ACS Macro Letters

pubs.acs.org/macroletters Viewpoint

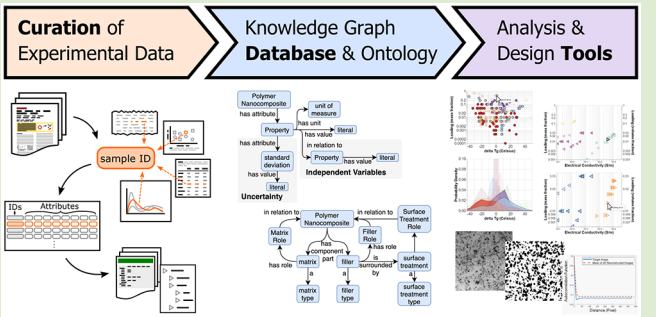
Polymer Nanocomposite Data: Curation, Frameworks, Access, and Potential for Discovery and Design

L. Catherine Brinson,* Michael Deagen, Wei Chen, James McCusker, Deborah L. McGuinness, Linda S. Schadler, Marc Palmeri, Umar Ghumman, Anqi Lin, and Bingyin Hu

Cite This: *ACS Macro Lett.* 2020, 9, 1086−1094 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: With the advent of the materials genome initiative (MGI) in the United States and a similar focus on materials data around the world, a number of materials data resources and associated vocabularies, tools, and repositories have been developed. While the majority of systems focus on slices of computational data with an emphasis on metallic alloys, NanoMine is an open source platform with the goal of curating and storing widely varying experimental data on polymer nanocomposites (polymers doped with nanoparticles) and providing access to characterization and analysis tools with the long-term objective of promoting facile nanocomposite design. Data on over 2500 samples from the literature and individual laboratories has been curated to date into NanoMine, including 230 samples from the papers bound in this virtual issue. This virtual issue represents an experiment of the flexibility of the data repository to capture the unique experimental metadata requirements of many data sets at one time and to challenge the authors to participate in the curation of their research data associated with a given publication. In principle, NanoMine offers a FAIR platform in which data published in papers becomes directly **F**indable and **A**ccessible via simple search tools, with open metadata standards that are **I**nteroperable with larger materials data registries, and allows easy **R**euse of data, e.g. benchmarking against new results. Our hope is that with time, platforms such as this one could capture much of the newly published data on materials and form nodes in an interconnected materials data ecosystem which would allow researchers to robustly archive their data, add to the growing body of readily accessible data, and enable new forms of discovery by application of data analysis and design tools.



The formation of data repositories in the materials domain is a response to the urgent need to harness data driven tools to accelerate materials discovery and design.[1−7] The majority of such data resources are focused on metallic material systems and computational materials data, where software prediction tools can rapidly sweep through compositional space to predict specific structures and properties of interest, thereby providing uniformly aligned data readily classified into metadata structures. Examples of these data resources can be found in a recent perspective article.[8] Experimental and computational data from polymeric material systems offer significant challenges to centralized data platforms because of the complexity and high dimensionality of the data as well as the lack of standardization for many of the experimental and simulation tools and approaches. However, to make progress in polymer design and move away from the expensive trial and error process in the infinite design space, it is essential to meet this challenge.[7] Similarly, for polymer nanocomposites, we need a framework that can robustly categorize and store the vast amount of experimental data currently held in .pdf files of publications and individual research computers in a way that makes it easy to plot, find correlations, and import into other platforms for analysis. NanoMine[9−11] is an example of a living data resource dedicated to nanocomposite research data which has been built using a structured vocabulary, a robust data representation, multiple data ingress pathways, nascent search and visualization tools, and a suite of characterization and analysis tools.

The materials vocabulary used to organize the metadata framework for NanoMine also forms part of the high level "polymer data core"[12] and makes possible indexing from other data stores such as the Materials Data Facility (MDF).[13,14] The metadata was initially built into an XML schema-based

Received: April 3, 2020
Accepted: June 16, 2020

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

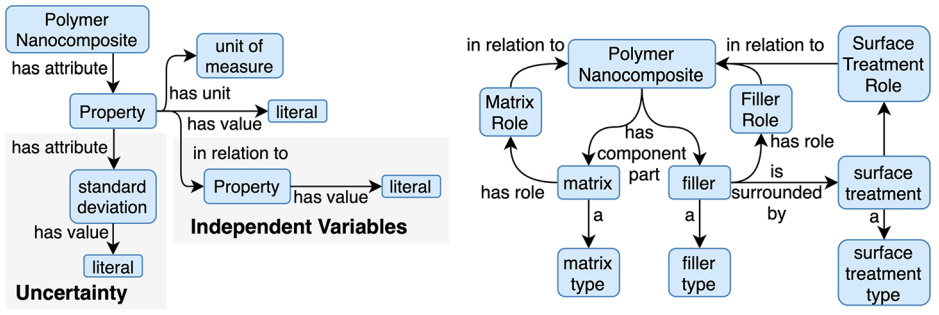


Figure 1. Two conceptual maps of different parts of the NanoMine knowledge graph as a template of how instances of particular types can relate in the graph. In (a), Polymer Nanocomposite instances have attributes of type Property, which can in turn have units of type "unit of measure". In (b), we show that matrix and filler parts of the polymer nanocomposite have a type (like polystyrene or silica) and are specific instances relating to individual samples. These themselves can also have properties, and their role in the nanocomposite is expressed using role objects, like Matrix Role, Filler Role, or Surface Treatment Role. Note that "a" in the graph is shorthand for "type", which is the link between an instance of a class and the class itself.

representation[10,11] starting from the MDCS framework developed by NIST.[15] We extended that schema, maintaining compatibility, and developed an ontology for materials science, leveraging existing science[16] and provenance ontologies.[17] This ontology serves as an extensible knowledge representation model for material science. This approach allows the tools we develop for search, visualization, and data sharing to extend across multiple materials science domains and interoperate with existing standards for scientific metadata. Below, we present details on the data framework, the curation of data into that framework, and visualization and analysis tools, with examples and lessons from the data collected from participating authors with highlighted articles[18−32] in this virtual "Polymer Data" issue of *Macromolecules* and *Macro Letters*. We focus here on the NanoMine data framework. However, the methods, challenges, and opportunities exemplified herein extend generally for this new frontier of FAIR materials data.

