



NATIONAL SCIENCE FOUNDATION
4201 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22230

NSF 17-112

Dear Colleague Letter: Data-Driven Discovery Science in Chemistry (D3SC)

July 14, 2017

Dear Colleagues:

In 2016, NSF unveiled a set of "Big Ideas" — 10 bold, long-term research ideas that identify areas for future investment at the frontiers of science and engineering. Among them, the "Harnessing the Data Revolution" idea aims to promote engagement of NSF's research community in the pursuit of fundamental research in data science and engineering, the development of a cohesive, federated, national-scale approach to research data infrastructure, and the development of a 21st-century data-capable workforce.¹ Within this context, the Division of Chemistry (CHE) through this Dear Colleague Letter (DCL) invites submission of research proposals that seek to capitalize on the data revolution and promote data-driven discoveries to advance fundamental understanding of complex chemical systems.

The amount and variety of data generated in the chemical sciences, and the rate at which it is being produced, are rapidly increasing, driving the need for corresponding growth in our ability to extract useful insight from interrelated sources. In order to meet this challenge, the chemistry community must effectively share, mine, and repurpose its rapidly-growing chemical datasets in order to apply state-of-the-art data analysis tools to expand chemical understanding and make modern data science an integral part of the chemical research of the future.

Successful D3SC proposals will emphasize new information that can be obtained from better utilization of data (including data from multiple laboratories, techniques, and/or chemical systems), and how this can lead to new research directions. Proposals that foster and strengthen interactions among chemists (especially experimental chemists) and data scientists to advance research goals are strongly encouraged. The most competitive proposals will provide detailed discussion of specific data-enabled approaches to be used, the significant chemical problem to be studied, new fundamental chemical knowledge to be gained, as well as the broader relevance of the proposed activities to other areas of chemical research. Proposal elements that consider error and uncertainty analysis, record and store appropriate metadata, and determine the robustness and reliability of data are encouraged. Examples of possible topics include (but are not limited to) using tools of data visualization, data mining, machine learning (including emerging approaches such as deep learning and active learning), or other data analysis approaches to:

- Accelerate the discovery of more efficient or selective catalysts;
- Advance the design of new chemical species and/or synthetic reactions, and forecast improved

synthetic conditions;

- Map the mechanisms by which chemicals interact and transform, both covalently and noncovalently, and predict structure/property relations based on existing chemical datasets;
- Discover principles of multiscale organization underlying emergent chemical phenomena in macromolecular systems;
- Enable real-time feedback loops between chemical data collection and processing for rapid identification and correlation of key events during chemical measurements;
- Harness chemistry's rich, diverse but distributed datasets and identify novel ways of sharing and utilizing chemical data derived from multiple instruments, datatypes, and locations;
- Develop innovative approaches for integrating, correlating, and analyzing chemical simulation or measurement data to provide new chemical insights.

Note that the construction or maintenance of large-scale databases per se is not the focus of this DCL, although such databases may be required as a means to the endpoint of using the data to provide insights and predictions. Proposals focused on developing cheminformatics for biomedical or materials research applications are outside the scope of this DCL. Proposals whose primary focus is on the development of new algorithms and software should be submitted to the Computational and Data-Enabled Science and Engineering (CDS&E) program.² Proposals on the development of general-purpose data mining or analysis algorithms not aimed at addressing a specific chemical question are more appropriate for programs supporting general tool development outside of the CHE division.³

Proposals in response to this DCL should be submitted to the existing program of interest in CHE⁴ during the existing submission windows (deadlines) of the programs. The proposal title must be tagged with "D3SC:". Other than the proposal title, the cover page should be prepared as a regular proposal submission to the program. Principal Investigators (PIs) are strongly encouraged to contact the cognizant D3SC Program Officers⁵ prior to submission to determine the appropriateness of the work for consideration.

Proposals may be submitted in combination with other solicitations. For example, proposals may be submitted in combination with the Facilitating Research at Primarily Undergraduate Institutions: Research in Undergraduate Institutions (RUI) and Research Opportunity Awards (ROA) solicitation.⁶ These proposals should be submitted to the appropriate solicitation with "D3SC" added to the title (for example, RUI: D3SC: Name of your proposal). Submission of other types of proposals such as EAGER⁷ (EARly-concept Grants for Exploratory Research) and RAISE⁸ (Research Advanced by Interdisciplinary Science and Engineering) proposals may also be appropriate, but principal investigators are required to check with the cognizant program officers⁵ for additional guidance. If there are strong collaborations with industry, the Grant Opportunities for Academic Liaison with Industry (GOALI)⁹ proposals can be used in conjunction with this effort. For EAGER, RAISE, or GOALI proposals, the title of the proposal should have "EAGER:", "RAISE:", or "GOALI:" specified, followed by the "D3SC:" designation. Proposals including international collaboration are encouraged when those efforts enhance the merit of the proposed work. NSF typically supports the costs of the U.S. team and foreign partners are typically supported by their own funding agencies. Requests for supplemental funding may also be appropriate; again, please check with the cognizant program officers⁸ for additional guidance.

The Division of Chemistry is excited by the opportunities in the D3SC area and looks forward to working with the chemistry community to develop new approaches to gain insights from existing data, as well as new experimental and theoretical results. For recent D3SC awards, please search NSF award database

with the keyword "D3SC". For general questions about this DCL, email the cognizant Program Officers5 at ChemData@nsf.gov.

References

1. "Harnessing Data for 21st Century Science and Engineering" in 10 Big Ideas for Future NSF Investments: https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf.
2. Computational and Data-Enabled Science and Engineering (CDS&E): https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504813.
3. See solicitations for Critical Techniques, Technologies and Methodologies for Advancing Foundation and Application of Big Data Sciences and Engineering (BIGDATA, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767), Data Infrastructure Building Blocks (DIBBS, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504776), and Big Data Regional Innovation Hubs: Establishing Spokes to Advance Big Data Applications (BD Spokes, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505264).
4. CHE research programs: <https://www.nsf.gov/funding/programs.jsp?org=CHE>.
5. D3SC cognizant Program Officers: Lin He (lhe@nsf.gov), David Rockcliffe (drockcli@nsf.gov), Susan Atlas (satlas@nsf.gov), and Robert Cave (rjcave@nsf.gov).
6. Facilitating Research at Primarily Undergraduate Institutions: Research in Undergraduate Institutions (RUI) and Research Opportunity Awards (ROA): https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5518.
7. EARly-concept Grants for Exploratory Research (EAGER): https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_2.jsp#IIE2.
8. Research Advanced by Interdisciplinary Science and Engineering (RAISE): https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_2.jsp#IIE3.
9. Grant Opportunities for Academic Liaison with Industry (GOALI): https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_2.jsp#IIE4.