National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230

NSF 14-108

Dear Colleague Letter - Joint NSF/ENG and AFOSR Funding Opportunity - EAGERs to Energize Innovative Research and Development on *Dynamic Data Systems*

July 29, 2014

The Division of Electrical, Communications and Cyber Systems (ECCS) of the Engineering Directorate at the National Science Foundation (NSF), in collaboration with the Air Force Office of Scientific Research (AFOSR), seek Early Concept Grants for Exploratory Research (EAGER) proposals with the potential to transform our ability to understand, manage and control the operation of complex, multi-entity natural or engineered systems, through innovative approaches that consider new dimensions in Big Data, Big Computing, and a symbiotic combination of Data and Computing. NSF's and AFOSR's interests lie in highly innovative projects in their early stages that address unique challenges and identify fruitful directions for analytics to transform engineering and scientific practice across various relevant disciplines and scales. The joint interests of both AFOSR and NSF include novel research in technical areas fostered by the DDDAS (Dynamic Data Driven Applications Systems) Program.

The present letter aims to identify opportunities for new capabilities along three key, innovative, and intertwined science and engineering directions for Big Data, Big Computing, and the relation between Data and Computing. Specifically:

- The next wave of Big Data is Dynamic Data arising from ubiquitous sensing and control in engineered and natural systems, through multitudes of heterogeneous sensors and controllers instrumenting these systems. This emerging paradigm calls for seamless and dynamic integration of traditional big data with real-time and archival data relating to ubiquitous sensing and control. The resulting *Large-scale Dynamic Data* environment makes it imperative to dynamically and adaptively manage and schedule these heterogeneous and distributed large-scale data resources, such that these multimodal and multifidelity data are intelligently collected, correlated, and utilized to optimize the understanding, analysis, and operational conditions of engineered and natural multi-entity systems.
- Adding to this important dimension of Dynamic Data is an extended view of Big Computing, which includes a new dimension of computing at the sensor and controller levels. The resulting *Large-scale Big Computing* environment consists of a set of computing platforms ranging from high-end and mid-range computing, to computing on the multitudes of sensors and controllers, with this collective set integrated in a dynamic and seamless manner and viewed as a unified platform.
- In the context of Large-scale Dynamic Data and Large-scale Big Computing, the present letter seeks to foster innovative ideas for new capabilities which result from *dynamic integration* between Data and Computing whereby systems instrumentation data and executing models of these systems are dynamically integrated in a feedback control loop, for improved understanding, analysis, and optimized management of operational conditions of engineered and natural multi-entity systems.

The nature and scope of this initiative are multidisciplinary, drawing from theoretical and applied research in signal processing, machine learning, mathematics, statistics, optimization, complex systems modeling and simulation, distributed and fault-tolerant computing, storage, communications and networking, and large-scale systems engineering and integration. The objective of this EAGER call is to engage

scientists and engineers experienced in big-scale natural or engineered systems as key drivers for the aforementioned innovative directions, and transform our ability to understand and exploit such innovations to enable new capabilities. Examples of possible topics that researchers may investigate on can be found in Reference 1.

EAGER SUBMISSION REQUIREMENTS

1. PI Eligibility: EAGER proposal inquiries will be accepted from a Principal Investigator (PI) or any consortium of investigators led by a PI at an eligible U.S. institution. An investigator may be included in only one submission in response to this Dear Colleague Letter; if more than one is submitted, only the first one submitted will be considered.
2. Deadlines: A) The deadline for submitting EAGER proposals is September 12, 2014. B) No later than August 29, 2014, potential proposers must submit an email , to both of the Program Officers listed below. The email shall include a tentative proposal title, a short abstract, and a list of PI and co-PIs along with their institutions. Use the following subject line in the email: “Intent for Dynamic Data Systems submission.”
3. Submission procedure: Submission of EAGER proposals must be done via Fastlane or Grants.gov following the NSF's Grant Proposal Guide instructions in GPG Chapter II.D.2. The proposal should clearly indicate the reason that the work is appropriate for EAGER support. Proposers must identify the program announcement number for the ECCS division: PD 13-7564. The project title must begin with the tag “EAGER-DynamicData:”.
4. Budgetary information: It is anticipated that each EAGER award may range between $80K and $150K per 12-month year, for up to 24 months. The requested budget must be clearly justified and must be commensurate with the goals of the effort. Estimated program budget, number of awards and average award size/duration are subject to the availability of funds.
5. Proposal processing and review procedure: NSF and AFOSR will convene a joint panel to review the submitted EAGER proposals, and those ideas that best meet the goals of this Dear Colleague Letter will be awarded. Proposers are referred to the website in Reference 1 for additional information.

