



Does Big Data Change the Privacy Landscape?

Collaboration as Big Data Ethics
September 19, 2016

Sallie Keller

Professor of Statistics & Director



<https://www.bi.vt.edu/sdal>



Biocomplexity Institute of Virginia Tech

- The study of life and environment as a **complex system**
- Understanding biology **in the context of** ecosystems and human-created systems
- **Transdisciplinary** team science

“From molecules to policy”



Problem-Driven Science

Our information biology approach is putting research to work in the real world, breaking down barriers between science and policy.

Social and Decision Analytics Lab

The Social and Decision Analytics Laboratory brings together statisticians and social and behavioral scientists to embrace today's data revolution, developing evidence-based research and quantitative methods to inform policy decision-making.

- **Science of *ALL* Data**
- **Community Analytics**
 - Education and Labor Force Analytics
 - Health Measurement Analytics
 - Industrial Innovation Analytics
- **Information Diffusion Analytics**

Does Big Data Change the Privacy Landscape: A Review of the Issues

Sallie Ann Keller
Stephanie Shipp
Aaron Schroeder



ANNUAL REVIEW OF STATISTICS AND ITS
APPLICATION

VOLUME 3, 2016

Let's begin with the bottom line upfront

- The *all* data revolution is changing the focus of the privacy discussion from the **masking and suppression** of data in order to maintain confidentiality, **to trust, policy,** and **governance**
- This is in itself a **revolution in thinking** about privacy

Privacy, confidentiality, and security

- **Privacy** refers to the amount of personal information individuals allow others to access about themselves
- **Confidentiality** is the process that data producers and researchers follow to keep individuals' data private
- **Security** applies to data storage and transport

Data definition circa 1993

- “*Data* is the term used to describe the various representations that have been developed to characterize the world and the **ways in which communication can occur between its individuals.**”
- “One frequently thinks of data as numbers representing some measurements, but data should also be thought of as the **relationships between those measurements**”



Database systems:
inferential security,
Keller and Unger

Today's *ALL* data revolution

Designed Data



Administrative Data



Opportunity Data



Procedural Data



Social and behavioral data flows

Infrastructure



- Condition
- Operations
- Resilience
- Sustainability

Environment



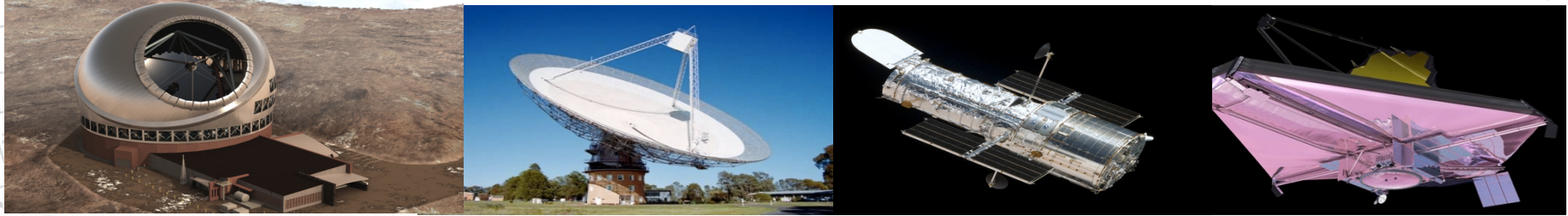
- Climate
- Pollution
- Noise
- Flora/ Fauna

People



- Relationships
- Location
- Economic Condition
- Communication
- Activities
- Health

ALL Data-The new lens for social observing



- Data collected faster, while individuals are in the act of behaving in real life situation
- Adapt methods to make the best use of these data
- New data streams produce new discoveries but should not be allowed to degrade the scientific approach

The privacy journey through U.S. legislation

- Constitution, Article 1, Section 2 (1787)
- Census Act (1790)
- Title 13 U.S. Code (1954, 1962, 1990)
- Freedom of Information Act of 1966
- Privacy Act of 1974.
- Title 26 U.S. Code (1986) of Internal Revenue Code (1939, 1954)
- The Confidential Information Protection and Statistical Efficiency Act of 2002

