



Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE: ANNOUNCES \$200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation’s most pressing challenges.

To launch the initiative, six Federal departments and agencies today announced more than \$200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.

“In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education, and national security,” said Dr. John P. Holdren, Assistant to the President and Director of the White House Office of Science and Technology Policy.

To make the most of this opportunity, the White House Office of Science and Technology Policy (OSTP)—in concert with several Federal departments and agencies—created the Big Data Research and Development Initiative to:

- Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data.
- Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning; and
- Expand the workforce needed to develop and use Big Data technologies.

Today's initiative responds to recommendations by the President's Council of Advisors on Science and Technology, which last year concluded that the Federal Government is under-investing in technologies related to Big Data. In response, OSTP launched a Senior Steering Group on Big Data to coordinate and expand the Government's investments in this critical area. Today's announcement describes the first wave of agency commitments to support this initiative, including:

National Science Foundation and the National Institutes of Health - Core Techniques and Technologies for Advancing Big Data Science & Engineering

"Big Data" is a new joint solicitation supported by the National Science Foundation (NSF) and the National Institutes of Health (NIH) that will advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large and diverse data sets. This will accelerate scientific discovery and lead to new fields of inquiry that would otherwise not be possible. NIH is particularly interested in imaging, molecular, cellular, electrophysiological, chemical, behavioral, epidemiological, clinical, and other data sets related to health and disease.

National Science Foundation: In addition to funding the Big Data solicitation, and keeping with its focus on basic research, NSF is implementing a comprehensive, long-term strategy that includes new methods to derive knowledge from data; infrastructure to manage, curate, and serve data to communities; and new approaches to education and workforce development. Specifically, NSF is:

- Encouraging research universities to develop interdisciplinary graduate programs to prepare the next generation of data scientists and engineers;
- Funding a \$10 million Expeditions in Computing project based at the University of California, Berkeley, that will integrate three powerful approaches for turning data into information - machine learning, cloud computing, and crowd sourcing;
- Providing the first round of grants to support "EarthCube" – a system that will allow geoscientists to access, analyze and share information about our planet;
- Issuing a \$2 million award for a research training group to support training for undergraduates to use graphical and visualization techniques for complex data.
- Providing \$1.4 million in support for a focused research group of statisticians and biologists to determine protein structures and biological pathways.
- Convening researchers across disciplines to determine how Big Data can transform teaching and learning.

Department of Defense – Data to Decisions: The Department of Defense (DoD) is "placing a big bet on big data" investing approximately \$250 million annually (with \$60 million available for new research projects) across the Military Departments in a series of programs that will:

- Harness and utilize massive data in new ways and bring together sensing, perception and decision support to make truly autonomous systems that can maneuver and make decisions on their own.

- Improve situational awareness to help warfighters and analysts and provide increased support to operations. The Department is seeking a 100-fold increase in the ability of analysts to extract information from texts in any language, and a similar increase in the number of objects, activities, and events that an analyst can observe.

To accelerate innovation in Big Data that meets these and other requirements, DoD will announce a series of open prize competitions over the next several months.

In addition, the Defense Advanced Research Projects Agency (DARPA) is beginning the XDATA program, which intends to invest approximately \$25 million annually for four years to develop computational techniques and software tools for analyzing large volumes of data, both semi-structured (e.g., tabular, relational, categorical, meta-data) and unstructured (e.g., text documents, message traffic). Central challenges to be addressed include:

- Developing scalable algorithms for processing imperfect data in distributed data stores; and
- Creating effective human-computer interaction tools for facilitating rapidly customizable visual reasoning for diverse missions.

The XDATA program will support open source software toolkits to enable flexible software development for users to process large volumes of data in timelines commensurate with mission workflows of targeted defense applications.

National Institutes of Health – 1000 Genomes Project Data Available on Cloud:

The National Institutes of Health is announcing that the world's largest set of data on human genetic variation – produced by the international 1000 Genomes Project – is now freely available on the Amazon Web Services (AWS) cloud. At 200 terabytes – the equivalent of 16 million file cabinets filled with text, or more than 30,000 standard DVDs – the current 1000 Genomes Project data set is a prime example of big data, where data sets become so massive that few researchers have the computing power to make best use of them. AWS is storing the 1000 Genomes Project as a publically available data set for free and researchers only will pay for the computing services that they use.

Department of Energy – Scientific Discovery Through Advanced Computing:

The Department of Energy will provide \$25 million in funding to establish the Scalable Data Management, Analysis and Visualization (SDAV) Institute. Led by the Energy Department's Lawrence Berkeley National Laboratory, the SDAV Institute will bring together the expertise of six national laboratories and seven universities to develop new tools to help scientists manage and visualize data on the Department's supercomputers, which will further streamline the processes that lead to discoveries made by scientists using the Department's research facilities. The need for these new tools has grown as the simulations running on the Department's supercomputers have increased in size and complexity.

US Geological Survey – Big Data for Earth System Science: USGS is announcing the latest awardees for grants it issues through its John Wesley Powell Center for Analysis and Synthesis. The Center catalyzes innovative thinking in Earth system science by providing scientists a place and time for in-depth analysis, state-of-the-art computing capabilities, and collaborative tools invaluable for making sense of huge data sets. These Big Data projects will improve our understanding of issues such as species response to climate change, earthquake recurrence rates, and the next generation of ecological indicators.

Further details about each department's or agency's commitments can be found at the following websites by 2 pm today:

NSF: http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607

HHS/NIH: <http://www.nih.gov/news/health/mar2012/nhgri-29.htm>

DOE: <http://science.energy.gov/news/>

DOD: www.DefenseInnovationMarketplace.mil

DARPA: <http://www.darpa.mil/NewsEvents/Releases/2012/03/29.aspx>

USGS: <http://powellcenter.usgs.gov>

###

OSTP was created by Congress in 1976 to serve as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government.

For more information about OSTP, visit WhiteHouse.gov/OSTP