Data Framework: The NanoMine data platform began with a nanocomposite specific schema where the structure and the fields were developed using a sampling of papers from the polymer nanocomposites literature. In the initial deployment, over 250 individual terms or parameters were defined in NanoMine to capture and organize data on provenance (e.g., "author name"), composition (e.g., "molecular weight"), processing (e.g., "annealing temperature"), characterization (e.g., "tensile test"), microstructure (e.g., "interparticle distance"), and properties (e.g., "Young's modulus").[10] As data from additional papers were ingested into the platform, new fields were added to the schema on an as-needed basis with a current state of ∼350 individual terms, including ∼10 new terms added in the process of curation for this virtual issue.[11] While the schema approach provided an initial stable organizational platform, to support the larger vision for continued growth, extension of the platform, and ready ability to overlay and extract knowledge, the data terms and structure were transformed into a richer knowledge graph framework supported by an ontology.

Ultimately, we want to store not only data but also knowledge about what the data may imply and how it may be interconnected. For a machine to "understand" a set of data, the data must be properly annotated with metadata describing what the data represent. In Semantic Web standards, this is captured in a machine-readable form as a *knowledge graph*, often stored in a database using the Resource Description Framework (RDF).[33] The knowledge graph comprises a multitude of nodes representing entities in a particular domain along with edges that capture the relationships between pairs of nodes, and every node and edge is defined by a globally unique identifier in the form of a Universal Resource Identifier (URI) such as a Universal Resource Locator (URL). Nodes in the knowledge graph can be things like individual samples, properties, journal articles, or people. Links represent relationships between two nodes and are labeled with a predicate. These "triples" can be expressed as statements (links) such as "Sample X was attributed to Author Y". Each of the entities are identified using URIs and are defined in ontologies which capture the meanings of the terms using a computer understandable language. Critically, ontologies do not need to be fully defined at the outset; nodes and edges (links) in the knowledge graph can be added as the domain evolves and new insights emerge. This structure is flexible and extensible, allowing multiple communities to simultaneously develop and use ontologies that can later be linked together. We followed a standard ontology engineering method for designing the ontology based on the use cases of the project.[34] In Figure 1, we show a conceptual map or diagram that shows how different types (classes) of entities in NanoMine relate to each other using different predicates.

To support this effort, we expanded on the development of the Whyis knowledge graph management framework[35] and have established a materials ontology for nanocomposites.[36] Whyis provides capabilities to translate knowledge from many different sources, including tabular, XML, JSON, and plain text, into consistent, unambiguous integrated knowledge representations. Each nanomaterial sample, property of that sample, material constituents, and all of their relationships are identified using URIs. Further, Whyis uses a representation for managing fragments of knowledge graphs and their provenance, called nanopublications,[37] to ensure consistent versioning, tracking, and justification for each new piece of knowledge. This approach also allows Whyis to implement a distributed knowledge inference system that supports knowledge translation as well as tools for deductive inference, entity recognition and resolution from text. Inference tools can be readily developed and deployed in the knowledge graph. One

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

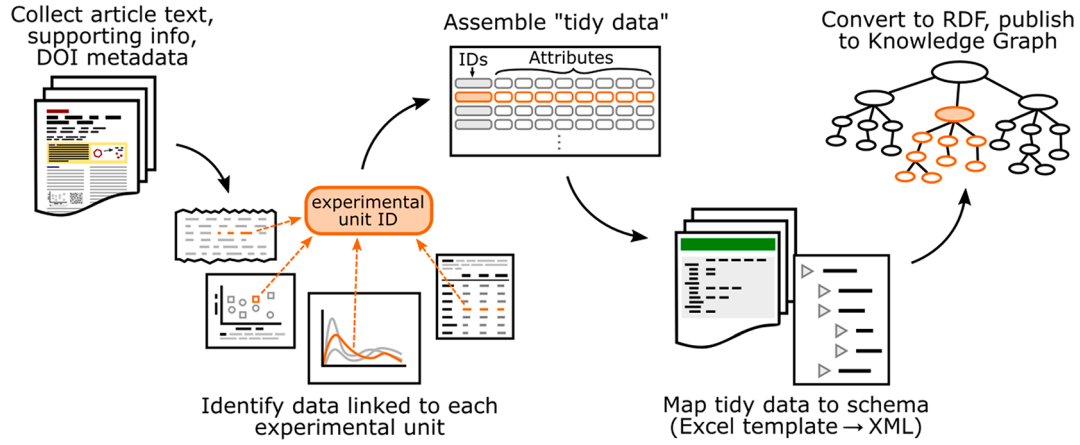


Figure 2. Schematic overview of the workflow for curating semistructured experimental data from the literature into the structured data schema for NanoMine.

example currently under construction is a tool to support automated analysis of microstructural images which can then add new quantitative descriptors of the microstructures of samples to the graph. More complex tools such as machine learning approaches can also be employed, using the rigorous knowledge representation to discover relationships between composition, processing, structure, and properties.

Curation: Curating data "from the wild" (e.g., research articles, images, supporting documents, spreadsheets) into the NanoMine framework (or any data framework) remains a key challenge on several fronts. The curation process must maximize the amount of information extracted from the source data while also ensuring that incoming data meet the standards of quality for the platform. The process must enable retrieval and organization of data from diverse formats reliably into the data framework. Additionally, as the curation process is the initiation point for addition of new fields and descriptors to the data framework, the process must allow for development and extension of the platform.

At present, data curation in NanoMine is performed manually by researchers with domain expertise and familiarity with the NanoMine schema (Figure 2). In development of this virtual issue, we approached a set of authors of papers in *ACS Macro Letters* and *Macromolecules* over the past several years with the offer to highlight their data in this volume. Bringing these external researchers into the process (those without familiarity with the schema but an intimate knowledge of their data) provided valuable insights into knowledge extraction. In past cases, researchers have curated their own data into NanoMine through customized schema templates in the form of Excel workbooks.[11] For this issue, while some authors utilized customized excel templates, several external authors provided their data in a tabular format, which were then mapped by the NanoMine team to the schema. In nearly all cases, bridging the gap between individual data sets and the NanoMine schema involved an iterative, back-and-forth dialogue between these external researchers and the NanoMine team. To make available data accessible and fully leverage the suite of tools provided by NanoMine, the curation platform will be further adapted to new pathways for data ingestion.