Technical Points of Contact:

Zhi Tian
Program Director
NSF/ENG/ECCS Division
ztian@nsf.gov
703-292-8839

Frederica Darema
AFOSR
Frederica.Darema@us.af.mil
703-588-1926

References

1. APPENDIX
2. https://community.apan.org/afosr/w/researchareas/7661.dynamic-data-driven-applications-systems-dddas.aspx; http://www.nsf.gov/cise/cns/dddas/: Dynamic Data-Driven Application Systems (DDDAS) -InfoSymbiotic Systems
3. http://www.dtc.umn.edu/bigdata/: NSF Workshop on Big Data-From Signal Processing to Systems Engineering, March 21-22, 2013.

APPENDIX

In 2012 the White House announced the *Big Data Research and Development Initiative* to mobilize the research and development enterprise towards big data analytics for solving some of the nation's most pressing challenges. NSF has responded with Programs which have emphasized the class of Big Data which "come from many disparate sources, including scientific instruments, medical devices, telescopes, microscopes, satellites; digital media including text, video, audio, email, weblogs, twitter feeds, image collections, click streams and financial transactions; dynamic sensor, social, and other types of networks; scientific simulations, models, and surveys; or computational analyses of observational data." The present Dear Colleague Letter (DCL) seeks to create new capabilities through synergistic multidisciplinary research along the three research axes discussed in the DCL.

Specifically, within the context of these three new directions, as articulated in the DCL, environments which involve data across multiple scales, computing across multiple scales, and dynamic integration of data and computing are expected to enable significant science and engineering advances, and yet also give rise to big challenges, some of which are accentuated by the complexity of the scales, modalities, and dynamic integration of data and computing considered here. Systems analysis and decision support often need to be performed in real time in response to the underlying dynamics of the phenomena of interest; as a result, performing analytics on voluminous datasets through centralized processing and storage units becomes infeasible, as the system environments considered here are by definition heterogeneous and distributed, combining many modalities of computing - high-end, real-time, multi-core based platforms and across multiple distributed data systems.

The complex systems and support environments discussed here present a disruptive paradigm for scientists, engineers and optimization experts who are traditionally trained to seek exact or high-quality solutions from modest amounts of data. Large, dynamic and diverse scales of computing and systems instrumentation intrinsically create situations of incomplete and corrupted computations and data. In order to understand, analyze and support such complex systems, data and computation approximation methodologies are required to quickly generate results subject to time constraints and incomplete and corrupted data, while maintaining quantifiable fidelity. For robust, decentralized and online data analytics in the presence of missing data or corrupted measurements, there is an urgent need to explicitly account for the storage, query, time constraint, node failure, and communication burden.

This DCL is aimed to identify new opportunities and foster research and development for novel and advanced analytical and experimental methods and capabilities, based on and exploiting new directions in Big Data and Big Computing articulated above. The following are examples of topics that researchers may consider for developing new theories, advanced computational methods and analytics coupled with innovative experimental techniques. These examples are only provided to engage researchers on novel research and to map out solutions towards key problems with ENG and AFOSR relevance; they are not intended to be prescriptive nor imply any special emphasis or priority.

- Dynamic data driven statistical learning and other numeric and non-numeric multifidelity and scalable methodologies (e.g. PCA and POD) for robust, decentralized and real-time data analytics, where the classical resources of the theory of computation (e.g., time, space, energy, communication capacity) vie with the data resource in dynamic and adaptive ways.
- Methods for adaptive computations of application systems models and simulations, in which dynamic data inputs drive executing models and in reverse to the executing applications controlling the instrumentation. Engineering of systems and simulation models for data collection and utilization by leveraging the unique characteristics and structures of engineered data.
- Multi-entity systems analysis and decision support in the presence of errors or incompleteness in the collected and computed data.
- Identification of appropriate data-centric approximation methodologies to exchange complexity for accuracy in massive, multimodal, multifidelity, decentralized signal and data analysis tasks and

cooperative sensor planning.

- Hardware-software and analog-digital hybrid systems for Big Data and Big Computing applications (as defined in this call). New methodologies for programming across a range of computing and data intensive tasks, integrating real-time communications, and building distributed Large-scale Dynamic Data and Large- scale Big Computing infrastructures.
- Identification of emerging engineering disciplines and appropriate pathways where traditional disciplinary research and development can be revolutionized by Large-scale Dynamic Data and Large-scale Big Computing.