Tech America Congressional Briefing
Big Data: What it Means and How it Drives Innovation
May 2, 2012 10:05-10:15 a.m.
Duration: 10 minutes, and Q&A with panel

- Thank you very much, Kevin. Good morning. It is my pleasure ...

- Advances in information technologies are transforming the fabric of our society and **data** represents a *transformative new currency* for science, engineering, education and commerce.

- To optimize the ability to extract knowledge and insights from this data, the White House Office of Science and Technology Policy led federal science agencies in the launch of a "big data" research and development initiative just over a month ago.

But what is Big Data?

- "Big Data" is characterized not only by the enormous **volume** of data but also by the **diversity** and **heterogeneity** of the data and the **velocity** of its generation.

- This explosive growth in the size, complexity, and rate of digital data is generated/produced/created by:
    - modern experimental methods and observational (**longitudinal**) studies,
    - large-scale simulations,
    - scientific instruments such as telescopes and particle accelerators,
    - Internet transactions,
    - email, videos, images, click streams...

- o Not to mention … ubiquitous widespread deployment of **sensors** all around us:
    - in the environment,
    - in our critical infrastructure such as bridges and smart grids,
    - in our homes, and even on our clothing!

- There are enormous opportunities to harness these large-scale diverse data sets, to **extract knowledge** from them, and to provide powerful new approaches to **drive discovery** and **decision-making**, and to make increasingly **accurate prediction** based on the data.

- The President's Council of Advisors on Science and Technology recently concluded that the Federal government needs to:

  1. increase R&D investments for collecting, storing, preserving, managing, analyzing, and sharing the increasing quantities of data, and

  2. Futhermore, PCAST observed that the potential to gain new insights … to move from data to knowledge to action has tremendous potential to transform all areas of national priority.

Why is Big Data important?

- First, insights and more accurate predictions from large and complex collections of data have important implications for the economy.
    - Access to information is transforming traditional businesses and is creating opportunities in new markets.
    - Big Data is driving the creation of new IT products and services based on business intelligence and data analytics, and ... is boosting the productivity of firms that use it to make better decisions and identify new business trends.

- Second, advances in Big Data are critical to accelerate the pace of discovery in almost every science and engineering discipline. From new insights about protein structure, biomedical research and clinical decision-making, and climate modeling, to new ways to mitigate and respond to natural disasters, and develop new strategies for effective learning and education – there are enormous opportunities for data-driven discovery!

- Third, Big Data also has the potential to solve some of the Nation's most pressing challenges -- in science, education, environment and sustainability, [government,] medicine (OR healthcare), commerce, and cyber and national security -- with enormous societal benefit and laying the foundations for U.S. competitiveness for many decades to come.

What does this mean for scientific discovery?

- Data are motivating a profound transformation in the culture and conduct of scientific research. Data-driven discovery is revolutionizing scientific exploration and engineering innovations.

What is NSF doing?

- NSF's strategic efforts for Big Data build on NSF's legacy supporting the fundamental science and underlying infrastructure enabling big data science and engineering. Our strategy includes: (1) advances in foundational techniques and technologies (that is, new methods) to derive knowledge from data; (2) cyberinfrastructure to manage, curate and serve data to science and engineering research and education communities; (3) new approaches to education and workforce development; and (4) support for [OR nurturing] new types of collaborations -- multi-disciplinary teams and communities enabled by new data access policies -- to make advances in the grand challenges of the computation- and data-intensive world today.

- The cornerstone of this strategy is a joint NSF – NIH solicitation, *Core Technologies and Techniques for Advancing Big Data Science & Engineering,* or *Big Data*, which 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.

- Three foundational research areas are targeted through this NSF-NIH programs are: (1) large-scale data collection and management, (2) data analytics, and (3) e-science collaboration environments.

- Data access and analysis is already having enormous impact. You can imagine a day when:
    - Complete health/disease/genome/environmental knowledge bases enable biomedical discovery, and patient-centered therapy.
      The data can be mined to spot unwanted drug interactions, or predict onset of diseases.
    - By linking together finance, human resources, supply chain, customer management systems, companies can use data mining techniques to get a complete picture of their operations – identify new business trends, operate more efficiently, and improve forecasting.
    - Accurate high-resolution models support forecasting and management of increasingly stressed watersheds and ecosystems;
    - Consumers have the information they need to make optimal energy consumption decisions in their homes and cars;
    - Accurate predictions of natural disasters, such as earthquakes, hurricanes, and tornadoes, enable life-saving and cost-saving preventative actions;
    - Our cyber-enabled world is safe, secure, and private, enabling assured use of our critical infrastructure and on-line commerce; and
    - … why it's transformative for education and WFD:
    - Students and researchers have access to intuitive tools to view, understand, and learn from publicly-available, large scientific data sets on everything from genome sequences to astronomical star surveys, from public health databases to particle accelerator simulations, and our teachers and educators use student performance analytics to improve learning and enhance assessment.

- Realizing the enormous potential of Big Data requires a long-term bold, sustainable and comprehensive approach, not only by NSF, but also throughout the government and our nation's research institutions.

- All of this is enabled by a thriving discovery and innovation ecosystem that includes:
    - Leading-edge universities and research labs;
    - Scientists and engineers in a flexible talent-rich labor market;
    - Investments in research and education - the building blocks of discovery and innovation; and
    - A vibrant private sector catalyzed by American entrepreneurial spirit.

- The work we do today will lay the groundwork for new enterprises, promote economic growth, improve our citizens' quality of life, and fortify the foundations for U.S. competitiveness for decades to come.

- Thank you again for your attention.