At the current rate of global scientific output, the number of cited references in the scientific literature doubles approximately every nine years.[38] According to the Web of Science, there exist nearly 50 000 peer-reviewed publications with the specific topic "polymer nanocomposites" since the field began in the early 1990s with current output approaching 5000 new articles per year. To maintain pace with growing amounts of available data, the process of identifying relevant information from the literature and incorporating into the NanoMine scheme will require a greater degree of automation. New advances in the task of information extraction are continuously being developed and hold promise for automating components of the data curation process in the future. For example, recent research efforts applying modern natural language processing techniques to materials literature have shown promise in automated extraction of synthesis information for inorganic crystalline materials and identification of potential candidate materials of interest.[39−41] Future developments aim to streamline aspects of the curation process and allow researchers to more intuitively import their research results. Moving forward, new tools for data curation are being pursued that aim to more seamlessly integrate data collection, analysis, and sharing into the everyday workflow of materials scientists.

The curation process does not necessarily end when the data have been submitted to NanoMine. Despite any efforts to create a comprehensive schema or ontology, new knowledge or information inevitably arrives that does not conform to the current versions. An open-world assumption, acknowledging that information evolves, and thus any vocabulary is inherently incomplete, is integral to the decision to adopt an ontology-based approach to the materials data framework. Given the flexibility and extensibility of ontologies, we can assign metadata to the materials knowledge graph as that information is needed. In some cases, the addition of new terms to the ontology may enable latent information from earlier curated works to be imported into NanoMine. Furthermore, analysis of individual or aggregated data sets may generate new insights, which themselves may become new metadata in the framework. In the following section, we discuss the ongoing development of visualization and analysis tools that aim to generate additional value from individual and aggregated data sets.

Visualization and Analysis: Once data from an array of samples exists in the knowledge graph framework, it is possible to search, visualize, download and analyze the data in different ways. While these tools are under continuous development, we present here a few examples that intersect with the data from the papers in this virtual issue.

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

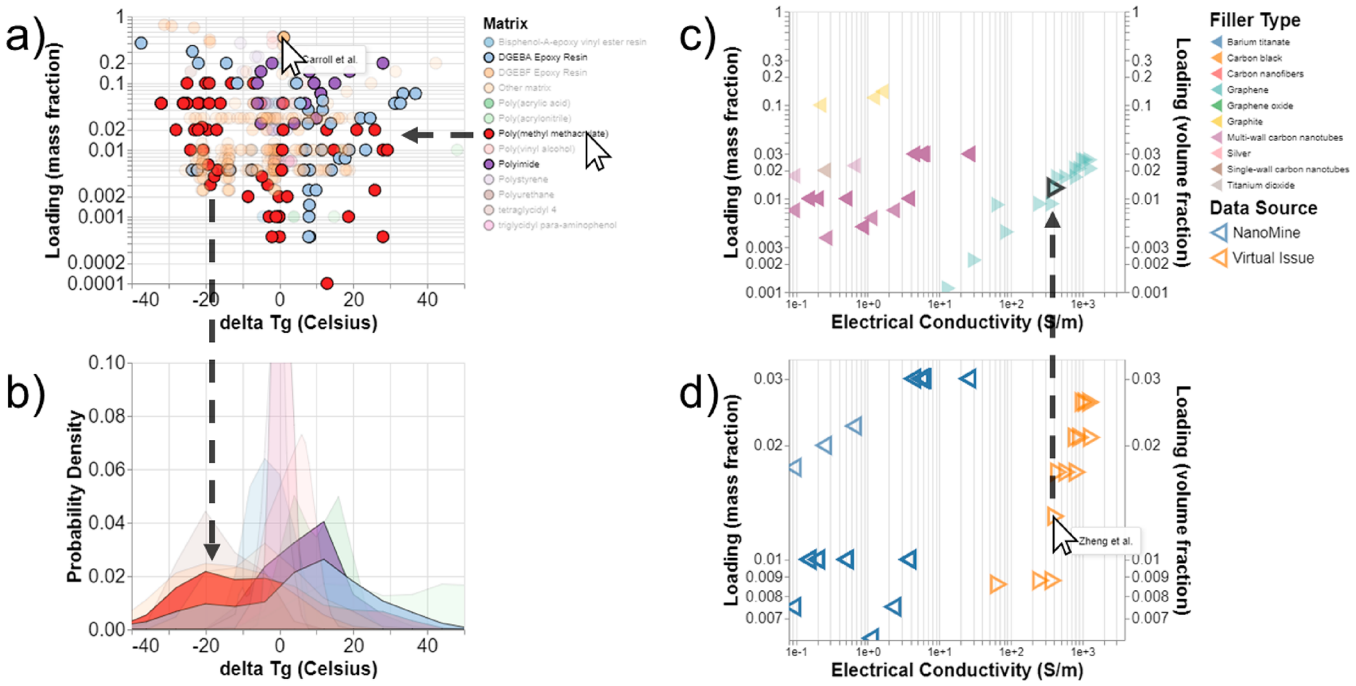


Figure 3. Example interactive charts of aggregated NanoMine data, where linked plots enable simultaneous interaction with multiple views of the data. Dashed arrows are shown here to illustrate interactivity. In (a) and (b), the change in glass transition temperature is displayed as scatter and density plots. Here, DGEBA epoxy, PMMA, and polyimide were selected using shift+click interactions with the legend. In (c) and (d), both scatter plots feature dynamic zooming and highlighting. When the user hovers over a point in the lower plot, the corresponding point in the above plot is highlighted with a bold outline. Hovering over a point in (a) or (d) displays a tooltip with basic provenance information as shown here with data points from Zheng et al.[32] (adapted from ref 32, Copyright 2018 American Chemical Society) and Carroll et al.[20] (adapted from ref 20, Copyright 2017 American Chemical Society), respectively, that were curated into NanoMine for this virtual issue.

Search and Visualization: Users may access NanoMine data through a faceted browser GUI that enables data exploration among select properties and attributes (the facets, e.g., polymer name, physical property of interest, author name, etc.), with ability to intuitively narrow the results according to desired intersections across facets. Visualizations are created using Vega-Lite, a declarative specification language for interactive visual charts.[42] In the faceted browser, users can select properties and material types to narrow their search and then visualize selected properties using simple charts using Vega-Lite. Advanced users can access the raw data directly by querying the publicly available NanoMine SPARQL endpoint and provide Vega and Vega-Lite chart specifications for more advanced visualization. These charts encode the provided data as static elements (e.g., *x*-position, *y*-position, shape, color) as well as interactive elements (e.g., mouse-hover tooltips, mouse-click, or click-and-drag selections). Visitors to the site can browse the *NanoMine Gallery of Interactive Charts*,[43] which showcases examples of NanoMine data visualized through customized Vega-Lite specifications.

