

UBIQUITOUS COLLECTION AND ITS DISCONTENTS

**A Presentation to the
Government-University-Industry Research Roundtable
The National Academies**

**Alvaro Bedoya, Executive Director
Georgetown Center on Privacy & Technology
October 15, 2014**

WHAT ARE WE TALKING ABOUT?

- “Big Data” is a **big**, amorphous term.
- Many descriptive and *normative* claims.
- I want to focus on a *specific* set of normative claims that I think have received less attention.

THE NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
- Algorithmic decision-making is less biased and more accurate than human decision-making.
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.

THE NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
→ **The Raspberry Theory of Big Data**
- Algorithmic decision-making is less biased and more accurate than human decision-making.
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.





3rd Ave & Pine St Mezzanine

3rd Ave & Pine St Mezzanine

6:17 PM

verizon wireless

NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
→ The Raspberry Theory of Big Data
- Algorithmic decision-making is less biased and more accurate than human decision-making.
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.

NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
→ The Raspberry Theory of Big Data
- Algorithmic decision-making is less biased and more accurate than human decision-making.
→ The Data Theory of Big Data
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.



NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
→ **The Raspberry Theory of Big Data**
- Algorithmic decision-making is less biased and more accurate than human decision-making.
→ **The Data Theory of Big Data**
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.

NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
- Algorithmic decision-making is less biased and more accurate than human decision-making.
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.

NORMATIVE CLAIMS OF BIG DATA

- We should rely on correlation, not causation.
- Algorithmic decision-making is less biased and more accurate than human decision-making.
- We should move away from consent-for-collection model & rely on use restrictions.
- We should retain data, not delete it.

ARGUMENT 1: INEVITABILITY

“[A] sea of ubiquitous sensors, each of which has legitimate uses, make the notion of limiting information collection challenging, if not impossible.”

~ White House Big Data Report (2014)

“[S]o much data is being collected, in so many ways, that it is practically impossible to give people a meaningful way... to consent to its collection in the first place.”

~ Craig Mundie (2014)

ARGUMENT 2: DESIRABILITY

“The beneficial uses of near-ubiquitous data collection are large, and they fuel an increasingly important set of economic activities.”

“A policy focus on limiting data collection will not be a broadly applicable or scalable strategy – nor one likely to achieve the right balance between beneficial results and unintended negative consequences (such as inhibiting economic growth).”

~ PCAST Report (2014)

WHAT I AM ARGUING

- Ubiquitous data collection is *not* inevitable.
- Ubiquitous data collection is *not* desirable.

HOW I AM ARGUING IT

- Will focus on location information.
- Why?
 - Sensitive data.
 - *Seems* inevitable – in fact, often used as an example for how ubiquitous data collection is inevitable.

ARGUMENT 1: INEVITABILITY

- Is the ubiquitous collection of location data inevitable?
- It would certainly *seem* so!

WHEN IS LOCATION DATA CREATED?

- Use a mapping app or check-in
 - Phone call → carrier.
 - Use Internet → carrier.
 - Non-location oriented app gets location
 - App company, advertisers

WHEN IS LOCATION DATA CREATED?

- Use a mapping app or check-in
 - Phone call → carrier.
 - Use Internet → carrier.
 - Non-location oriented app gets location
 - App company, advertisers

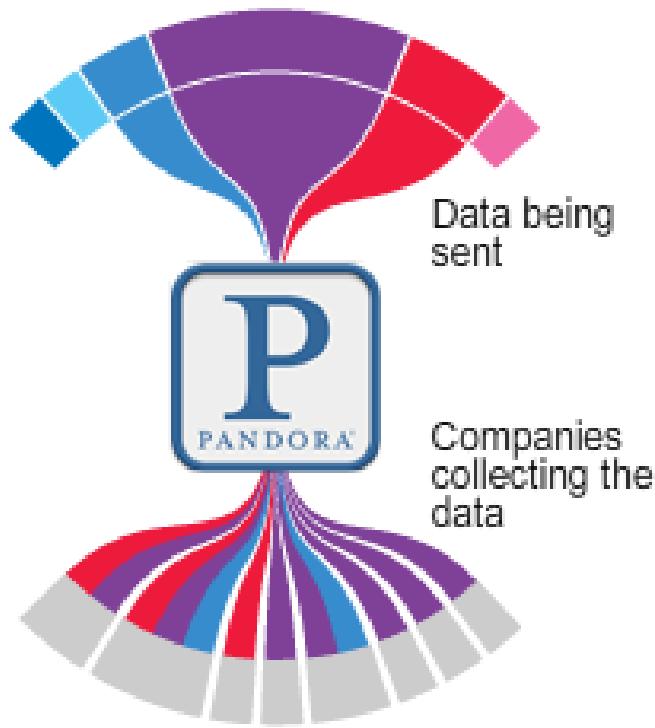
WHEN IS LOCATION DATA CREATED?

