Big Data / Privacy: Pick One?

A. Michael Froomkin
University of Miami School of Law
froomkin@law.miami.edu

Privacy, Quickly

- Has multiple elements including control of access to body, to thoughts, protection of private sphere, and especially intimate relations
- Today will concentrate on one aspect: ability to control data about oneself
  - Primarily an instrumental value
  - Enhances human freedom, flourishing, experimentation, innovation, self-realization; protects against discrimination, and also distant, often invisible, exercises of (often algorithmic) power

Data == Power

- People with data about you can exercise power over you
  - Market power
  - Government power
- Special case of both above: Sorting power
  - Can be invisible
  - Can be lifelong
  - Can be very empowering – or very damaging
    - Credit scores
    - Citizen safety scores
    - Issues of accurate, inaccurate, and predictive (i.e. speculative) scoring

Big Data, Quickly

- US government definition: the growing technological ability to capture, aggregate, and process data
- EU: “the massive and rapid processing of data (through modern data analytics) in the search for information (including unforeseen information) The practice of data mining poses a significant challenge due to the degree of opacity characterising many contemporary data processing activities. ...
  - Data mining practices may result in “behavioural targeting” and further encourage a “datafication” of society that poses significant challenges for privacy and digital rights in general. Due to such risks as statistical discrimination...
- Bottom line: Big Data is *both a technology and a process* with many (reinforcing) parts, each of which can be enabled or, in theory, regulated. It is also, arguably, an ideology.

Big Data as a Process

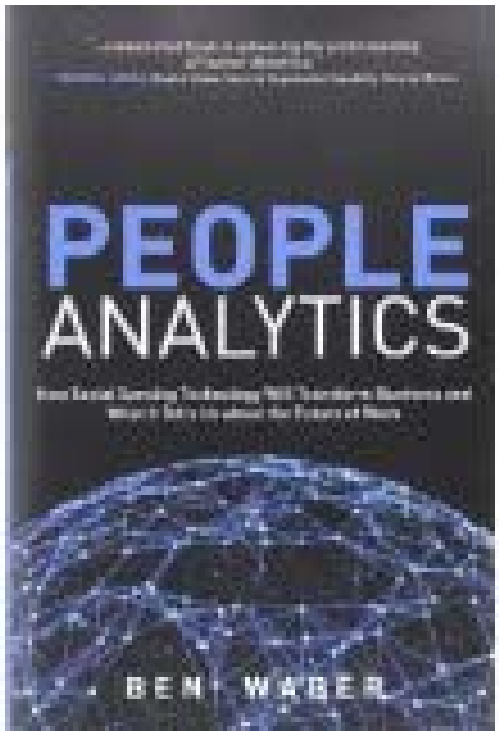
- Data collection
  - Sensors: Smart city, Cell Phones
  - Self-surveillance
  - Transactional data: Communications, Shopping, Medical
- Storage
- Processing
  - Correlation / Data Mining
    - Issues of identification/ de-identification/ re-identification
- Re-Use
  - Nature of Big Data projects is to seek unexpected connections

The Trend is to Ever-More (Big) Data

- More apps creating occasions to create and collect data
- More monitoring
  - So called ‘self-surveillance’ e.g. Fitbits, Nests
- Ubiquitous hardware/sensors
- Bigger aggregations of data
  - Smart cities
  - Fusion centers, DHS, NSA
- Cost/benefit motivates private sector: costs are dropping, perceived benefits growing

The Lure of Unexpected Insights

"What if I told you that ...one of the biggest decisions a company makes revolves around the size of its lunch tables?"



Are You Trusted? Or 'Dangerous'?

"Beware," a program made by West Safety Services is capable of quickly sorting through billions of publicly available commercial records to alert first responders to potentially dangerous situations, according to the product website. Beware calculates "threat scores" by assigning people and addresses green, yellow, or red scores with red being highest threat and green being the safest.

Mixed Blessings of Big Data

Benefits – Numerous, Usually Concentrated

- Medical, Epidemiology
- Security
- Finance
- Marketing
- Urban planning
- “Quantified Self”

Costs – Widespread, Diffuse, Harder to Quantify

- Privacy reduced or eliminated
  - Increases risk of reprisal for acts
    - Less experimentation
  - Dossiers are bigger
- Increased opacity in decision-making
- Polity
  - Self-censorship
    - Unpredictable consequences of speech, civic choices
  - Less whistleblowing

More & ‘Better’ Algorithms

As they are fed more data

- Can infer results about you based on people who share observed traits about you, another way in which there is no occasion for consent on your part, if other have consented to deeper dive into their data
- Thus, big data is a power shift - away from data subject towards those who control and can access the data.

Common Privacy Protection Techniques Fail

- Consent to data collection loses what little meaning it has
  - "Consent" is meaningless for remote sensing
  - In US 'consent' not required for most transactional collection unless "sensitive"
  - "Informed consent" is *impossible if one cannot predict what effect the data use might have*
- Processing limits are inadequate
  - US law bans 'discrimination' against certain classes e.g. race, religion
  - But no law against discriminating against people who pay for pizza with credit
  - Battle over no-fly lists is going very slowly
- Data anonymization easily undermined
  - Re-identification is easy and likely profitable

Metaproblem: Attitudes

- As Julie Cohen says, Big Data repackages surveillance as innovation. Plus, “Big Data ... equates information with truth and more information with more truth, and that denies the possibility that information processing designed simply to identify a ‘pattern’ might be systematically infused with a particular ideology.”
- “[P]rivacy is increasingly cast as the spoiler in this tale, the obstacle to the triumphant march of predictive rationalism. ... [I]f information processing is rational, then anything that disrupts information processing, including privacy protection, is presumptively irrational.”

What is to be Done?

- “Big Data cannot be entirely defanged and its users cannot be entirely trusted” – James Grimmelman
- If Cohen is correct that Big Data is the result of an ideology then we should be wary of the argument that the value of insights drawn from huge datasets justify the creation of centralized repositories and their use.
- Privatizing, distributing, or ‘federalizing’ the data makes little difference.
- The specter of David Brin (*The Transparent Society*) haunts privacy studies?

Try Everything – And Do It Soon

- Regulate sensors/data collection (my favorite)
- Regulate storage or flows (sharing)
  - Age out data
  - RTBF
- Regulate use/re-use once holding data (but, 1st Amendment)
- Focus on discrimination: ban 'bad' algorithms (but enforcement hard, and unconscious biases in algorithmic design or in data set construction)
- Demand a new due process in data processing
  - Right to contest distant, invisible algorithmic decision making
  - Get notice of when, how and why your data is being processed
- Invent new ways to opt-in and opt-out
  - (But opt-out options need to be realistic not formalistic)

Thank You