Example demonstrations for the NanoMine Gallery were prepared by querying NanoMine data and merging with curated data from select articles in this virtual issue (Figure 3). In the first example, filler loading is plotted against the normalized glass transition temperature, T_g (Figure 3a). The normalization is performed by subtracting the T_g of the neat matrix reference (control value) from the T_g of the polymer nanocomposite sample. By selecting one or more points in the scatterplot, all samples containing the same matrix material are highlighted along with the corresponding probability distributions in the density plot below (Figure 3b). In another example, electrical conductivity is plotted against filler loading (Figure 3c, d). While the selected data shown in the plots span 4 orders of magnitude, a log scale was selected for the *x*-axis to capture the nearly 20 orders of magnitude spanned by all of the electrical conductivity data within NanoMine. While some polymer nanocomposites are engineered as electrical insulators, others are designed for maximum electrical conductivity such as the polystyrene/graphene composites from Zheng et al.[32] in this virtual issue.

Microstructure Characterization and Reconstruction (MCR) Tools: A collection of module tools for MCR and simulation software to model bulk nanocomposite material response in NanoMine augments knowledge generated by experimental data. Integrating these different sources of knowledge from both experiments and simulation is critical for establishing processing−structure−property relationships and subsequently enabling material design. The MCR techniques developed in our research[44−46] have been incorporated into NanoMine to provide parsimonious microstructure analysis workflow for researchers as shown in Figure 4. We provide two popular binarization tools namely, Otsu's method[47] and Niblack's method,[48] as well as three microstructure characterization and reconstruction techniques (correlation function, physical descriptors, and spectral density function),[44] applicable for two-phase, isotropic nanocomposites, and each providing quantitative descriptors of the structure that can be used for data comparison or analysis.

Each tool is accompanied by detailed instructions on how to interact with the tool as well as recommendations for how to select tools best suited for a particular microstructure. The newly added "Intelligent Characterization" tool selects the

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

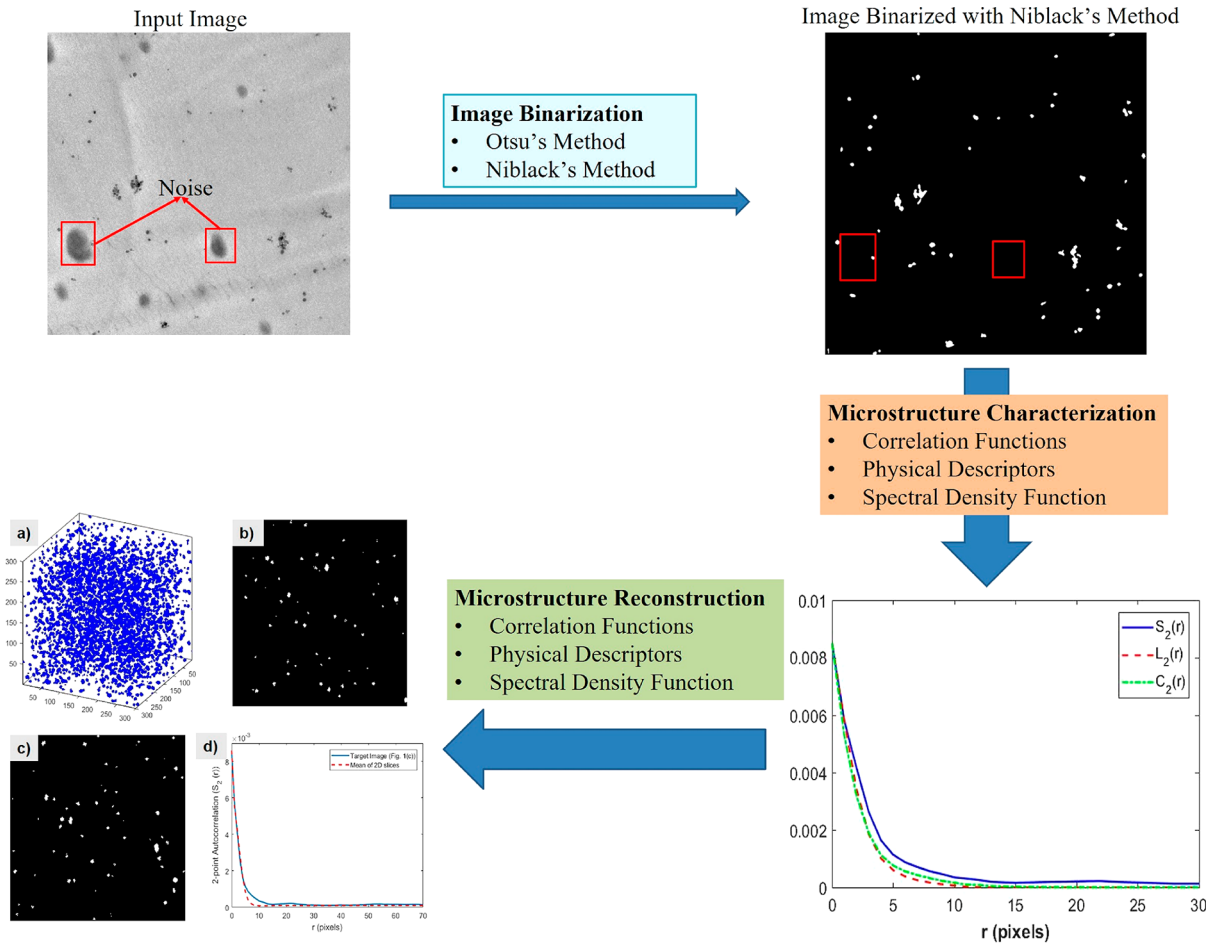


Figure 4. Workflow depicted highlights the key features of MCR tools in NanoMine, starting from a TEM image (upper left[49]) binarized using either the Otsu method or Niblack method (upper right[49]). After that, the microstructure can be characterized using the Correlation Function, Physical Descriptors, or Spectral Density Function (lower right[49]). Lastly, a statistical equivalent of the microstructure image can be reconstructed in 2D or 3D for use in FEA simulations (lower left). Adapted from ref 49, Copyright 2020 World Scientific.

