

Big Data and Privacy

Rebecca Balebako

Advisor: Dr. Lorrie Cranor

**Engineering &
Public Policy**



Agenda

- Quiz
- Important dates
- Exam feedback
- Benefits of big data
- Concerns about big data
- What is PCAST
- PCAST big data recommendations

black	1965	male	02141	short of breath
black	1965	male	02141	chest pain
black	1965	female	02138	painful eye
black	1965	female	02138	wheezing
black	1964	female	02138	obesity
black	1964	female	02138	chest pain
white	1960-69	male	02138	short of breath
white	1960-69	human	02139	hypertension
white	1960-69	human	02139	obesity
white	1960-69	human	02139	fever
white	1960-69	male	02138	vomiting
white	1960-69	male	02138	back pain

Important Dates

- November 18th – Homework 7 due
 - Time consuming: **start today**
- November 20 – Draft project paper due
- Dec 4th – Poster Fair
- Dec 11th – Final exam 1-4pm
- Dec 12th – Final project due

Comments on Exam

Benefits of big data

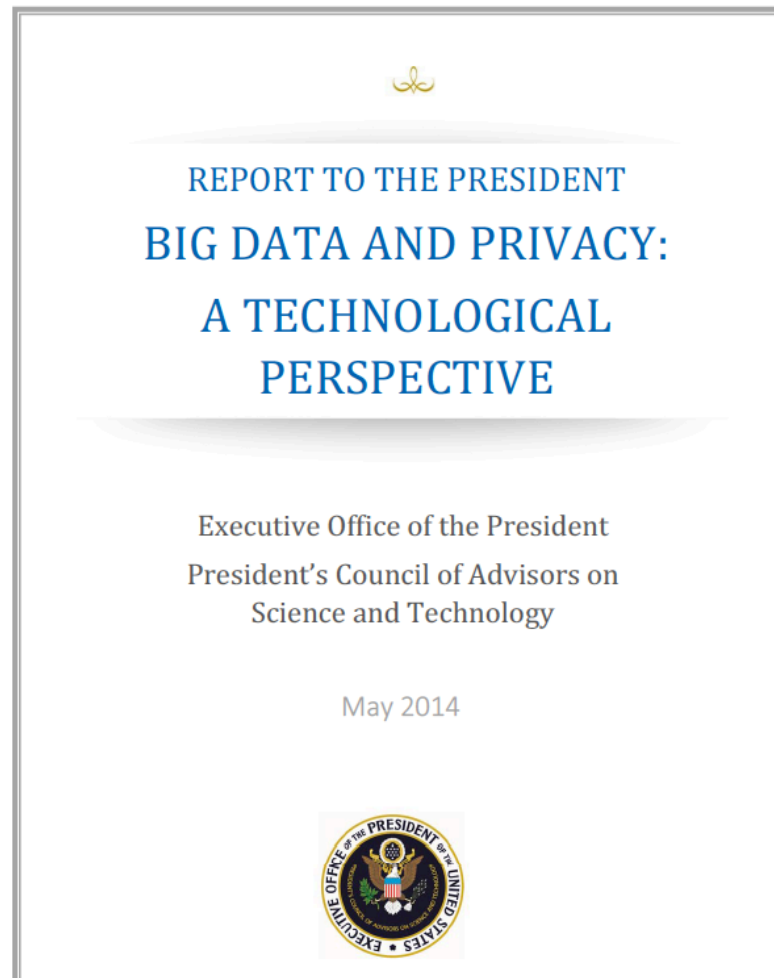
- Scientific American “How Big Data Can Transform Society for the Better’ Oct 13
- Understanding the spread of Malaria in Kenya through mobile phone usage patterns (Wesolowski, *Science* 2012)
- Better public transportation through GPS tracking
- Better public health through search queries

Concerns about big data

- Incremental Effect
- Automated Decision-Making
- Predictive Analysis
- Lack of Access and Exclusion
- Analytics
- Chilling Effect

Solutions to the concerns?

PCAST report on big data



PCAST = President's advisory committee on science and tech.