- Use a mapping app or check-in
 - App company, advertisers.
- Phone call → carrier.
- Use Internet → carrier.
- Non-location oriented app gets location
 - App company, advertisers
- Phone is “on,” with defaults
 - OS company

WHEN IS LOCATION DATA CREATED?

- Use a mapping app or check-in
 - App company, advertisers
- Phone call → carrier.
- Use Internet → carrier.
- Non-location oriented app gets location
 - App company, advertisers
- Phone is “on,” with defaults
 - OS company
- Walk through a mall or store
 - Tracking firm, retailers

HOW IS LOCATION DATA SHARED?



Categories of data:

User, password

Contacts

Age, gender

Location

Phone ID

Phone number

- *WSJ* 2010 survey: 47 of 101 top apps: Sent user location data to 3rd parties, typically without consent.

HOW DOES THE LAW PROTECT IT?

- It doesn't, really... at least not historically.
- Law enforcement can obtain your location data without a warrant, even after *Jones*.
 - **In re Application of U.S. for Historical Cell Site Data, 5th Cir. (2013)**
- Generally, firms can sell your location to *any other company*, without your consent.
 - **18 U.S.C. § 2702(c)(6).**

ARGUMENT 1: INEVITABILITY

- ...except it's *not*.
 - 1) There's a consensus rule: get consent to collect.

« Apps



Angry Birds

ROVIO MOBILE LTD. ▾

Accept & download

PERMISSIONS

Storage

Modify or delete the contents of your USB storage



Phone calls

Read phone status and identity



Network communication

Full network access



Your location

Approximate location (network-based)



ARGUMENT 1: INEVITABILITY

- ...except it's *not*.

- 1) There's a consensus rule: get consent to collect.

ARGUMENT 1: INEVITABILITY

- ...except it's *not*.
 - 1) There's a consensus rule: get consent to collect.
 - 2) When companies *break* that rule, *other* companies have *innovated* to protect it.

Apple Moves To Stop Location-Tracking By Mobile Analytics Companies

by Wendy Davis, June 9, 2014, 5:11 PM

[Comment \(6\)](#) [Recommend \(13\)](#)[News](#) | [US](#) | [World](#) | [Sports](#) | [Comment](#) | [Opinion](#)[News](#) > [Technology](#) > [Apple](#)

Apple's iOS 8 will stop re-tracking customers via Wi-Fi

New software will randomise the unique identifiers on phones, protecting users against spying

**TRENDING NOW**

Colin Kaepernick tapes over Beats logo to appease #brands

23
NEW ARTICLES[LOG IN](#) | [SIGN UP](#) [LONGFORM](#) [VIDEO](#) [REVIEWS](#) [TECH](#) [SCIENCE](#) [CULTURE](#) [DESIGN](#) [BUSINESS](#) [US & WORLD](#) [FORUMS](#)**PREVIOUS STORY**[Next-generation condoms might feel like human skin](#)**NEXT STORY**[Surprising new study reveals dangers of anesthesia in kids](#)

APPLE

TECH

iOS 8 strikes an unexpected blow against location tracking

By [Russell Brandom](#) on June 9, 2014 09:25 am [Email](#) [@russellbrandom](#)291
COMMENTS**QUARTZ****BIG MAC**

A tiny technical change in iOS 8 could stop marketers spying on you

Alex Hern[Follow @alexhern](#)[Follow @guardiantech](#)[theguardian.com](#), Tuesday 10 June 2014 10.30 EDT

Schneier on Security

Blog**Newsletter****Books****Events**[← Security Risks from Remote-Controlled Devices](#)

iOS 8 is Randomizing Mac OS X's Unique Identifier

[This](#) seems like a good idea.

