

Reference ID: 11217048519_Dorfman

Reference ID: 11217048519_Dorfman

Submission Date and Time: 12/12/2019 9:35:48 AM

This contribution was submitted to the National Science Foundation in response to a Request for Information, <https://www.nsf.gov/pubs/2020/nsf20015/nsf20015.jsp>. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Consent Statement: "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publicly available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

Consent answer: I consent to NSF's use and display of the submitted information.

Author Names & Affiliations

Submitting author: Kevin Dorfman - University of Minnesota

Additional authors: Frank Bates, University of Minnesota

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

polymers

Title of Response

Impact on Polymer Research

Abstract

There are significant challenges and limitations in applying data science for polymer science and engineering. These can only be addressed by tight integration of experiment and theory/simulation.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

The key challenge in polymers is the limited ability to produce real materials relative to the ability to simulate polymer properties. At the same time, the ability to produce massive amounts of simulation data for polymers (say, relative to hard materials) is the importance of entropy and non-equilibrium effects on the macroscopic properties of polymers.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

The number of parameters that describe polymers is highly variable (molecular weight, dispersity, composition), which makes it challenging to create coherent data structure that can be easily translated to machine learning, for example. Looking from the theory side, the simulations are typically formulated in terms of universal parameters, such as the statistical segment length and chi parameter. These are helpful for establishing global ways to think about the problem, but challenging to use as predictive tools. These problems are amplified exponentially for block polymers due to the number of possible combinations and architecture of the blocks.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

It is essential that all data-driven approaches to polymers have a very strong connection between the theory/simulation people and experimentalists. One needs to take care in the conversion from one type of data (e.g. simulation) to the other (e.g. experiment) and vice versa. Curation of data, in particularly data quality, is also a challenge that requires expertise in both areas.

-- End Submission --

Response to NSF 20-015, *Dear Colleague Letter: Request for Information on Data-Focused Cyberinfrastructure Needed to Support Future Data-Intensive Science and Engineering Research*

Reference ID: «Respondent_ID»_«Primary_Last»
