Big Data and the Government:

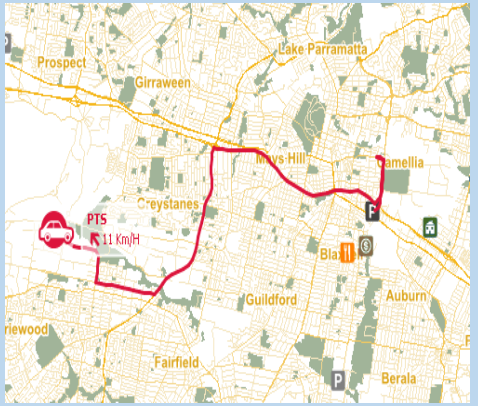
Implications for Individuals and Businesses

Christopher Slobogin
Vanderbilt University Law School

Mumbai, India
December 14, 2018

Government Access to Data about Individuals

- What type of justification is needed? What type of procedure is needed?
- Five regulatory scenarios
- American law
- Indian law???



ChoicePoint
GOVERNMENT SERVICES

The Fourth Amendment

- Prohibits unreasonable “searches” and “seizures”
- Generally, non-exigent searches must be authorized by a judicially-issued warrant based on probable cause
- Probable cause = ~50% probability evidence will be discovered
- Some types of searches and seizures (e.g., stop and frisk) can take place based on reasonable suspicion, which = ~30% probability
- And some can take place upon a mere showing of relevance/that search will “satisfy official curiosity” (e.g., with a subpoena)

Supreme Court Jurisprudence Until Recently: Two Doctrines defining "search"

- *Katz v. United States* (1967): A search for 4th A. purposes = police infringement of "expectations of privacy society is prepared to recognize as reasonable" (REP). Cf. Indian law
- Knowing exposure doctrine: no REP when traveling public roads (cf. GPS tracking, ALPR and CCTV)
- General public use doctrine: no REP vis a vis technology used by general public (cf. Startron)
- Contraband-specific doctrine: no REP vis a vis technology that detects only items with no privacy significance (cf. Raytheon device)
- Assumption of risk/third party doctrine: no REP vis a vis information surrendered to third party (cf. bank, phone and Choicepoint records)

Some Cracks in the Jurisprudence

- *Jones v. United States* (2012)
  - Planting GPS device on car is a search because it involves a "trespass"
  - 5 justices: *any* GPS tracking is a search, at least if "prolonged"
- *Carpenter v. United States* (2018)
  - Accessing 7 days of cell-site location data requires a warrant
  - Less than 7 days?
  - Majority: we're not addressing access to phone, bank or ISP records
- Definition of search has expanded—but is a warrant always required?
- Five types of data searches:
  - Investigating an identified suspect (suspect-driven)
  - Profiling potential suspects using algorithms, artificial intelligence (profile-driven)
  - Accessing data re time and place of crime (event-driven)
  - Data collection (program-driven)
  - Data holder disclosure (volunteer-driven)

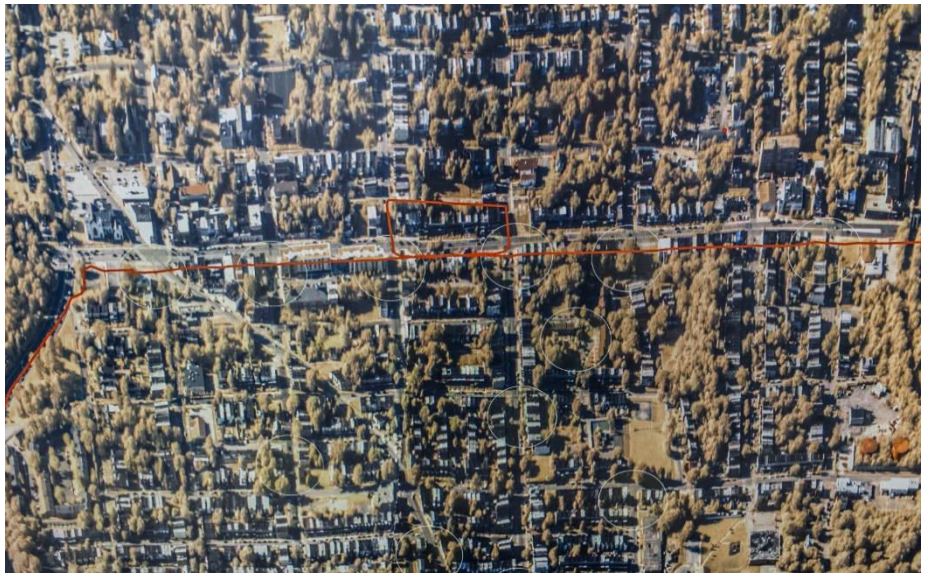
Suspect-Driven--Proportionality

- Police have a suspect and want more information (is a warrant required?)
- Why a warrant should not always be required for suspects
  - Not all searches are equally intrusive
  - A probable cause requirement might irrationally handcuff the police
- The proportionality principle: the level of justification should be roughly proportionate to the level of intrusion (cf. *Jones* and *Carpenter*)
- But how to implement it?
  - Justification hierarchy:
    - Warrants based on probable cause
    - Judicially-issued subpoena (based on relevance or reasonable suspicion)
    - Written authorization by executive superior based on relevance
  - Intrusiveness hierarchy:
    - Focus of information? (e.g., re an individual rather than a business)
    - Nature of information? (medical records v. utility records; content v. metadata)
    - Amount of information? (28 days worth v. < 7 days worth)

