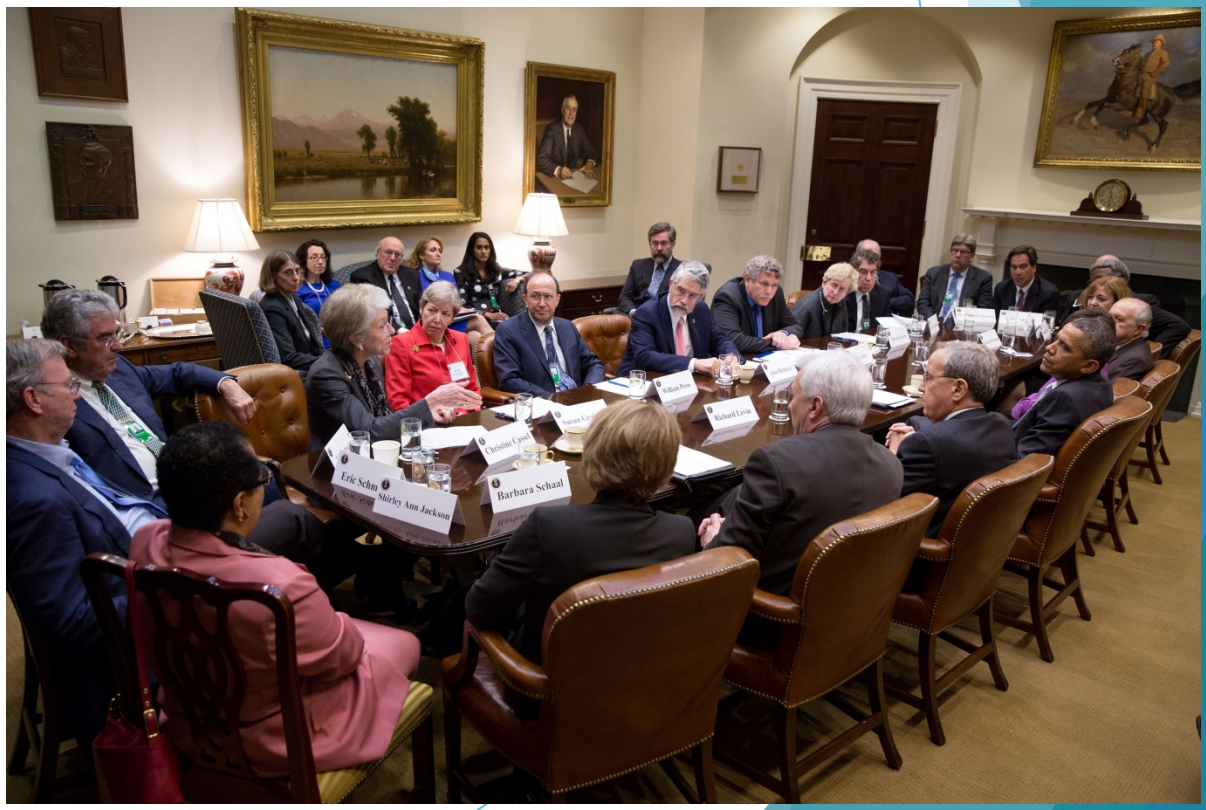
REPORT TO THE PRESIDENT
BIG DATA AND PRIVACY:
A TECHNOLOGICAL PERSPECTIVE

June 2014

Charge from President Obama

- January 17, 2014 speech requesting analysis of big-data implications for policy
  - Scoping study
  - PCAST report to inform and accompany White House report
- Objectives of the PCAST report
  - Assess current technologies for managing and analyzing big data and preserving privacy
  - Consider how such technologies are evolving
  - Explain what technological capabilities and trends imply for design and enforcement of public policy to protect privacy in big-data contexts



Presidents Council of Advisors on Science and Technology (PCAST)

PCAST Working Group Members & Staff

Susan Graham, Co-Chair, UC Berkeley
William Press, Co-Chair, University of Texas
S. James Gates, Jr., University of Maryland
Mark Gorenberg, Zetta Venture Partners
John P. Holdren, OSTP Director
Eric Lander, Broad Institute of Harvard and MIT
Craig Mundie, Microsoft Corp.
Maxine Savitz, National Academy of Engineering
Eric Schmidt, Google, Inc.

