Unleashing the Power of Big Data: Why We Need All Hands on Deck

Thomas Kalil
Deputy Director for Technology and Innovation
White House Office of Science and Technology Policy &
National Economic Council
tkalil@ostp.eop.gov

May 3, 2013

Administration interest in Big Data (1)

- Economic impact – demonstrated productivity increases for firms that are mastering Big Data and analytics
- Accelerate pace of discovery in science and engineering
- National security/intelligence
- Health – better care at lower costs, new infrastructure for clinical research + drug discovery that integrates EHRs, m-health, omics

Administration interest in Big Data (2)

- Education:
  - Move from data-poor to data-rich domain.
  - Online courses that improve the more students use them.
  - Online courses as "wind tunnel" for learning science
- Smart grid, energy efficiency
- Real-time labor market information

Government role (1)

- Invest in R&D related to Big Data technologies
- Support efforts to expand “Big Data” workforce (e.g. data scientists)
- Support applications of Big Data in national priorities and for agency missions, be a “smart user” of Big Data approaches
- Use challenges to demonstrate what is possible (see, e.g. http://www.kaggle.com and http://www.topcoder.com

Government role (2)

- Policy issues:
  - –Privacy
  - –Transborder data flow
  - –Global applications (see, e.g. UN Global Pulse)

- Make more government data available in bulk-downloadable, machine-readable format consistent with President's Open Government Initiative

- Be a catalyst for investments by private sector

March 29, 2012 announcement

- Over $200 million in investments in R&D related to Big Data, both grants and solicitations
- DARPA XDATA program
- Joint solicitation between National Science Foundation and National Institutes of Health
- Department of Energy-funded institute in scientific visualization
- NIH 1,000 Genome Project, joint with European Bioinformatics Institute

All Hands on Deck

- Make the "big data" effort a national initiative as opposed to a federal initiative.

- Identify steps that all stakeholders (companies, regions, early adopters, universities, researchers, investors, non-profits, foundations, professional societies, skilled volunteers) have taken and can take to make the most of the opportunities created by Big Data

- High-profile event in Fall 2013 to celebrate and catalyze new and expanded commitments

Examples of actions to advance Big Data (1)

- Universities:
  - New courses and courses of study – e.g. CMU Department of Machine Learning, Northwestern Master's in Analytics, NCSU MSA
  - MOOCs
- University-industry collaborations
  - MIT + Intel – bigdata@CSAIL
  - UC Berkeley – AMPLAB – NSF + 17 industry collaborators
- Regional efforts
  - Massachusetts Big Data Initiative
- Data challenges (e.g. Hewlett Foundation on automated student assessment)

Examples of actions to advance Big Data (2)

- Open sourcing Big Data software, publishing papers on key results from private R&D (e.g. Hadoop, R, Hbase, Cassandra, Drill, etc.)

- Data philanthropy
  - Orange making 2.5 billion anonymized cellphone records from Cote D'Ivoire available to researchers

- Collaboration with Big Data oriented non-profits (e.g. Datakind)

Examples of actions to advance Big Data (3)

- Collaborations to demonstrate value of Big Data (e.g. treat and prevent risk factors for stroke, heart attack, diabetes)
  - Pilots with rigorous evaluation
  - Strategies for scaling if successful
- Help with national agenda-setting (e.g. Computing Community Consortium, TechAmerica Foundation)
- Your idea here