Profile-Driven: Hit Rates

- May police try to identify suspects using a profile?
  - Drug courier/credit card fraud/terrorists?
  - “Strategic Subject list”
  - Beware/A.I.?
- Five considerations. Profiles should:
  - Be transparent and interpretable
  - Not use racially-biased data
  - Demonstrate a “hit rate” (rate at which the profile identifies current or imminent criminals) proportionate to the data intrusion
  - Only authorize police action proportionate to the hit rate
  - Be applied in a universal or pre-specified neutral manner

Event-driven: Hassle Rates



- A technology-driven search based on an event (e.g., Oklahoma case/Baltimore/"data dump")
- Proportionality reasoning suggests that, in the typical case, little regulation is needed
- However, to avoid high "hassle rates" of innocent people, police should seek judicial authorization before wide-ranging event-driven cases. Judges should consider:
  - "the gravity of the public concerns served by the seizure [search],
  - "the degree to which the seizure [search] advances the public interest,
  - "the severity of the interference with individual liberty."

  See *Illinois v. Lidster*, 540 U.S. 419, 427 (2004)

Program-Driven: Democratic Authorization

- Mass data collection programs
  - Total Information Awareness (TIA)
  - Prism, Stellar Wind, NSA metadata program
  - Fusion centers

National Fusion Center Association map of fusion centers nationwide. Does not include all fusion centers.

- Should such data *collection* programs be permitted?: *If* maintained outside government when possible, and *if* subject to. . .
- Accountability mechanisms (cf. Administrative Procedure Act)
  - Authorizing legislation (identifying harms, means and targets)
  - Administrative regulations subject to notice & comment
  - Judicial "hard look" review to ensure rules are rational and that
  - Even application of regulations (through universal or pre-specified criteria)
  - Auditing, security, and notice provisions

Summary

- **Suspect-driven**: If a policing agency seeks non-public data about an identified person it should have to demonstrate suspicion of wrongdoing proportionate to the intrusion involved, as measured by the type and amount of data obtained.
- **Profile-driven**: If a law enforcement agency is accessing data for the purpose of executing a profile to identify suspects, it should ensure the profile's hit rate is proportionate to the data intrusion and to the police action contemplated, that it uses an understandable algorithm that is not illegitimately discriminatory, and that it is used in connection with a specific crime or to investigate the entire profiled group.
- **Event-driven**: If policing agencies are relying on a crime rather than a suspect or a profile as the starting point of the investigation, they should keep the number of people investigated to the minimum dictated by the time, place and gravity of the crime.
- **Program-driven**: Collections of data needed by law enforcement should be maintained outside of government to the extent consistent with governing needs, but wherever maintained should be authorized by specific legislation and administrative rules transparently and democratically arrived at.

Relationship to Indian Law

- Background--*Kharak Singh* (1964); *Gobind* (1975); *Canara Bank* (2005) ; *Puttaswamy* (2017/2018). There is no explicit constitutional right to privacy, but
  - Articles 19 & 21 guarantee the rights to freedom and liberty that may only be infringed to achieve a "compelling state interest" in a "narrowly tailored" way
  - Further, because the Indian Constitution "protects people, not places;" the third party doctrine apparently does not apply in India
- Suspect-driven--*PUCL* (1997). Individualized suspicion is required, but:
  - It is defined as a "reasonable basis [to believe] documents will lead to the discovery of [crime]" (*Canara Bank*)
  - Judicial review might not be required where "public safety" is concerned (*Gobind*, Personal Data Protection law). Compare exigency requirement in American law.
- Profile-driven and event-driven--*Malak Singh* (1981)
  - Approves surveillance of "persons with previous criminal record . . . or [who] are reasonably believed to be habitual offenders . . whether they have been convicted or not" (although perhaps limited to "public safety" cases)
  - Judicial review of the grounds for a surveillance order may be necessary

Relationship to Indian Law, cont'd

- Program-driven (*PUCL*)
  - Requires a proclamation of an "Emergency" via public notification before large-scale interceptions for national security reasons may take place.
  - But the "bulk collection" issue is not settled (Bhatia, 2014)
- Aadhaar card
  - Not explicitly authorized legislatively until 2016
  - Only mandatory for those seeking government benefits (do legislators have a sufficient stake in the program?)
  - "Regulations" determine which biometric and demographic data is obtained, when required, and the measures to secure it
  - Police need a court order to access Aadhaar data, although for "national security" a 3-bureaucrat Oversight Committee decides