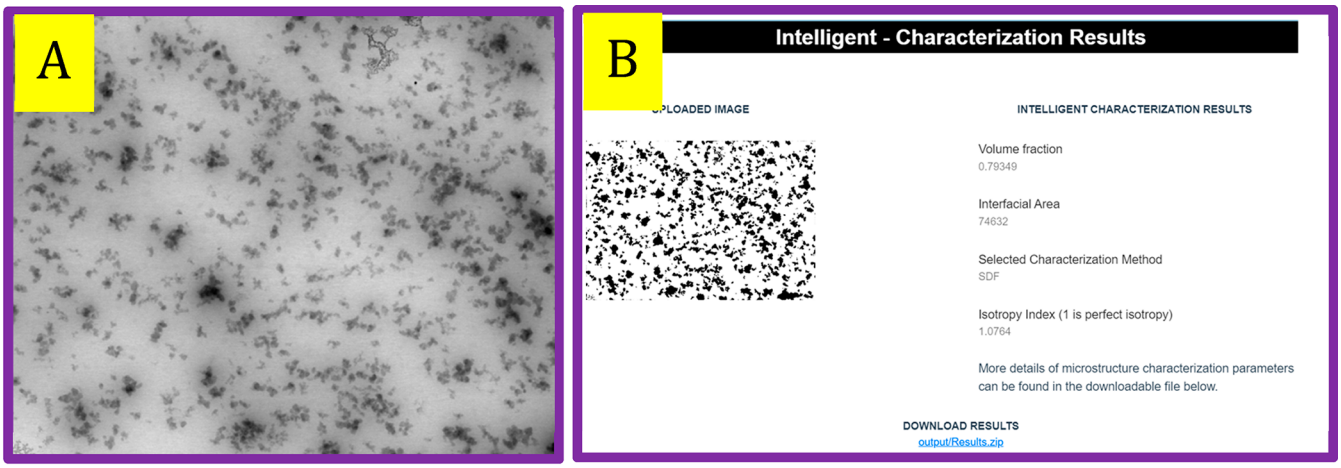
most suitable characterization method between the "physical descriptors"[50] and the "spectral density function (SDF)"[51] approaches based on analyzing the user uploaded image(s). The example result in Figure 5 shows that SDF is preferred over physical descriptors for this particular microstructure. Characterization results, such as those shown in Figure 5c, generated on SDF parametrization, can be easily passed to the NanoMine Database for subsequent use in machine learning and other data mining procedures.

The MCR tools can also be used for reconstructing statistically equivalent 2D or 3D images for a given isotropic microstructure image. Figure 6 shows an example of the result page for microstructure reconstruction using the physical descriptor-based method for a sample from this virtual issue.[22] Minor discrepancies of the two correlation functions (original versus reconstructed) are contributed by the noise in the original image and the approximation introduced by using the descriptors. Such reconstructed images can serve as inputs to structure−property simulations to predict material behavior and therefore are key components of material design workflows.

Our MCR tools have several user-friendly features that make these tools attractive for researchers. For example, all computations are performed on the NanoMine server, all tools support commonly used image file formats and the ability to analyze a batch of images, and NanoMine provides e-mail alerts to users upon completion of their requests. It is also important that the data framework can readily store many images from the same material sample, providing a complete representation of the material microstructure not possible in a typical published paper with limited figures. These batches of images can be processed in the MCR workflow and are very useful in gaining insight into the material heterogeneity through an informative statistical description. This capability was capitalized when extra images that are not part of the original publications were provided by multiple authors for this virtual issue.

Challenges and Opportunities: The papers in this virtual issue represent a successful experiment in the growing web of accessible materials data: the authors contributed data from their publications to a common platform, which encoded the data for each of their samples in a searchable and accessible format, allowing quick visualization of data from these papers integrated with data from a body of previously curated samples. The relative ease of creating such benchmarking visualizations (e.g., Figure 3) demonstrates a critically important function of the knowledge graph. At the same time, this static representation is just the beginning of the analysis possibilities that are enabled by capture of large swaths of data in a robust, accessible framework.

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094



C

Sample name	bw.mat		
Universal descriptors	**Volume fraction**	**Interfacial area**	**Isotropy**
	0.793486	74632	1.076363
Characterization method selected	SDF		
SDF			
Fitting function	Chi2		
PARAM (i)	0.23368	0.257512	3.580483
Fitting goodness	R^2	rmse	
	0.990339	0.001692	
Function definition	PARAM1 * chi2pdf(PARAM2 * X, PARAM3)		
Remarks	PARAM 1 - peak height; 2,3 - curve shape		

Figure 5. (A) Transmission electron microscopy (TEM) image of a nanocomposite reproduced from Rishi et al.,[26] Copyright 2018 American Chemical Society. (B) The results page of Intelligent Characterization using the input image in (A) shows the binarized image as well as some basic information. (C) The output (downloadable excel file) containing detailed description of the SDF parameters.