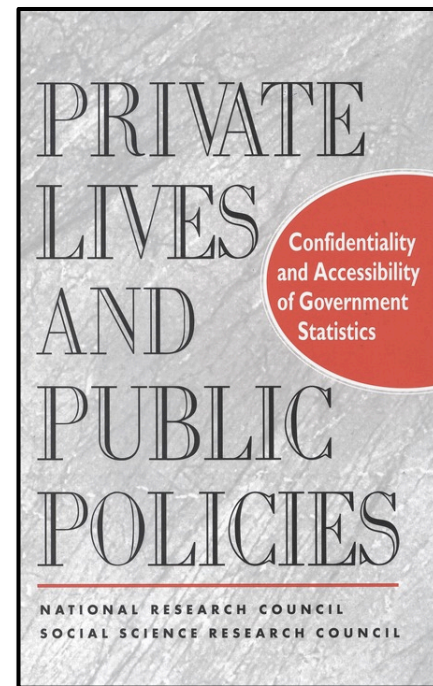
Research community joins privacy journey



- Disclosure-limited data dissemination, 1986
 - Duncan and Lambert
 - *J. Am. Stat. Assoc.*

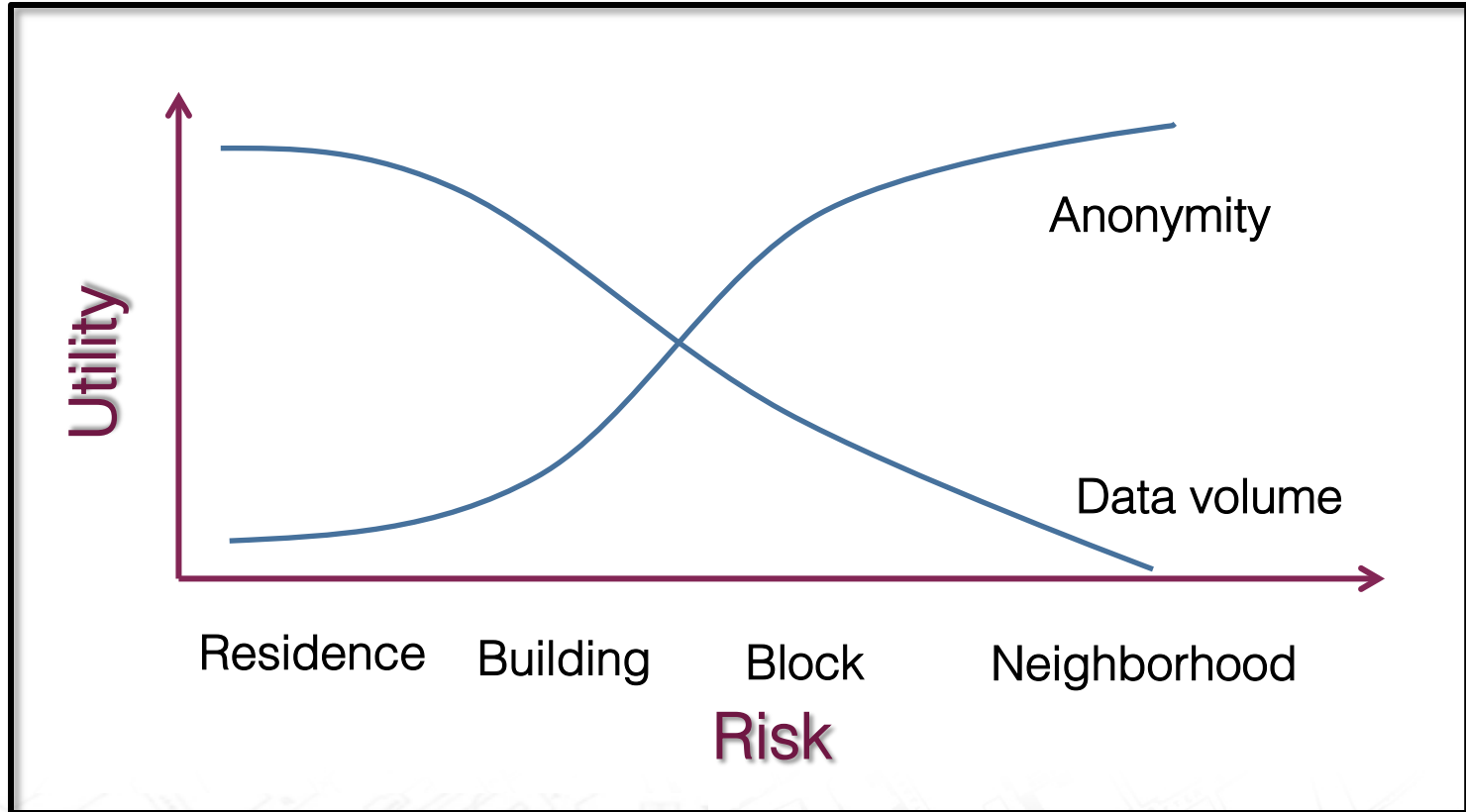


- The risk of disclosure for microdata, 1989
 - Duncan and Lambert
 - *J. Bus. Econ. Stat*