Tags: [Apple](#), [iOS](#), [marketing](#), [tracking](#)

Posted on June 10, 2014 at 4:40 PM • 43 Comments

[Like](#)[Tweet](#)

ARGUMENT 1: INEVITABILITY

- ...except it's *not*.
 - 1) There's a **consensus rule: get consent to collect.**
 - 2) When companies *break* that rule, *other* companies have *innovated* to protect it.

ARGUMENT 1: INEVITABILITY

- ...except it's *not*.
 - 1) There's a **consensus rule**: get **consent** to collect.
 - 2) When companies *break* that rule, *other* companies have *innovated* to protect it.
 - 3) Developing rule: Consent to *share*.

ARGUMENT 1: INEVITABILITY

- ...except it's *not*.
 - 1) There's a **consensus rule**: get **consent** to collect.
 - 2) When companies *break* that rule, *other* companies have *innovated* to protect it.
 - 3) Developing rule: Consent to *share*.
 - 4) The law is developing around this area.

FIRMS COMPETE ON LOCATION PRIVACY

Apple iOS

Collects location data in background by default

Google Android

Collects location data in background by default

FIRMS COMPETE ON LOCATION PRIVACY

Apple iOS

Collects location data in background by default

Google Android

Collects location data in background by default

Blocks certain WiFi tracking by default

FIRMS COMPETE ON LOCATION PRIVACY

Apple iOS

Collects location data in background by default

Blocks certain WiFi tracking by default

Bans developers from sharing location data without consent

Google Android

Collects location data in background by default

ARGUMENT 1: INEVITABILITY

- The ubiquitous collection of location data is *not* inevitable.
- Taking action that assumes inevitability...
 - Undercuts industry consensus on collection and growing action on sharing.
 - Undercuts legal progress.
 - Stymies innovation and competition on privacy.

ARGUMENT 2: DESIRABILITY → BENEFITS

“The beneficial uses of near-ubiquitous data collection are large, and they fuel an increasingly important set of economic activities.”

“A policy focus on limiting data collection will not be a broadly applicable or scalable strategy – nor one likely to achieve the right balance between beneficial results and unintended negative consequences (such as inhibiting economic growth).”

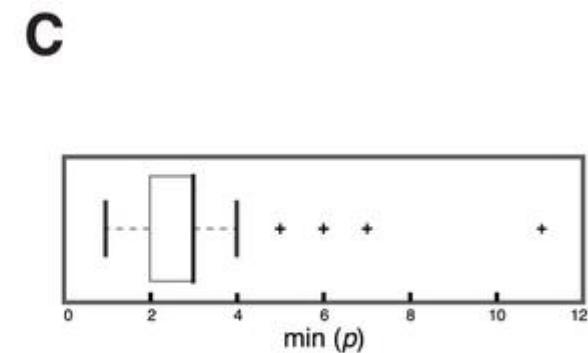
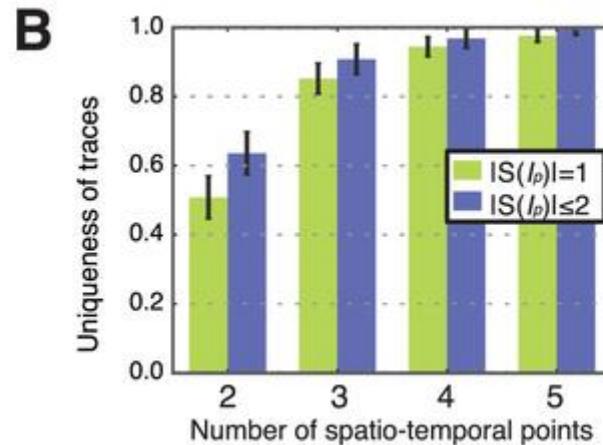
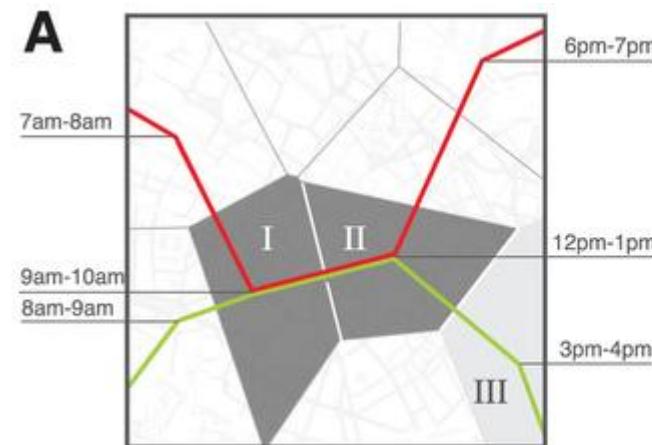
~ PCAST Report (2014)

VALUABLE USES OF LOCATION DATA

- Give you directions.
- Improve maps.
- Track traffic.
- Diagnose you.
- Identify you.