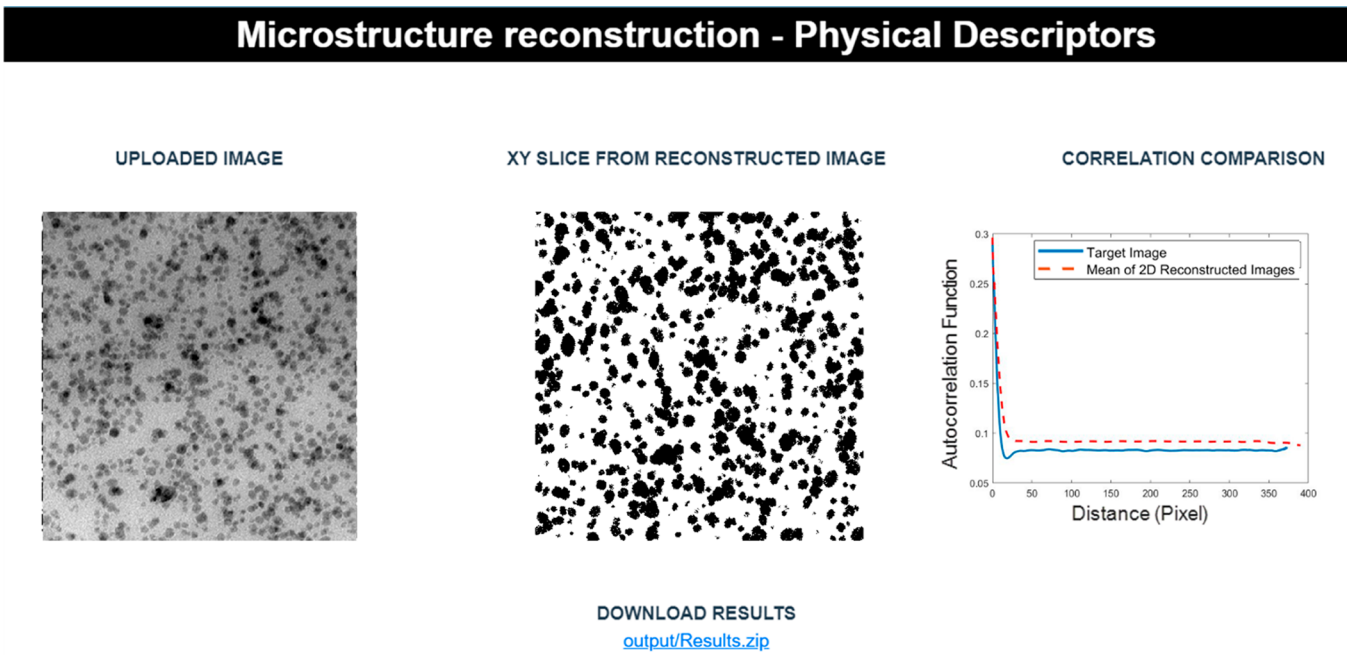


Figure 6. TEM image reconstruction using NanoMine. (Left) TEM micrograph of silica in poly(ethylene oxide) from Jimenez et al.,[22] Copyright 2019 American Chemical Society. (Middle) One 2D slice from the 3D reconstruction of the original image using Physical Descriptor technique. (Right) Comparison of two-point correlation functions between reconstructed and original (target) images.

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

Of the many opportunities enabled by coalescence of materials data in this manner, the most critical surround the concepts of discovery and design. With composition, processing, structure, and property parameters uniformly tagged, it will be possible to apply data analysis methods to shed light on mechanistic underpinnings for interesting phenomena. For example, as the MCR tools are applied to all current and future samples, meaningful and consistent quantification of the microstructure, such as quality of nanoparticle dispersion or alignment, can be used as features in variable importance studies for specific property values. Performing these analyses in conjunction with processing features will allow important processing−structure−property relationships to be highlighted and then targeted for more in depth investigation. The ability to perform these studies over wide ranging data from many different laboratories under many different experimental conditions will open unique opportunities for discovery not currently available in the traditional methods of research. On the front of material design, the knowledge graph framework and assembled data enable new methods for optimization and design of new materials with desired suites of physical properties. Our MCR tools will help extract the most effective material design representations, e.g., descriptors versus spectral density functions, and use them as materials design variables in the iterative optimization process integrated with processing−structure−property relations.

While the vision for discovery and design enabled by data sharing platforms is sweeping and compelling, a number of challenges remain. Most critical among them are rigorous curation of both archived and newly generated data, interoperability across the growing ecosystem of materials data platforms, and usability to ensure researchers can easily access, search, visualize, upload, or download data and information, enabling rapid and detailed analyses across data sets. The growing acceptance and enthusiasm for materials data sharing, and the explosion of the intersection of AI/ML across all sciences, has created a fertile environment and there is common understanding of the importance of interoperability and sharing of frameworks, vocabularies and tools. Encouragement and support from funding agencies on interoperability and platform usability would be especially helpful to leverage the funds spent on the initial platform development and research questions into long-term, sustained benefit for all. A significant challenge is that a large fraction of the essential work needed for usability and interoperability typically falls outside the scope of PhD/postdoc research work.

Perhaps the most significant challenge, especially for experimental data is on the curation side. Curating experimental data into any framework remains a time-consuming bottleneck to realization of large repositories of experimental materials research data in a consistent metadata framework. The benefits to an author of depositing well-annotated data in a repository are significant, including additional avenues for discovery, use and citation of the work by others, the ability to readily benchmark and visualize that data with similar work, and the ready integration with coupled analysis tools. However, the avenues of exposure and networking are not yet well established, and the benefit of the coupled resources requires an investment of time. Overall, the nontrivial overhead of time required to sift, clean, tag, and upload data creates substantial barriers to participation. Therefore, efforts in automated information extraction, crowdsourcing, and data validation are essential to assist with scale and robustness of data imported into NanoMine and other data platforms. The experience of this virtual issue, and assembly of its associated underlying data sets, indicates that new systems and processes for crowdsourcing data into the platform are critical to significantly increase the quantity of data archived and thus the value of the platform.

In summary, NanoMine presents a powerful framework for future materials discovery and design for nanocomposites and provides an example of structure, methods, and tools that generalize to any material system. Lessons learned in the experiment of creating this virtual issue include the value of a flexible underlying structure, the importance of user-friendly search and visualization tools, and the advantage of ready storage of many images for microstructural analysis. The largest challenge facing NanoMine and all data resources for experimental materials data are creation of tools that enable users to easily curate and access the data. Despite the FAIR principles used in developing NanoMine, practical issues such as ease of data ingestion and data query and visualization remain. Solutions include recruiting and supporting the coding community to participate in the ongoing materials genome initiative experiment, partnerships with publishers to unify expected data standards and formats and even requirements from funding agencies on durable accessibility of data. Rich future development opportunities beckon to expand this and other platforms with machine learning and design optimization tools to enable fully automated materials discovery.

AUTHOR INFORMATION

Corresponding Author

L. Catherine Brinson − *Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States;* orcid.org/0000-0003-2551-1563; Email: cate.brinson@duke.edu

Authors

Michael Deagen − *Department of Mechanical Engineering, University of Vermont, Burlington, Vermont 05405, United States*

Wei Chen − *Department of Mechanical Engineering, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-4653-7124

James McCusker − *Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States*

Deborah L. McGuinness − *Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States*

Linda S. Schadler − *Department of Mechanical Engineering, University of Vermont, Burlington, Vermont 05405, United States*

Marc Palmeri − *Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States*

Umar Ghumman − *Department of Mechanical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Anqi Lin − *Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States*

Bingyin Hu − *Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States*

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094

Complete contact information is available at: https://pubs.acs.org/10.1021/acsmacrolett.0c00264

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge support of the NSF CSSI program (OAC-1835648, OAC-1835677, OAC-1835782) and the NSF DMREF program (CMMI-1818574, CMMI-1729743, CMMI-1729452). Ongoing collaborations with NIST, CHiMaD (NIST 70NANB19H005) at Northwestern University, and the Materials Data Facility at Argonne have provided important insight for this work. We are also grateful for the enthusiastic participation of the authors in this virtual issue and the support of the ACS publishing team and editors.