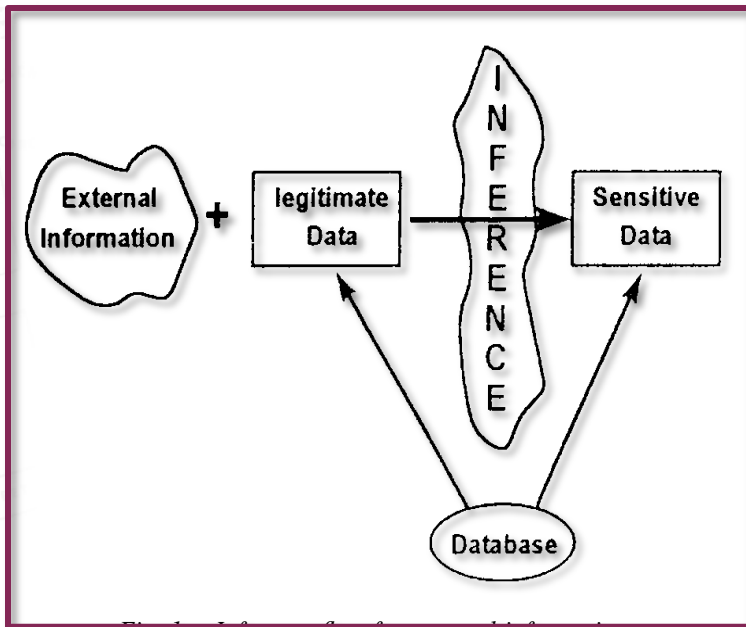


NRC 1993

Foundation based on Data Risk-Utility maps



Introduce sequential query remote access



A database system prototype for remote access to information based on confidential data
Keller and Unger
J of Official Statistics

- Novel idea for “Official Statistics,” 1998
- Research supported by:



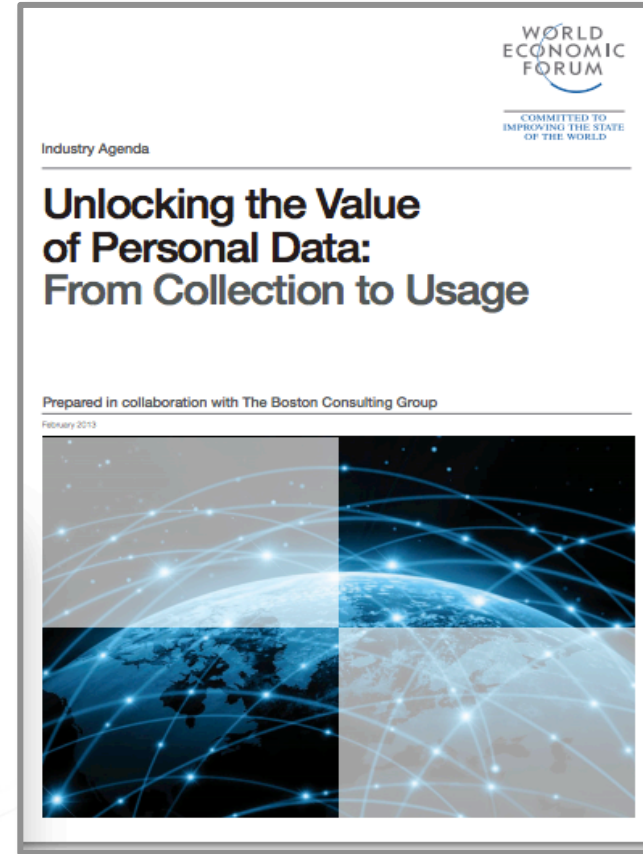
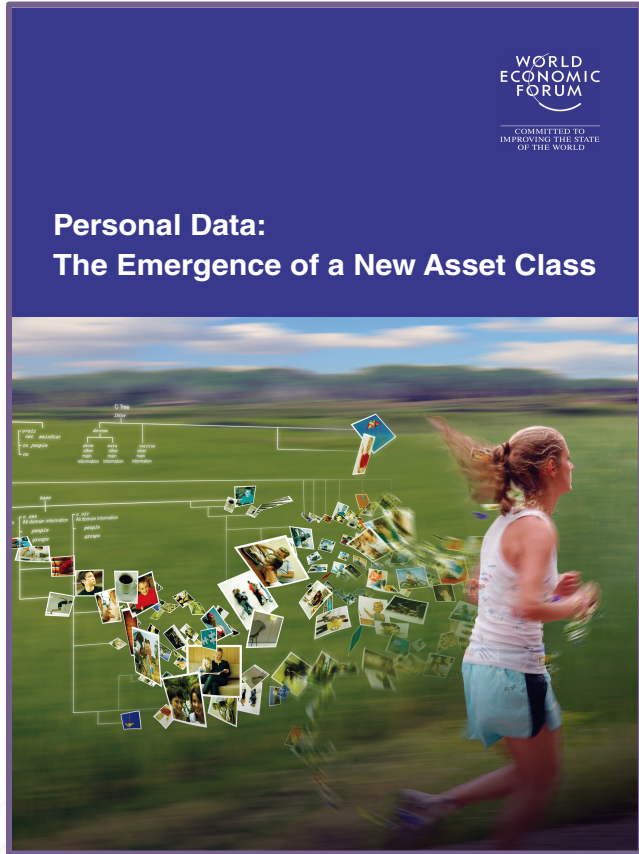
Census Bureau