LOCATION SOCIAL SCIENCE RESEARCH

- Individuals move more erratically before falling ill. (S. Pentland)
- A record of someone's location data is unique to that person (de Montjoye et al.)
 - Location information can function as a quasi-biometric.



VALUABLE USES OF LOCATION DATA

- Give you directions. →Help me.
- Improve maps. →Help me & others like me.
- Track traffic.
- Diagnose you. →Conduct health & science research.
- Identify you.
- Consumer profiling. →Help others grow revenue.
- Targeting ads.
- Why should the privacy protections for *my* data turn on the financial benefit of *my* information for *others*?

NOT ALL BENEFITS OF DATA ARE EQUAL

“[W]e should distinguish between research that benefits the public and that which serves only narrow and private gain.”

“[W]e should separate benefits built upon data sets that are full of information about people from those built upon data that has almost nothing to do with personal information”

~ Paul Ohm (2013)

ARGUMENT 2: DESIRABILITY → HARMS

“The beneficial uses of near-ubiquitous data collection are large, and they fuel an increasingly important set of economic activities.”

“A policy focus on limiting data collection will not be a broadly applicable or scalable strategy – nor one likely to achieve the right balance between beneficial results and unintended negative consequences (such as inhibiting economic growth).”

~ PCAST Report (2014)

ARGUMENT 2: DESIRABILITY → HARMS

“The beneficial uses of near-ubiquitous data collection are large, and they fuel an increasingly important set of **economic activities**.”

“A policy focus on limiting data collection will not be a broadly applicable or scalable strategy – nor one likely to achieve the right balance between beneficial results and **unintended negative consequences (such as inhibiting economic growth)**.”

~ PCAST Report (2014)



(A) Evenings 9pm-7am:
→ Where you live.

(B) Weekdays 7:30am:
→ Your child's school.

(C) Weekdays 9am-6pm
→ Where you work.

(D) Thurs.-Sat. nights:
→ Sexual orientation.

(E) Weekly visits:
→ HIV status.

(F) Sundays, 11am:
→ Religion, race.

USE LIMITS & VULNERABLE COMMUNITIES

- A position advocating a de-emphasis on collection controls and an increased reliance on use restrictions assumes that we – our industries and our government – can adequately prohibit, *ex ante*, uses of data that are harmful to vulnerable communities.
- Yet our society is especially slow to condemn – or even acknowledge – uses of data that hurt marginalized communities.

WESTERN DEFENSE COMMAND AND FOURTH ARMY
WARTIME CIVIL CONTROL ADMINISTRATION
Presidio of San Francisco, California
April 1, 1942

INSTRUCTIONS
TO ALL PERSONS OF
JAPANESE
ANCESTRY

Living in the Following Area:

All that portion of the City and County of San Francisco, State of California, lying generally west of the north-south line established by Junipero Serra Boulevard, Worcester Avenue, and Nineteenth Avenue, and lying generally north of the east-west line established by California Street, to the intersection of Market Street, and thence on Market Street to San Francisco Bay.

All Japanese persons, both alien and non-alien, will be evacuated from the above designated area by 12:00 o'clock noon Tuesday, April 7, 1942.

No Japanese person will be permitted to enter or leave the above described area after 2:00 a.m., Thursday, April 2, 1942, without obtaining special permission from the Provost Marshal at the Civil Control Station located at

1701 Van Ness Avenue
San Francisco, California

The Civil Control Station is equipped to assist the Japanese population affected by this evacuation in the following ways:

1. Give advice and instructions on the evacuation.
2. Provide services with respect to the management, leasing, sale, storage or other disposition of most kinds of property including: real estate, business and professional equipment, buildings, household goods, boats, automobiles, livestock, etc.
3. Provide temporary residence elsewhere for all Japanese in family groups.
4. Transport persons and a limited amount of clothing and equipment to their new residence, as specified below.