Marjory S. Blumenthal, PCAST Executive Director
Michael Johnson, OSTP (NSIA Assistant Director)

Other PCAST Members

Rosina Bierbaum, University of Michigan
Christine Cassel, National Quality Forum
Christopher Chyba, Princeton University
Shirley Ann Jackson, RPI
Chad Mirkin, Northwestern University
Mario Molina, UC San Diego
Ed Penhoet, Alta Partners
Barbara Schaal, Washington University
Daniel Schrag, Harvard University

Structure of Report

- Chapter 1: Introduction (what is new, what is enduring)
- Chapter 2: Examples and Scenarios (illustrating current and potential big data and privacy issues)
- Chapter 3: Collection, Analytics, and Supporting Infrastructure
- Chapter 4: Technologies and Strategies for Privacy Protection and building blocks for future privacy protection policies
- Chapter 5: PCAST Perspectives and Conclusions

What is Privacy?

- "Right to be left alone"
- Ability to share information selectively but not publicly
- Ability to make intimate personal decisions without government interference
- Protection from discrimination on the basis of personal characteristics (e.g., race)
- Intersection with anonymity
- Long history of interaction with technology

- **Invasion of private communication…byproduct of social networking**
- **Public disclosure of inferred private facts…byproduct of analytics**
- **Tracking, stalking…byproduct of locational tracking**
- **False conclusions about individuals…byproduct of group and sometimes personal profiles from big-data analytics**
- **Foreclosure of self-determination…byproduct of long-lived data and analyses**
- **Inhibition of private association…byproduct of concern about potential disclosures**

Changing Technological Contexts

- Privacy history conditioned on “small data”
  - Collection of data/development of data sets used w/conventional statistics
  - Context of a personal relationship (e.g., personal physician, local shop)
- Big data attributes
  - Quantity and variety of data available to be processed (3 Vs)
  - Scale of analysis that can be applied to those data (“analytics”)
  - Expansion of metadata
- Laws have not always kept pace w/technological realities

People Emit Data Continuously . . .

Born digital

- Generated for computer(s)
- Clicks and taps, GPS, cookies

Born analog

- Byproduct of the physical world
- Sensors collect (often invisibly)

Over-collection? Digital convergence?

Big-data analytics create new information

- Data mining and machine learning
- Data fusion and integration (data from different sources)
- Image/speech recognition
- Social-network analysis (self-censorship won't help…)

The Cloud as Dominant Infrastructure

- From commoditized data centers to a complex of software and communications to allow data to be ingested, accessed, and used efficiently
- Replication and distribution
- Infrastructure for mobility (e.g., smart-phone apps)
- Potential security benefits from automation, procedures, oversight
- “Producer-users”—companies that perfect the tech for themselves, may also offer to others

Cybersecurity and Privacy: Distinctions and Dependency

- Cybersecurity: technologies enforce policies for computer use and communication*
  - Systems to protect identity and to authenticate (are you who you say)
  - Systems to protect confidentiality, integrity, availability, non-repudiation; auditability (prove that something occurred)
- Poor cybersecurity is a threat to privacy, but . . .
- Violations of privacy are possible with no failure in computer security
  - Misuse of data, fusion of data
- * More challenging to codify privacy policies than security policies

Technologies and Strategies for Privacy Protection

- Cryptography and encryption
- Anonymization and de-identification
- Data deletion and ephemerality
- Notice and consent

Areas of Concern: Examples

- Healthcare:
  - Personalized medicine (including genetic info); mobile devices that monitor
    - New understanding, better diagnoses, and better treatment (esp. w/massive amounts of data)
    - De-identified data can be re-identified with growth in volume and variety of data
- Education:
  - New online platforms collect masses of data, enable longitudinal datasets
    - New insights are possible into learner abilities and learning styles, more customized support
    - Personal development may be undermined by the persistence of older information
- Home:
  - More ways of collecting, storing, and communicating
    - Fourth Amendment “persons, houses, papers, and effects”
    - Sensors and monitors, home WiFi networks, and Internet connections for more and more objects

What Might the Future Look Like?

