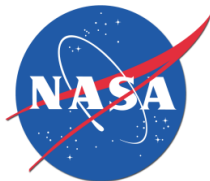




Federal Networking and Information Technology R&D Program



Big Data is a Big Deal
Big Data Senior Steering Group
Wendy Wigen, Technical Coordinator
May 3, 2012



Member Agencies

◆ DOC

- NIST
- NOAA

◆ DoD

- DARPA
- NSA
- OSD
- Service Research Orgs

◆ DOE

- NNSA
- SC

◆ DHHS

- AHRQ
- NIH
- ONC

◆ Independent Agencies

- DHS
- EPA
- NASA
- NARA
- NSF

"We don't have a data problem... we have an analysis problem..."

Or

"We have a filtering problem..."

Or

"We have an opportunity..."

Big Data is a term applied to data sets whose size or complexity is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time.

Big Data Senior Steering Group's (BDSSG) Purpose

- ◆ Describe what a Big Data National Initiative would look like, its vision, goal, scope.
- ◆ Develop four main areas:
 - Core Technologies
 - Domain Research Projects
 - Challenges/Competitions
 - Workforce Development

Vision

A future in which the ability to analyze and extract information from large, diverse, and disparate data sets accelerates the process of scientific discovery and innovation; promotes new economic growth; and leads to new fields of research and new areas of inquiry that would otherwise be impossible.

Goals

- ◆ Promote New Science...by harnessing big data.
- ◆ Exploit big data to address national needs and agency missions.
- ◆ Support stewardship of federal data.
- ◆ Develop the necessary workforce and infrastructure to advance data science.

Core Technologies

- ◆ Solicitation: *Core Technologies and Technologies for Advancing Big Data Science & Engineering (BIGDATA)* NSF 12-499
- ◆ <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>
- ◆ 9 NSF Directorates and 7 NIH Institutes

March 29 White House Event

◆ To Launch the Big Data R&D Initiative

http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf

◆ To announce six new Big Data programs

- NSF: http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607
- HHS/NIH: <http://www.nih.gov/news/health/mar2012/nhgri-29.htm>
- DOD: www.DefenseInnovationMarketplace.mil
- DARPA: <http://www.darpa.mil/NewsEvents/Releases/2012/03/29.aspx>
- USGS: <http://powellcenter.usgs.gov>

White House Event con't

◆ To highlight existing programs

➤ Big Data Fact Sheet:

<http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>

➤ 87 current programs identified

➤ DoD, DHS, DOE, HHS, NASA, NARA, NSF

Challenges/Competitions

- ◆ Series of Challenges: Ideation -> New Tool
- ◆ Workshops to decide the parameters
- ◆ NASA Center of Excellence for Collaborative Innovation - > Plan (April 12)
- ◆ BD SSG Approval and Funding

Recap of the Challenge

- ◆ Develop tools or solutions that allow heterogeneous collections of data to become more homogeneous and searchable “on the fly” or “at first presentation.”
- ◆ Involves more than one agency and the resulting tool could be easily generalized for use by multiple agencies.
- ◆ Incorporates agency mission statements.
- ◆ Involves more than one data set, of which:
 - At least one must be a “big data” data set
 - At least one must be an active or streaming data set (this could be a requirement or an option)
- ◆ Promotes Data to Knowledge to Action

Domain Research Projects

- ◆ Identify current projects that could benefit from cross-agency collaboration.
- ◆ Propose new cross-agency projects.
- ◆ *Examples:* Earth Cube, XCEDE, NIH Electronic health record Collaboratory, Climate Change and Health

Workforce Development

- ◆ 20 projects across 7 agencies that may be suitable for adapting to Big Data, (Grants, Fellowships, Summer Internships, Scholarships, etc.)
- ◆ Undergraduate, Doc, Post-Doc, Mid-Career
- ◆ Building a Data Science Community: Meetings at annual conferences, professional associations etc.

Contacts

◆ Wendy Wigen, Technical Coordinator

◆ wigen@nitrd.gov

◆ Co-Chairs:

➤ Karin Remington, NIH

➤ Suzanne Iacono, NSF