- PCAST, staffed by OSTP
- Since FDR in 1933, many US Presidents have had a science and technology advisory board
- PCAST named and chartered by George H. W. Bush in 1990
- Topics: Climate, MOOCs, agriculture, nuclear security, big data

Seed Magazine, Recasting PCAST by Robert Koenig
President's Advisory Committee Straddles Worlds Of Politics, Science
Politics, Science (The Scientist, Vol:10, #6, p. 3 & 5, March 18, 1996)

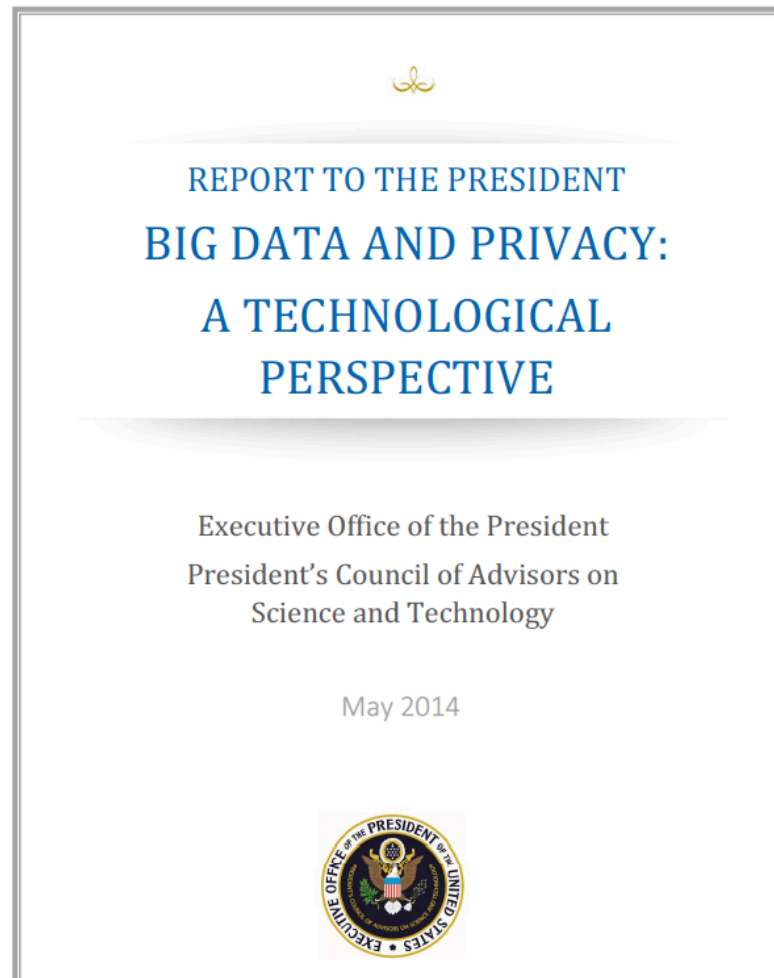
20 current PCAST members

- Three Nobel laureates,
- Four MacArthur Prize fellows,
- Two university presidents
- 4 members of the national scientific, engineering, or medical academies
- Google's chief executive, Eric Schmidt,
- Microsoft's research chief, Craig Mundie.

Decisions about and by PCAST

- Who is a member
- Number of members
- Access to president
- Funding available
- Private or public meetings?
 - The next public PCAST meeting will take place on November 14, 2014.
- Big studies or small reports

PCAST report on big data



What's new about big data

- The quantity and variety of data that are available to be processed.
- The scale of analysis, inferences, and conclusions
- Data fusion: “when data from different sources are brought into contact and new facts emerge”

~6000 mega Data centers



<http://www.Google.com/about/datacenters>

Data: “Born digital” versus “born analog”

- Born digital - data packets, such as intentionally typed text, clicks, GPS location
 - Overcollection of data is a red flag
- Born analog – pictures, video, health info, voice calls
 - Signal to noise = More information than is strictly needed
 - Overcollection of data may be a good design choice

Security Options for Big Data

- Technologies that allow encrypted databases to be queried
 - E.g. Zero-knowledge systems
 - difficult to extend to complex/unstructured systems
- Differential privacy
 - Trade-off between anonymity and accuracy
- Anonymization
 - “PCAST does not see it as being a useful basis for policy”
- Deletion
 - “The only viable assumption ...is that data, once created, are permanent”

Other FIPPs that don't seem viable

- Notice and Consent
- Collection Limitation
 - Except perhaps attention to collection practices
- A right to forget/ expiration

Specific options that might work

- Privacy profiles
 - E.g. Developed by ACLU or Consumer Reports
 - standard, machine-readable interfaces
- Focus on use
 - With metatags

PCAST Policy Recommendations

1. Focus more on use of data than collection and analysis
2. Policy should be on intended outcomes, not technology solutions

PCAST Policy Recommendations cont.

3. Strengthen U.S. research in privacy-related technologies
4. Encourage increased education and training opportunities concerning privacy protection

PCAST Policy Recommendations, cont.

5. US should take the lead through standards and procurement practices