and



National Security Agency

World Economic Forum 2012 and 2013



World Economic Forum 2013

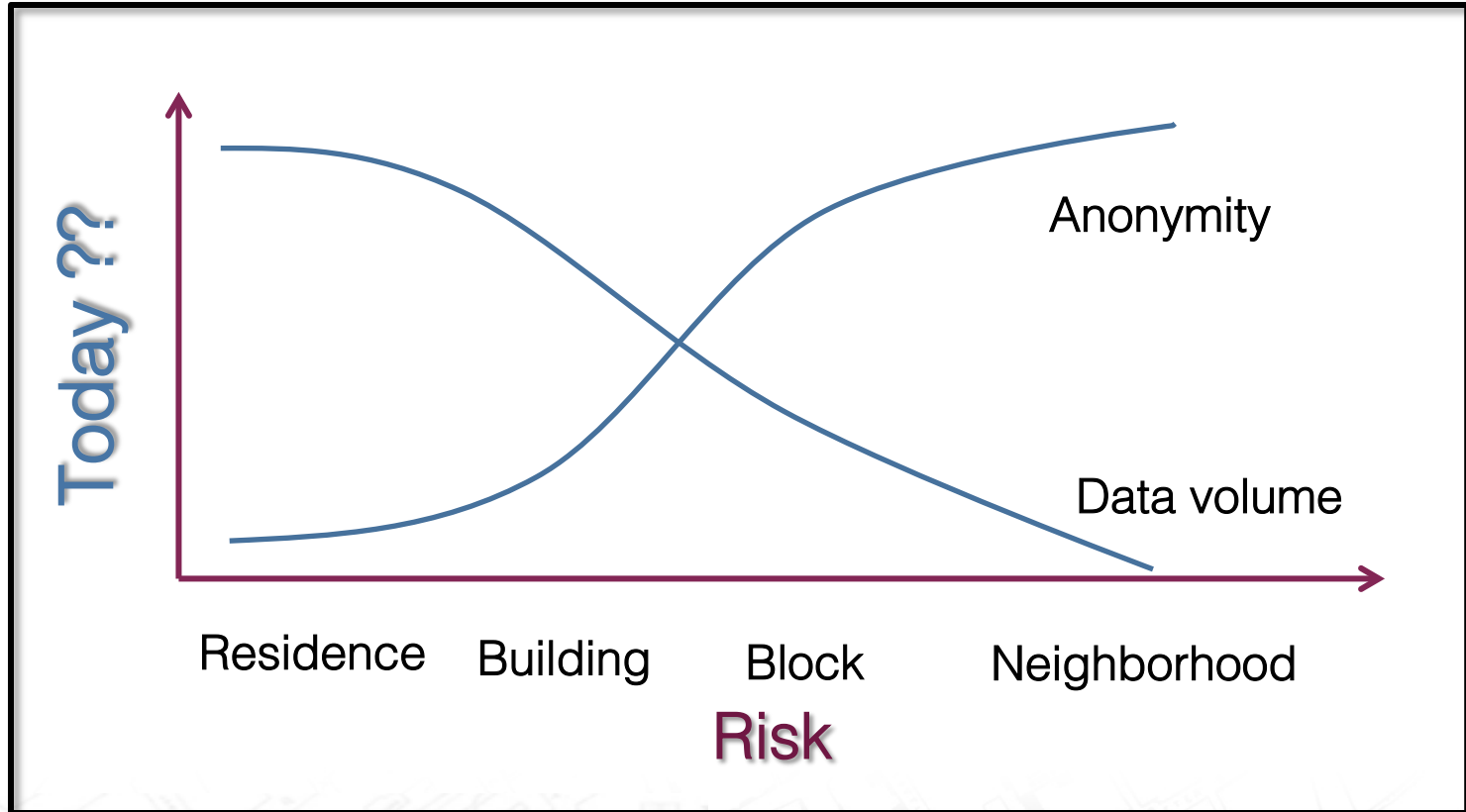
Past

- Definition of personal data is **predetermined** and binary
- Individual provides legal consent but **not truly engaged**
- Policy framework focuses on **minimizing risk to individual**

Future

- Definition of personal data is **contextual** and dependent on social norms
- Individual **engaged** and understands how data is used and value created
- Policy needs to focus on **balancing protection with innovation and economic growth**

