Roadmap for Safer Chemicals for Sustainable Material Chemistries through Informatics, was developed under Assistance Agreement No. 84045101 awarded by the U.S. Environmental Protection Agency to K. Rajan. It has not been formally reviewed by EPA. The views expressed in this document are solely those of A. Fong, A. McPherson, M. Rossi, and K. Rajan, and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. None of the funding sources played a role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparations, review or approval of the manuscript; or decision to submit the manuscript for publication.

Data availability

Not applicable.

Code availability

Not applicable.

Declarations

Conflict of interest

No conflicts of interest.

Open Access

REFERENCES

1. L. Persson, B.M.C. Almroth, C.D. Collins, S. Cornell, C. de Wit et al. Environ. Sci. Technol. (2022). https://doi.org/10.1021/acs.est.1c04158
2. American Chemistry Council, 2020 Guide to the Business of Chemistry (2020) https://www.americanchemistry.com/chemistry-in-america/data-industry-statistics/resources/2020-guide-to-the-business-of-chemistry, Accessed 9 Oct 2023
3. J. Tickner, K. Geiser, S. Baima, Environ. Sci. Policy Sustain. Dev. (2021). https://doi.org/10.1080/00139157.2021.1979857
4. P. Grandjean, M. Bellanger, Environ. Health (2017). https://doi.org/10.1186/s12940-017-0340-3
5. S. Swan, S. Colino, ***Count Down*** (Scribner, New York, 2020), p.201
6. Economist Impact, New surveys reveal heightened concern about ocean pollution (Economist Impact) (2021), https://impact.economist.com/ocean/sustainable-ocean-economy/the-economist-intelligence-unit-surveyed-more-than-1-000-global-executives. Accessed 23 Oct 2023
7. M.K.M. Lane, H.E. Rudel, J.A. Wilson et al., Green Chem. Just Chem. Nat. Sustain. **6**, 502–512 (2023)
8. B.I. Escher, H.M. Stapleton, E.L. Schymanski, Tracking complex mixtures of chemicals in our changing environment. Science **367**, 388–392 (2020)
9. J.B. Zimmerman, P.T. Anastas, H.C. Erythropel, W. Leitner, Science **367**, 397–400 (2020)
10. P.T. Anastas, J.C. Warner, ***Green Chemistry: Theory and Practice*** (Oxford, Oxford, 1998), p.34
11. P.T. Anastas, J.B. Zimmerman, Peer reviewed: design through the 12 principles of green engineering. Environ. Sci. Technol. **37**(5), 94–101 (2003)
12. S. Franjevic, M. Rossi, A. Hunsicker, M.W. Turner, GreenScreen® for Safer Chemicals: Hazard Assessment Guidance, Version 1.4 (Clean Production Action) (2018), https://www.greenscreenchemicals.org/images/ee_images/uploads/resources/GreenScreen_Guidance_v1_4_2018_01_Final.pdf. Accessed 22 Oct 2023
13. People and Environment in Our Supply Chain: 2021 Annual Progress Report (Apple) (2021), https://www.apple.com/supplier-responsibility/pdf/Apple_SR_2021_Progress_Report.pdf. Accessed 22 October 2023
14. K. von Borries, H. Holmquist, M. Kosnik, K.C. Beckwith, O. Jolliet, J.M. Goodman, P. Fantke, Potential for machine learning to address data gaps in human toxicity and ecotoxicity characterization. Environ. Sci. Technol. **57**(46), 18259–18270 (2023)
15. K. Rajan, A. McPherson, M. Rossi, Elements of change: moving forward together toward a cleaner, safer future (University at Buffalo, Niagara Share, and Clean Production Action) (2020) https://www.cleanproduction.org/images/ee_images/uploads/resources/HR_CORE20001_Report_M2.pdf. Accessed 22 Oct 2023
16. K. Rajan, Materials informatics: big data and the materials gene. Ann. Rev. Mater. Res. **45**, 153–169 (2015)
17. K. Rajan, Nanoinformatics: Materials Design for Health and Environmental Needs, in ***Nanotechnology Environmental Health and Safety: Risks, Regulation and Management***, 3rd edn., ed. by M. Hull, D. Bowman (Elsevier, Oxford, 2019), pp.119–150
18. S. Hodl, W. Robinson, Y. Bachrach et al. Explainability techniques for chemical language models (2023), arXiv:2305.16192v1 [cs.LG]
19. R. Song, A.A. Keller, S. Suh, Rapid life-cycle impact screening using artificial neural networks. Environ. Sci. Technol. **51**(18), 10777–10785 (2017)
20. X. Zhu, C.-H. Ho, X. Wang, Application of life cycle assessment and machine learning for high-throughput screening of green chemical substitutes. ACS Sustain. Chem. Eng. **8**(30), 11141–11151 (2020)
21. W. Cheng, C.A. Ng, Using machine learning to classify bioactivity for 3486 per-and polyfluoroalkyl substances (PFASs) from the OECD list. Environ. Sci. Technol. **53**(23), 13970–13980 (2019)
22. C. Cavasotto, V. Scardin, Machine learning toxicity prediction: latest advances by toxicity end point. ACS Omega **7**(51), 47536–47546 (2022)
23. T.V. Tran, A.S. Wibowo, H T, Kil and T. Chong, Artificial intelligence in drug toxicity prediction: recent advances, challenges, and future perspectives. J. Chem. Inf. Model. **63**(9), 2628–2643 (2023)
24. Raza et al., A machine learning approach for predicting defluorination of per- and polyfluoroalkyl substances (PFAS) for their efficient treatment and removal. Environ. Sci. Technol. Lett. **6**, 624–629 (2019)
25. X. Jia, X. et al., Advancing computational toxicology by interpretable machine learning. Environ. Sci. Tech. **57**(46), 17690–17706 (2023)
26. N.C. Kleinstreuer, I.V. Tetko, W. Tong, Introduction to special issue computational toxicology. Chem. Res. Toxicol. **34**, 171–175 (2021)
27. A. Babayigit, A. Ethirajan, M. Muller, B. Conings, Toxicity of organometal halide perovskite solar cells. Nat. Mater. **15**(3), 247–251 (2016)
28. E.N. Muratov, J. Bajorath, R.P. Sheridan et al., QSAR without borders. Chem. Soc. Rev. **49**, 3525 (2020)
29. K.M. Jablonka, L. Patiny, B. Smit, Making the collective knowledge of chemistry open and machine actionable. Nat. Chem. **14**, 365–376 (2022)
30. A.P. Bartók, S. De, C. Poelking, N. Bernstein, J.R. Kermode, G. Csányi, M. Ceriotti, Machine learning unifies the modeling of materials and molecules. Sci. Adv. **3**, e1701816 (2017)
31. A. Su, K. Rajan, A database framework for rapid screening of structure-function relationships in PFAS chemistry. Sci. Data **8**(1), 14 (2021)
32. A. Su, Y. Cheng, C. Zhang, Y.-F. Yang, Y.B. She, K. Rajan, An artificial intelligence platform for automated PFAS subgroup classification: a discovery tool for PFAS screening. Sci. Total. Environ. **921**, 171229 (2024)
33. M. Rossi, A. Blake, The Plastics Scorecard (Clean Production Action) (2014), https://www.bizngo.org/images/ee_images/uploads/resources/plastics_scorecard_2015_2_25e.pdf, p. 10. Accessed 6 Sept 2024
34. W. Cheng, C.A. Ng, Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list. Environ. Sci. Technol. **53**, 13970–13980 (2019)
35. National PFAS testing strategy: identification of candidate per- and poly-fluoroalkyl substances (PFAS) for testing. US Environmental Protection Agency (2021), p. 6
36. G. Patlewicz, R.S. Judson, A.J. Williams et al., Development of chemical categories for per- and polyfluoroalkyl substances (PFAS) and the proof-of-concept approach to the identification of potential candidates for tiered toxicological testing and human health assessment. Comput. Toxicol. **31**, 100327 (2024)
37. D. Giri, A. Mukherjee, K. Rajan, Informatics Driven Materials Innovation for a Regenerative Economy: Harnessing NLP for Safer Chemistry in Manufacturing of Solar Cells. In: A. Lazou et al. (eds), REWAS 2022: Developing Tomorrow's Technical Cycles (Volume I), The Minerals, Metals & Materials Series (2022). pp. 11–19
38. D. Giri, A. Mukherjee, K. Rajan, Uncertainty informed screening for safer solvents used in the synthesis of perovskite based solar cells via machine learning. https://doi.org/10.26434/chemrxiv-2022-zjzb2

