NSF Leads Federal "Big Data" Efforts

On March 29 2012, National Science Foundation (NSF) director Subra Suresh announced plans to expand NSF's funding, infrastructural, and scientific support of "big data" science. The announcement was made at an event held by the White House Office of Science and Technology Policy in Washington, D.C. The NSF was joined by the NIH and other federal agencies for the event.

Some people refer to this type of big data science as "e-science," though I've always preferred the term "big data science." Whichever label you use, it covers scientific disciplines from genomics to astronomy to geosciences and weather pattern modeling. Challenges to scientists involve storage, definition, and sharing of these massive datasets, and creation of statistical methods and algorithms for analyzing the data. The NSF requires all researchers applying for grants to submit Data Management Plans, and the NIH has a Data Sharing Policy: all investigator-initiated applications with direct costs greater than $500,000 in any single year are expected to address data sharing in their application or state why data sharing is not possible.

In his announcement, Suresh outlined details for a specific solicitation in joint partnership with the NIH for proposals for the program *Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)*. This program "aims to advance the core scientific and technological means of managing, analyzing, visualizing and extracting useful information from large, diverse, distributed and heterogeneous data sets."

In addition to the BIGDATA solicitation, other programs in the Big Data effort include:

- Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)
- CIF21 Track for IGERT
- Data Citation
- Data and Software Preservation for Open Science (DASPOS)
- Digging into Data Challenge
- EarthCube
- Expeditions in Computing
- Focused Research Group, stochastic network models
- Ideas Lab
- Information Integration and Informatics
- The Computational and Data-enabled Science and Engineering (CDS&E) in Mathematical and Statistical Sciences (CDS&E-MSS)
- Some Research Training Groups (RTG) and Mentoring through Critical Transition Points (MCTP) relate to big data
- Laser Interferometer Gravitational Wave Observatory (LIGO)
- The Open Science Grid (OSG)
- The Theoretical and Computational Astrophysics Networks (TCAN)

To see brief descriptions of these programs, and to read details of Suresh's announcement, see the announcement page on the NSF website.

You can also find more information on federal big data initiatives at the White House Office of Science and Technology Policy website.

Related Galter Library Biosciences and Bioinformatics Blog Posts:

- White House Releases Two Public Consultations on Open Access to Data and Publications
- What is E-Science?

Pamela Shaw
Biosciences & Bioinformatics Librarian
email Pamela

The Biosciences & Bioinformatics Blog highlights new tools and news items of interest to the biosciences research community at Northwestern University.