

**Press Release**  
**NSF Leads Federal Efforts In Big Data**

**At White House event, NSF Director announces new solicitation, \$10 million Expeditions in Computing award, and awards in cyberinfrastructure, geosciences, training**

**March 29, 2012**

National Science Foundation (NSF) Director Subra Suresh today outlined efforts to build on NSF's legacy in supporting the fundamental science and underlying infrastructure enabling the big data revolution. At an event led by the White House Office of Science and Technology Policy in Washington, D.C., Suresh joined other federal science agency leaders to discuss cross-agency big data plans and announce new areas of research funding across disciplines in this field.

NSF announced new awards under its Cyberinfrastructure for the 21<sup>st</sup> Century framework, Expeditions in Computing and EarthCube programs, as well as awards that expand statistical approaches to address big data. The agency is also seeking proposals under a Big Data solicitation, in collaboration with the National Institutes of Health (NIH), and anticipates opportunities for cross-disciplinary efforts under its Integrative Graduate Education and Research Traineeship program and an Ideas Lab.

NSF-funded research in these key areas will develop new methods to derive knowledge from data, and to construct new infrastructure to manage, curate and serve data to communities. As part of these efforts, NSF will forge new approaches for associated education and training.

"Data are motivating a profound transformation in the culture and conduct of scientific research in every field of science and engineering," Suresh said. "American scientists must rise to the challenges and seize the opportunities afforded by this new, data-driven revolution. The work we do today will lay the groundwork for new enterprises and fortify the foundations for U.S. competitiveness for decades to come."

NSF released a solicitation, "Core Techniques and Technologies for Advancing Big Data Science & Engineering," or "Big Data," jointly with NIH. This solicitation aims to extract and use knowledge from collections of large data sets in order to accelerate progress in science and engineering research. Specifically, it will develop and evaluate new algorithms, statistical methods, technologies, and tools for improved data collection and management, data analytics and collaboration.

"The Big Data solicitation creates enormous opportunities for extracting knowledge from large-scale data across all disciplines," said Farnam Jahanian, assistant director for NSF's directorate for computer and information science and engineering. "Foundational research advances in data management, analysis and collaboration will change paradigms of research and education, and promise new approaches to addressing national priorities."

One of NSF's awards announced today includes a \$10 million award under the Expeditions in Computing program to researchers at the University of California, Berkeley. The team will integrate algorithms, machines, and people to turn data into knowledge and insight. The objective is to develop new scalable machine-learning algorithms and data management tools that can handle large-scale and heterogeneous datasets, novel datacenter-friendly programming models, and an improved computational infrastructure.

NSF's Cyberinfrastructure Framework for 21st Century Science and Engineering, or "CIF21," is core to strategic efforts. CIF21 will foster the development and implementation of the national cyberinfrastructure for researchers in science and engineering to achieve a democratization of data. In the near term, NSF will provide opportunities and platforms for science research projects to develop the appropriate mechanisms, policies and governance structures to make data available within different research communities. In the longer term, what will result is the integration of ground-up efforts, within a larger-scale national framework, for the sharing of data among disciplines and institutions.

The first round of awards made through an NSF geosciences program called EarthCube, under the CIF21 framework, was also announced today. These awards will support the development of community-guided cyberinfrastructure to integrate big data across geosciences and ultimately change how geosciences research is conducted. Integrating data from disparate locations and sources with eclectic structures and formats that has been stored as well as captured in real time, will expedite the delivery of geoscience knowledge.

"EarthCube is a groundbreaking NSF program," said Tim Killeen, assistant director for NSF's geosciences directorate. "It represents a dynamic new way to access, share and use data of all types to accelerate and transform research for understanding our planet. We are asking experts from all sectors--industry, academia, government and non-U.S. institutions--to form collaborations and tell us what research topics they think are most important. Their enthusiastic and energetic response has resulted in a synergy of exhilarating and novel ideas."

NSF also announced a \$1.4 million award for a focused research group that brings together statisticians and biologists to develop network models and automatic, scalable algorithms and tools to tell us more about protein structures and biological pathways.

And, a \$2 million award for a research training group in big data will support training for undergraduates, graduates and postdoctoral fellows to use statistical, graphical and visualization techniques for complex data.

"NSF is developing a bold and comprehensive approach for this new data-centric world, from fundamental mathematical, statistical and computational approaches needed to understand the data, to infrastructure at a national and international level needed to support and serve our communities, to policy enabling rapid dissemination and sharing of knowledge," said Ed Seidel, assistant director for NSF's mathematical and physical sciences directorate. "Together, this will accelerate scientific progress, create new possibilities for education, enhance innovation in society and be a driver for job creation. Everyone will benefit from these activities."

In addition, anticipated cross-disciplinary efforts at NSF include encouraging data citation to increase opportunities for the use and analysis of data sets; participation in an ideas lab to explore ways to use big data to enhance the effectiveness of teaching and learning environments; and the use of NSF's Integrative Graduate Education and Research Traineeship, or IGERT, mechanism to educate and train researchers in data enabled science and engineering.

A full list of NSF data-enabled science and engineering projects may be found at: <http://www.nsf.gov/>.

A recording of the Big Data Rollout may be viewed at: <http://live.science360.gov/bigdata/>.

-NSF-

---

**Media Contacts**

Lisa-Joy Zgorski, NSF (703) 292-8311 [lisajoy@nsf.gov](mailto:lisajoy@nsf.gov)

**Program Contacts**

C. Suzanne Iacono, NSF (703) 292-8900 [siacono@nsf.gov](mailto:siacono@nsf.gov)