39. A. Su, H. Xue, Y. She, K. Rajan, AI informed toxicity screening of amine chemistries used in the synthesis of hybrid organic-inorganic perovskites. AIChE J. (2022). https://doi.org/10.1002/aic.17699
40. A. Mukherjee, A. Su, K. Rajan, Deep learning model for identifying critical structural motifs in potential endocrine disruptors. J. Chem. Inf. Model. **61**(5), 2187 (2021)
41. W.P. Dempsey, I. Foster, S. Fraser, C. Kesselman, Sharing begins at home: how continuous and ubiquitous FAIRness can enhance research productivity and data reuse. Harvard Data Sci. Rev. (2022). https://doi.org/10.1162/99608f92.44d21b86
42. L. Dencik, A. Hintz, J. Redden, E. Treré, Exploring data justice: conceptions, applications and directions. Inf. Commun. Soc. **22**(7), 873–881 (2019)
43. S.R. Carroll, E. Herczog, M. Hudson, K. Russell, S. Stall, Operationalizing the CARE and FAIR principles for indigenous data futures. Sci. Data **8**, 108 (2021)
44. S.R. Carroll, I. Garba, O.L. Figueroa-Rodríguez et al., The CARE principles for indigenous data governance. Data Sci. J. **19**(43), 1–12 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.