



NATIONAL SCIENCE FOUNDATION
2415 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22314

NSF 18-076

Dear Colleague Letter: Scalable Cyberinfrastructure to Accelerate Data-Driven Science and Engineering Research

May 14, 2018

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 for 21st Century Science and Engineering (HDR) Big 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 cyberinfrastructure; and the development of a 21st century data-capable workforce.^[1]

In parallel, NSF continues to invest in major facilities and platforms that drive research at national and international scales to catalyze new multi-disciplinary collaborations and discoveries. These investments require a robust, performant, and secure data cyberinfrastructure to address critical needs for the management, integration, delivery and analysis of priority community data to catalyze fundamental discoveries and address complex scientific questions across multiple research domains.

Through this Dear Colleague Letter (DCL), the Office of Advanced Cyberinfrastructure (OAC) encourages submission of proposals to the [Cyberinfrastructure for Emerging Science and Engineering Research \(CESER\) program](#) for scalable data-driven cyberinfrastructure (CI) exemplars that will accelerate discovery for one or more science and engineering research communities, capitalizing on and enhancing existing NSF priority investments. NSF is particularly interested in data CI exemplars that will significantly enhance the scientific utility of data produced by NSF-supported major multi-user research facilities (MMURF). Specific relevance to one or more NSF Big Ideas is highly encouraged.

Successful exemplars will demonstrate capabilities that:

- Address one or more major identified science and engineering research challenges,

particularly in support of NSF Big Ideas;

- Capitalize upon existing NSF investments in data CI, NSF MMURF, and data research;
- Have the potential to rapidly expand or scale capacity and impact within 18 months; and
- Substantially augment scientific impacts within the period of the award.

Competitive proposals will provide detailed discussion of the significant domain research problem(s) to be addressed and the new fundamental knowledge to be gained; the specific data-enabled CI approaches to be used to allow that domain research; and the broader relevance of the proposed activities to other areas of research. Successful proposals will also emphasize new information that can be obtained from better connections among data sources and utilization of data (including data from multiple facilities, techniques, and/or instruments), and how this can lead to new research directions. 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 potential topics include (but are not limited to):

- Incorporating streaming data, intelligent data delivery, and real-time feedback loops between data collection and processing to enable design of smart infrastructures and provision of real-time information for better analysis, visualization, and discovery; and
- Enriching scientific value of community data via integration of diverse and distributed datasets from multiple instruments in novel ways to enhance processing, analysis, sharing, and new science pathways.

Note that the creation or maintenance of large-scale databases per se is not the focus of this DCL, although the development and enhancement of such databases may be proposed as one element towards the ultimate provisioning of data CI that significantly increases the value of, and insights and predictions derived from, these data. Principal investigators (PIs) whose primary focus is on the development of new algorithms and software are encouraged to instead submit to the Computational and Data-Enabled Science and Engineering (CDS&E) program,^[2] although proposers are welcome to include software and middleware development that is integral to accomplishing a data CI.

SUBMISSION PROCESS

Awards pursuant to this DCL will be funded through OAC's [CESER](#) program. It is anticipated that all awards will be made in fall 2018.

Prior to submitting a proposal in response to this DCL, prospective PIs should consult with one or more of the cognizant NSF Program Officers listed below to ascertain that the focus and budget of the proposed work are appropriate.

Full proposals must be submitted via Fastlane or Grants.gov, following the instructions in

NSF's *Proposal and Award Policies and Procedures Guide* (PAPPG; [NSF 18-1](#)). To be eligible for funding in FY 2018, PIs are encouraged to submit proposals to CESER by June 20, 2018.

NSF anticipates making up to 10 awards through CESER pursuant to this DCL, with budgets of up to \$1.5 million and durations up to two years, depending on the quality of proposals and availability of funds.

Cognizant NSF Program Officers are:

- Amy Walton, Tel: (703) 292-4538, Email: awalton@nsf.gov
- Vipin Chaudhary, Tel: (703) 292-2294, Email: vipchaud@nsf.gov
- William Miller, Tel: (703) 292-7886, Email: wlmiller@nsf.gov

Sincerely,
Manish Parashar
Office Director, Office of Advanced Cyberinfrastructure

Erwin Gianchandani
Assistant Director (Acting), Computer and Information Science and Engineering

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.

NSF is an Equal Opportunity Employer Committed to Employing a Highly Qualified Staff that Reflects the Diversity of Our Nation.