The Following Instructions Must Be Observed:

1. A responsible member of each family, preferably the head of the family, or the person in whose name most of the property is held, and each individual living alone, will report to the Civil Control Station to receive further instructions. This must be done between 8:00 a.m. and 5:00 p.m., Thursday, April 2, 1942, or between 8:00 a.m. and 5:00 p.m., Friday, April 3, 1942.

NOT

Headqua
Western Defens
and Fourt

Presidio of San Fran
April 1, 1942

Civilian Exclusion

1. Pursuant to the provisions of Public Proclamation No. 1, dated March 26, 1942, respectively, it is hereby ordered that all persons of Japanese ancestry, be excluded from and after 12 o'clock noon, P. S. T., of Tuesday, April 7, 1942, from the area described as follows:

All that portion of the City and County of San Francisco, State of California, lying generally west of the north-south line established by Junipero Serra Boulevard, Worcester Avenue, and Nineteenth Avenue, and lying generally north of the east-west line established by California Street, to the intersection of Market Street, and thence on Market Street to San Francisco Bay.

2. A responsible member of each family, and each individual

COLLECTION CONTROLS ARE CRITICAL

“Far too often, today’s invidious discrimination was yesterday’s national security or public health measure.”

“The American public may never make up its mind about gay people, immigrants, minorities, and the poor – or how they and their data should be treated. Individual controls on data collection take that choice out of the hands of companies and the government, and into the hands of the individual.”

~ Bedoya & Vladreck (2014)

“The runaway slave came to my house and stopped outside, [...] Through the swung half-door of the kitchen I saw him limpsey and weak, And went where he sat on a log, and led him in and assured him, And brought water and filled a tub for his sweated body and bruised feet, [...] He staid with me a week before he was recuperated and passed north, I had him sit next me at table my firelock leaned in the corner.”

~ Walt Whitman, *Song of Myself* (1855)

COLLECTION CONTROLS ARE CRITICAL

- The survival of some vulnerable communities – communities we now venerate and cherish – has turned on their ability to avoid detection.
- In a world of ubiquitous collection, what would have happened to:
 - Civil rights activists?
 - Runaway slaves on the Underground Railroad?
 - American Revolutionaries (dissidents)

CONCLUSIONS

- **Ubiquitous collection is neither inevitable nor desirable.**
- **Accepting it as inevitable would undercut critical innovation – and progress – on privacy.**
- **Eliminating (or de-emphasizing) user controls also threatens to disproportionately harm vulnerable communities who have not been adequately protected by use restrictions.**

RECOMMENDATIONS

- Be skeptical of claims that consent is impossible.
- Examine the sources of your data.
- Distinguish between benefits of data analysis that accrue to data subjects – and those that do not.
- Seek out and evaluate non-economic costs to collection and retention.

REFERENCES

Solon Barocas, Andrew Selbst, *Big Data's Disparate Impact* (Aug. 2014)

Alvaro Bedoya and David Vladeck, *Comments on Big Data and Consumer Privacy in the Internet Economy*, *National Telecommunications & Information Administration* (Aug. 5, 2014)

Danielle Keats Citron, Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASHINGTON LAW REVIEW 1 (2014)

Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (May 2014)

Executive Office of the President, *President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective* (May 2014)

Victor Mayer-Schonberger, Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013)

Yves-Alexandre de Montjoye, Cesar A. Hidalgo, et al., *Unique in the Crowd: The privacy bounds of human mobility*, *Nature: Scientific Reports* (March 2013)

REFERENCES

Craig Mundie, *Privacy Pragmatism: Focus on Data Use, Not Data Collection*, Foreign Affairs (March/April 2014)

National Consumer Law Center, *Big Data: A Big Disappointment for Scoring Consumer Credit Risk* (March 2014)

Paul Ohm, Response, *The Underwhelming Benefits of Big Data*, 161 UNIV. OF PENNSYLVANIA LAW REVIEW ONLINE 339 (2013)

Robinson + Yu, *Civil Rights, Big Data, and Our Algorithmic Future* (Sept. 2014)

Scott Thurm, Yukari Iwatani Kane, *Your Apps Are Watching You*, Wall Street Journal (Dec. 18, 2010)

Claire Suddeth, *The Weather Channel's Secret: Less Weather, More Clickbait*, Bloomberg Businessweek (Oct. 9, 2014) (raspberries in Miami)

World Economic Forum, *Unlocking the Value of Personal Data: From Collection to Usage* (Feb. 2013)