■ REFERENCES

(1) Campbell, C. E.; Kattner, U. R.; Liu, Z.-K. The development of phase-based property data using the CALPHAD method and infrastructure needs. *Integr. Mater. Manuf. Innov.* **2014**, *3*, 158−180.
(2) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, No. 011002.
(3) Lukas, H.; Fries, S. G.; Sundman, B. *Computational Thermodynamics: The Calphad Method*; Cambridge University Press: Cambridge, 2007.
(4) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1*, 15010.
(5) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.
(6) de Pablo, J. J.; Jackson, N. E.; Webb, M. A.; Chen, L.-Q.; Moore, J. E.; Morgan, D.; Jacobs, R.; Pollock, T.; Schlom, D. G.; Toberer, E. S.; Analytis, J.; Dabo, I.; DeLongchamp, D. M.; Fiete, G. A.; Grason, G. M.; Hautier, G.; Mo, Y.; Rajan, K.; Reed, E. J.; Rodriguez, E.; Stevanovic, V.; Suntivich, J.; Thornton, K.; Zhao, J.-C. New frontiers for the materials genome initiative. *npj Comput. Mater.* **2019**, *5*, 41.
(7) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122*, 17575−17585.
(8) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.* **2019**, *6*, 1900808.
(9) NanoMine: Material Informatics for Polymer Nanocomposites. https://materialsmine.org/nm#/ (accessed March 26, 2020).
(10) Zhao, H.; Li, X.; Zhang, Y.; Schadler, L. S.; Chen, W.; Brinson, L. C. Perspective: NanoMine: A material genome approach for polymer nanocomposites analysis and design. *APL Mater.* **2016**, *4*, No. 053204.
(11) Zhao, H.; Wang, Y.; Lin, A.; Hu, B.; Yan, R.; McCusker, J.; Chen, W.; McGuinness, D. L.; Schadler, L.; Brinson, L. C. NanoMine schema: An extensible data representation for polymer nanocomposites. *APL Mater.* **2018**, *6*, 111108.
(12) NIST Materials Resource Registry. https://www.nist.gov/programs-projects/nist-materials-resource-registry; machine readable version https://www.rd-alliance.org/materials-vocabulary-skos-definition (accessed March 30, 2020).
(13) Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The Materials Data Facility: Data Services to Advance Materials Science Research. *JOM* **2016**, *68*, 2045−2052.
(14) Blaiszik, B.; Ward, L.; Schwarting, M.; Gaff, J.; Chard, R.; Pike, D.; Chard, K.; Foster, I. A data ecosystem to support machine learning in materials science. *MRS Commun.* **2019**, *9*, 1125−1133.
(15) Dima, A.; Bhaskarla, S.; Becker, C.; Brady, M.; Campbell, C.; Dessauw, P.; Hanisch, R.; Kattner, U.; Kroenlein, K.; Newrock, M.; Peskin, A.; Plante, R.; Li, S.-Y.; Rigodiat, P.-F.; Amaral, G. S.; Trautt, Z.; Schmitt, X.; Warren, J.; Youssef, S. Informatics Infrastructure for the Materials Genome Initiative. *JOM* **2016**, *68*, 2053−2064.
(16) Dumontier, M.; Baker, C. J.; Baran, J.; Callahan, A.; Chepelev, L.; Cruz-Toledo, J.; Del Rio, N. R.; Duck, G.; Furlong, L. I.; Keath, N.; Klassen, D.; McCusker, J. P.; Queralt-Rosinach, N.; Samwald, M.; Villanueva-Rosales, N.; Wilkinson, M. D.; Hoehndorf, R. The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. *J. Biomed. Semant.* **2014**, *5*, 14.
(17) Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; Zhao, J. PROV-O: The PROV Ontology, 2013. http://www.w3.org/TR/prov-o/ (accessed March 30, 2020).
(18) Bailey, E. J.; Griffin, P. J.; Composto, R. J.; Winey, K. I. Multiscale Dynamics of Small, Attractive Nanoparticles and Entangled Polymers in Polymer Nanocomposites. *Macromolecules* **2019**, *52*, 2181−2188.
(19) Beneš, H.; Popelková, D.; Šturcová, A.; Popelka, Š.; Jůza, J.; Pop-Georgievski, O.; Konefał, M.; Hrubý, M. Aqueous-Based Functionalizations of Titanate Nanotubes: A Straightforward Route to High-Performance Epoxy Composites with Interfacially Bonded Nanofillers. *Macromolecules* **2018**, *51*, 5989−6002.
(20) Carroll, B.; Cheng, S.; Sokolov, A. P. Analyzing the Interfacial Layer Properties in Polymer Nanocomposites by Broadband Dielectric Spectroscopy. *Macromolecules* **2017**, *50*, 6149−6163.
(21) Huang, J.; Chen, X.; Bai, P.; Hai, R.; Sun, C.; Xu, T. 45% Periodicity Reduction in Nanocomposite Thin Films via Rapid Solvent Removal. *Macromolecules* **2019**, *52*, 1803−1809.
(22) Jimenez, A. M.; Krauskopf, A. A.; Pérez-Camargo, R. A.; Zhao, D.; Pribyl, J.; Jestin, J.; Benicewicz, B. C.; Müller, A. J.; Kumar, S. K. Effects of Hairy Nanoparticles on Polymer Crystallization Kinetics. *Macromolecules* **2019**, *52*, 9186−9198.
(23) Kulshreshtha, A.; Modica, K. J.; Jayaraman, A. Impact of Hydrogen Bonding Interactions on Graft−Matrix Wetting and Structure in Polymer Nanocomposites. *Macromolecules* **2019**, *52*, 2725−2735.
(24) Liu, C.; Kadono, H.; Mayumi, K.; Kato, K.; Yokoyama, H.; Ito, K. Unusual Fracture Behavior of Slide-Ring Gels with Movable Cross-Links. *ACS Macro Lett.* **2017**, *6*, 1409−1413.
(25) Ning, X.; Jimenez, A. M.; Pribyl, J.; Li, S.; Benicewicz, B.; Kumar, S. K.; Schadler, L. S. Nanoparticle Organization by Growing Polyethylene Crystal Fronts. *ACS Macro Lett.* **2019**, *8*, 1341−1346.
(26) Rishi, K.; Beaucage, G.; Kuppa, V.; Mulderig, A.; Narayanan, V.; McGlasson, A.; Rackaitis, M.; Ilavsky, J. Impact of an Emergent Hierarchical Filler Network on Nanocomposite Dynamics. *Macromolecules* **2018**, *51*, 7893−7904.
(27) Sadeghi, S.; Arjmand, M.; Otero Navas, I.; Zehtab Yazdi, A.; Sundararaj, U. Effect of Nanofiller Geometry on Network Formation in Polymeric Nanocomposites: Comparison of Rheological and Electrical Properties of Multiwalled Carbon Nanotube and Graphene Nanoribbon. *Macromolecules* **2017**, *50*, 3954−3967.
(28) Sotta, P.; Albouy, P.-A.; Abou Taha, M.; Long, D. R.; Grau, P.; Fayolle, C.; Papon, A. Nonentropic Reinforcement in Elastomer Nanocomposites. *Macromolecules* **2017**, *50*, 6314−6322.
(29) Wanasekara, N. D.; Eichhorn, S. J. Injectable Highly Loaded Cellulose Nanocrystal Fibers and Composites. *ACS Macro Lett.* **2017**, *6*, 1066−1070.
(30) Xia, W.; Qin, X.; Zhang, Y.; Sinko, R.; Keten, S. Achieving Enhanced Interfacial Adhesion and Dispersion in Cellulose Nanocomposites via Amorphous Interfaces. *Macromolecules* **2018**, *51*, 10304−10311.
(31) Zhao, Y.; Liu, J.; Li, X.; Lu, Y.; Wang, S.-Q. How and Why Polymer Glasses Lose Their Ductility Due to Plasticizers. *Macromolecules* **2017**, *50*, 2024−2032.