- Taylor Rodriguez packs for a trip, leaves suitcase outside home for pick-up
  - Camera on streetlight watches the bag, suitcase has an RFID tag (anti-theft)
- Her suitcase is picked up at night by delivery company
  - Shipper knows Taylor's itinerary and plans
- Self-driving car arrives, its instructions for her itinerary delivered by the cloud
- No boarding passes or queues at the airport
  - Everyone is tracked by phone, facial recognition, gait, emotional state, RFID tags
- In this world, the cloud and robotic aides are trustworthy WRT personal privacy
  - Improvements in convenience and security of everyday life become possible . . .
  - *Not an endorsement, just food for thought!*

PCAST Perspectives and Conclusions

- Government role to prevent breaches of privacy that can harm individuals, groups
  - Tech *plus* law/regulation to generate incentives, contend with measure-countermeasure cycle
- New sources of big data are abundant; new analytics tools will emerge
  - New data aggregation and processing can bring enormous economic and social benefits.
  - Unintentional leaking of data and deliberate systemic attacks on privacy are potential risks
  - Cannot always recognize privacy-sensitive data when collected—may emerge w/analytics, may be able to home in on the moment of particularization to an individual
  - "Dual use" (same technologies usable for benefit or harm)
- Data collectors, data analyzers, and users of analyzed data as different actors
  - Policy can intervene at various stages of this value chain
  - Attention to collecting practices may reduce risk, but use is the most technically feasible place to apply regulation
- Technological feasibility matters

Recommendation 1: Policy attention should focus more on the actual uses of big data and less on its collection and analysis

- Any adverse consequences of big data arise from a program/app interacting with raw data or information refined via analytics
- Policies focused on the regulation of data collection, storage, retention, a priori limitations on applications, and analysis (absent identifiable actual uses of the data or products of analysis) are unlikely to yield effective strategies for improving privacy
- It is not the data themselves that cause the harm, nor the program itself (absent any data), but the confluence of the two

Recommendation 2: Policies and regulation should not embed particular technological solutions, but rather should be stated in terms of intended outcomes

- Technology alone is not sufficient to protect privacy
- To avoid overly lagging the technology, policy concerning privacy protection should address the purpose—the "what" —rather than prescribe the mechanism—the "how"
- Controlling the use of personal data is more effective than regulating technologies of data collection, storage, and retention (these may evolve rapidly)

Recommendation 3: With support from OSTP, the NITRD agencies should strengthen U.S. research in privacy-related technologies and in the relevant areas of social science that inform the successful application of those technologies

- Some of the technology for controlling uses already exists
- Research and research funding are needed for (1) technologies that help to protect privacy, (2) social mechanisms that influence privacy-preserving behavior, and (3) legal options that are robust to changes in technology and create appropriate balance among economic opportunity, national priorities, and privacy protection

Recommendation 4: OSTP, together with the appropriate educational institutions and professional societies, should encourage increased education and training opportunities concerning privacy protection

- Career paths for professionals (e.g., digital-privacy experts both on the software-development side and on the technical-management side)
- Programs that provide education leading to privacy expertise are essential and need encouragement

Recommendation 5: The United States should adopt policies that stimulate the use of practical privacy-protecting technologies that exist today. It can exhibit global leadership both by its convening power and also by its own procurement practices

- Nurture the commercial potential of privacy-enhancing technologies through U.S. government procurement and through the larger policy framework
- Promote the creation and adoption of standards
- Cloud computing offers positive new opportunities for privacy
  - Privacy-Preserving Cloud Services?
- PCAST is not aware of more effective innovation or strategies being developed abroad

White House Big Data Study

Big Data: Seizing Opportunities, Preserving Values

May 2014

Areas of Focus

- Preserving Privacy Values
- Educating Robustly and Responsibly
- Big Data and Discrimination
- Law Enforcement and Security
- Data as a Public Resource

Policy Recommendations

1. Advance the Consumer Privacy Bill of Rights
2. Pass National Data Breach Legislation
3. Extend Privacy Protections to Non-U.S. Persons
4. Ensure Data Collected on Students in School is used for Educational Purposes
5. Expand Technical Expertise to Stop Discrimination
6. Amend the Electronic Communications Privacy Act

Questions?

Marjory S. Blumenthal, Executive Director, PCAST
mblumenthal@ostp.eop.gov