Research foundation is being challenged

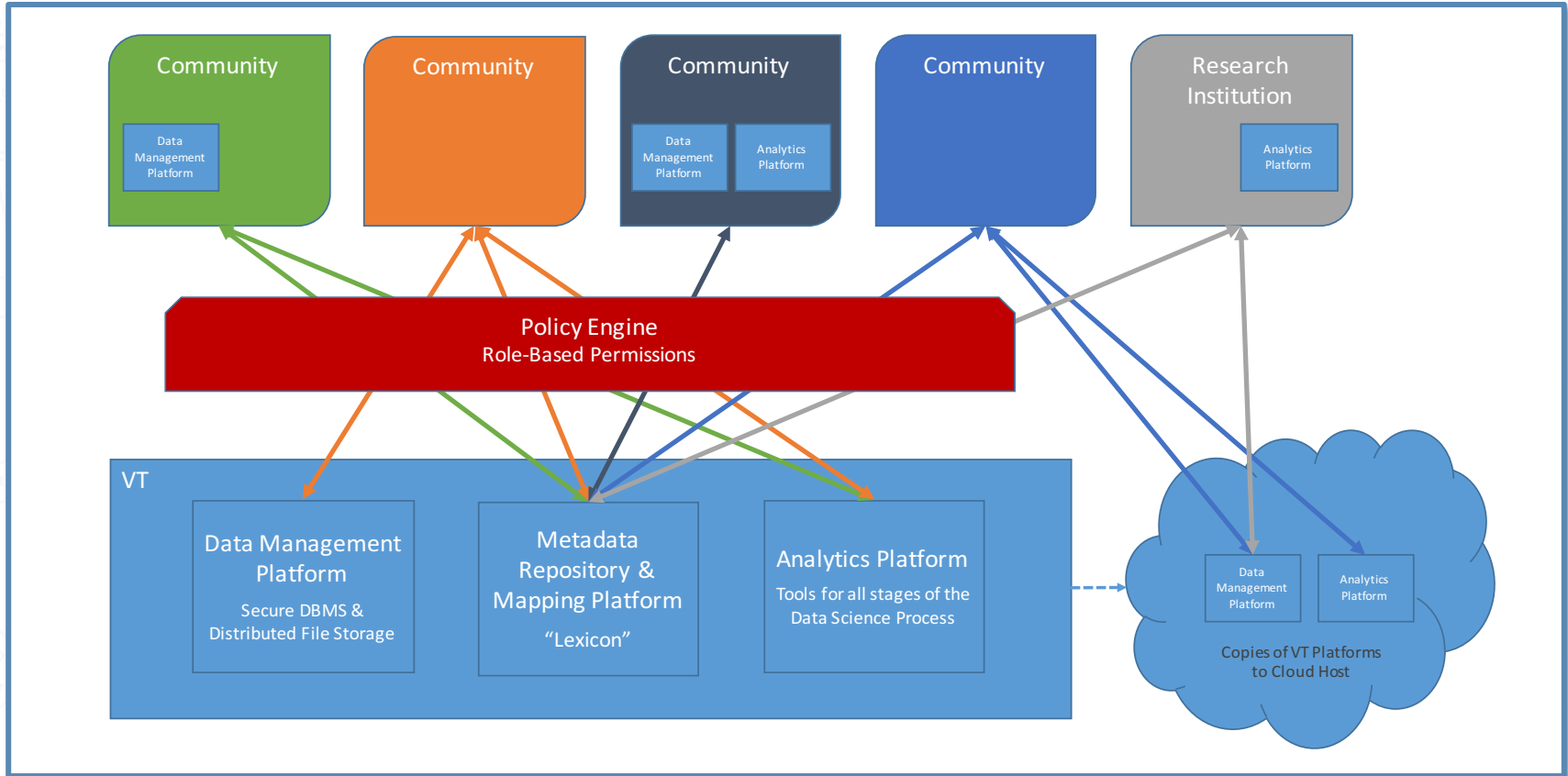


Need a paradigm shift

Indeed, there is hardly any part of one's life that does not emit some sort of "data exhaust" as a byproduct. And it has become virtually impossible for someone to know exactly how much of his data is out there or where it is stored. . . . The time has come for a new approach: shifting the **focus from limiting the collection and retention of data to controlling data at the most important point – the moment when it is used.**

Craig Mundie 2014

Trust & governance preserving infrastructure



Moving to a trust centric approach

- What should government be responsible for?
- How should non-government data influence government data development and reporting?
- What alliances need to be shaped and built
 - Why should the private sector “openly” participate?
- Transparency and information sharing critical
- Data reuse and reproducibility of research

Trust – Technology – Governance

Thanks to our sponsors



Global Forum on Urban and Regional
RESILIENCE



Local / State Government

Arlington County

Virginia Department of Emergency Management

Federal Statistical Agencies

Housing and Urban Development

National Science Foundation

National Center for Science and Engineering Statistics

U.S. Census Bureau

Department of Defense

U.S. Army Research Institute

Defense Manpower Data Center

Defense Advanced Research Project Agency

Industry

MITRE Corporation

Proctor & Gamble

Foundations

Robert Wood Johnson

Laura and John Arnold



laura and john arnold foundation®