(32) Zheng, Y.; Chen, W.; Wang, Z.; Wang, Q. Polystyrene/rGO Composites with Orientation-3D Network Binary Structure and Its Surprising Conductivity. *Macromolecules* **2018**, *51*, 7993−8000.
(33) Manola, F.; Miller, E. RDF primer. https://www.w3.org/TR/rdf-primer/ (accessed March 30, 2020).
(34) Kendall, E. F.; McGuinness, D. *Ontology Engineering*; Morgan & Claypool: 2019.
(35) McCusker, J.; Rashid, S. M.; Agu, N.; Bennett, K. P.; McGuinness, D. L. Developing Scientific Knowledge Graphs Using Whyis. SemSci@ISWC, 2018.
(36) https://materialsmine.org/nmf/nanomine.ttl (accessed March 30, 2020).
(37) Groth, P.; Gibson, A.; Velterop, J. The Anatomy of a Nanopublication. *Inf. Serv. Use* **2010**, *30*, 51−56.
(38) Bornmann, L.; Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2215−2222.
(39) Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z. H.; Román-Leshkov, Y.; Moliner, M.; Corma, A.; Olivetti, E. A Machine Learning Approach to Zeolite Synthesis Enabled by Automatic Literature Data Extraction. *ACS Cent. Sci.* **2019**, *5*, 892−899.
(40) Kononova, O.; Huo, H.; He, T.; Rong, Z.; Botari, T.; Sun, W.; Tshitoyan, V.; Ceder, G. Text-mined dataset of inorganic materials synthesis recipes. *Sci. Data* **2019**, *6*, 203.
(41) Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95−98.
(42) Satyanarayan, A.; Moritz, D.; Wongsuphasawat, K.; Heer, J. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 341−350.
(43) NanoMine. https://materialsmine.org/nm#/gallery (Accessed 2020-06-07).
(44) Bostanabad, R.; Zhang, Y.; Li, X.; Kearney, T.; Brinson, L. C.; Apley, D. W.; Liu, W. K.; Chen, W. Computational Microstructure Characterization and Reconstruction: Review of the State-of-the-art Techniques. *Prog. Mater. Sci.* **2018**, *95*, 1−41.
(45) Farooq Ghumman, U.; Iyer, A.; Dulal, R.; Munshi, J.; Wang, A.; Chien, T.; Balasubramanian, G.; Chen, W. A Spectral Density Function Approach for Active Layer Design of Organic Photovoltaic Cells. *J. Mech. Des.* **2018**, *140*, 111408.
(46) Xu, H.; Li, Y.; Brinson, C.; Chen, W. A Descriptor-Based Design Methodology for Developing Heterogeneous Microstructural Materials System. *J. Mech. Des.* **2014**, *136*, No. 051007.
(47) Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst., Man, Cybern. B. Cybern.* **1979**, *9*, 62−66.
(48) Niblack, W. *An Introduction to Digital Image Processing*; Prentice-Hall, Inc.: 1990.
(49) Chen, W.; Schadler, L.; Brinson, C.; Wang, Y.; Zhang, Y.; Prasad, A.; Li, X.; Iyer, A., Materials Informatics and Data System for Polymer Nanocomposites Analysis and Design. In *Handbook on Big Data and Machine Learning in the Physical Sciences*; Kalidindi, S., Kalinin, S. V., Lookman, T., Eds.; World Scientific: 2020; Chapter 3, pp 65−125.
(50) Xu, H.; Dikin, D. A.; Burkhart, C.; Chen, W. Descriptor-based methodology for statistical characterization and 3D reconstruction of microstructural materials. *Comput. Mater. Sci.* **2014**, *85*, 206−216.
(51) Yu, S.; Zhang, Y.; Wang, C.; Lee, W.-k.; Dong, B.; Odom, T. W.; Sun, C.; Chen, W., Characterization and Design of Functional Quasi-Random Nanostructured Materials Using Spectral Density Function. *J. Mech. Des.* **2017**, *139*. DOI: 10.1115/1.4036582

https://dx.doi.org/10.1021/acsmacrolett.0c00264
ACS Macro Lett. 2020, 9, 1086−1094