Search engines, GIS, and privacy

EPA Symposium

St. Louis, November 2007

Lee Tien
tien@eff.org

Electronic Frontier Foundation

Framing the privacy issues

- Surface
  - Law usually lags behind technology
  - Internet means widespread use
- Tough issues
  - Role of “intermediaries”
  - Problem of “privacy in public”

Electronic Frontier Foundation

Search histories are revealing

- Aug. 2006: AOL publicly released 3 months of search queries by 650,000 users
- Not identified by name — random ID #s
- But same ID # across each person
- Many people easily identified

Electronic Frontier Foundation

NYT: AOL searcher 4417749

- *Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.*

Electronic Frontier Foundation

- *No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from “numb fingers” to “60 single men” to “dog that urinates on everything.”*
- *There are queries for “landscapers in Lilburn, Ga,” several people with the last name Arnold and “homes sold in shadow lake subdivision gwinnett county georgia.”*

Electronic Frontier Foundation

- *It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs.*
- *"Those are my searches," she said, after a reporter part of the list to her.*

Electronic Frontier Foundation

The intermediary problem

- Search engines, banks, telcos . . . .
- We all use them, and they know what we do
- No 4th Amendment protection
- Maybe a statute, maybe company privacy policies or terms of service

Microsoft terms

- *When you register for certain Microsoft services, we will ask you to provide personal information. The information we collect may be combined with information obtained from other Microsoft services and other companies. We use cookies and other technologies to keep track of your interactions with our sites and services to offer a personalized experience.*

Electronic Frontier Foundation

Google Earth terms

- *Information collected by Google in connection with your use of the Software may be stored and processed . . . .*
- *Google may access, preserve, and disclose your account information if required to do so by law or in a good faith belief that such access, preservation or disclosure is reasonably necessary to . . . .* [list]

Electronic Frontier Foundation

Google data collection

- *We may combine personal information collected from you with information from other Google services or third parties . . . .*
- *Google's servers automatically record information when you visit our website . . . including the URL, IP address, browser type and language, and the date and time of your request.*

Electronic Frontier Foundation

So what's this mean?

- Since Google records everything, it will know every search/map query made by EPA
- And who at EPA made them (unique ID)
- Privacy issue if you're making personal searches, like Ms. Arnold
- Operational security if EPA personnel doing enforcement or investigations?

Electronic Frontier Foundation

Virtual mapping and location data

- Place is a linker like SSN
  - We can be linked to places we frequent
  - Arguably our “normal” pattern
  - Help identify our “abnormal” activities?
- Not just you: other people in your life
- Or others who go to the same places
- Data mining technology today . . . .

Electronic Frontier Foundation

Location: one court's views on GPS

- *a detailed record of travel to doctors' offices, banks, gambling casinos, tanning salons, places of worship, political party meetings, bars, grocery stores, exercise gyms, places where children are dropped off for school, play, or day care . . . the strip club, the opera, the baseball game, the 'wrong' side of town, the family planning clinic, the labor rally . . . . places that can reveal preferences, alignments, associations, personal ails and foibles.*

Electronic Frontier Foundation

Geodemographics

- DBs of public/private, individual/aggregate records on consumer identity and behavior
- GIS tools to analyze and graphically depict spatial distribution of our characteristics
- Segmentation schemes that identify consumer types through factor and cluster analysis of spatially referenced demographic and psychographic data

PRIZM: "You Are Where You Live"

- http://www.claritas.com/MyBestSegments/Default.jsp
- US neighborhoods are fit into 15 social groups
  - S1 Elite Suburbs
  - U3 Urban Cores
  - C1 2nd City Society
  - T2 Exurban Blues
  - R3 Rustic Living

Electronic Frontier Foundation

62 geodemographic clusters

- And then assigned into a specific cluster
- E.g. U1 Urban Uptown includes
  - Urban Gold Coast (elite urban singles)
  - Money and Brains (very affluent couples)
  - Young Literati (less $, more education)
  - American Dreams (immigrants)
  - Bohemian Mix (bohemian singles)

R1 Country Families includes:

- Cluster 44 “Shotguns and Pickups”
  - *Found in the Northeast, the Southeast, the Great Lakes and the Piedmont industrial regions . . . . lead the group in blue-collar jobs . . . . most are married with school-age children. They are church-goers who also enjoy hunting, bowling, sewing and attending auto races*

Location privacy underdeveloped

- Aerial surveillance Dow Chemical v US
- No 4th A "search" although *some of the photographs taken . . . at 1,200 feet are capable of enlargement to a scale of 1 inch equals 20 feet or greater, without significant loss of detail or resolution. When enlarged in this manner, and viewed under magnification, it is possible to discern equipment, pipes, and power lines as small as 1/2 inch in diameter.*

Possible limits?

- Pictures near home or *identifiable human faces . . . more serious privacy concerns*
- Lower courts: it's a search to use
  - High-power telescope to ascertain from 1/4 mile away what a person is reading inside high-rise apartment
  - Night scope to watch what couples are doing in their darkened bedrooms

Electronic Frontier Foundation

Exemplifies “privacy in public” issue

- Aerial surveillance merely one aspect
- Also tracking: US v Knotts
- By placing device inside container, federal agents tracked it from time of purchase in Minneapolis until delivery to cabin 100 miles away in rural Wisconsin.
- Was this a 4th Amendment search?

Electronic Frontier Foundation

No, because when driver . . .

- *traveled over the public streets he voluntarily conveyed to anyone who wanted to look the fact that he was traveling over particular roads in a particular direction, the fact of whatever stops he made, and the fact of his final destination when he exited from public roads onto private property*
- "voluntary exposure" rule

Electronic Frontier Foundation

Bad logic!

- “Anyone who wanted to look” couldn’t know that the container bought in Minneapolis was now at a secluded cabin in rural northern Wisconsin
- Need army of bystanders all along the route who pass on what they observe
- Why should fragmented disclosures destroy one’s travel privacy expectation?

Electronic Frontier Foundation

Fundamental error

- Privacy is not a discrete commodity, possessed absolutely or not at all
- Dramatic difference, in privacy terms, between revealing bits and pieces of information sporadically to a small and often select group for a limited purpose and a focused examination of all or much information about a particular person

Electronic Frontier Foundation

Canadian concern re Street View

- *does not appear to meet the basic requirements of knowledge, consent, and limited collection and use as set out in the legislation*
- *Our Office considers images of individuals that are sufficiently clear to allow an individual to be identified to be personal information within the meaning of PIPEDA*

Electronic Frontier Foundation

Google's response

- *Street View only features imagery taken on public property. This imagery is several months old and is no different from what any person can readily capture or see walking down the street.*
- Same logic we just saw in Knotts case
- But today data captures are routine and not done by law enforcement

Electronic Frontier Foundation

One of privacy's big problems today

- Much personal data in other's hands
- Routine surveillance of people, places; no need to plant trackers — we carry them already (e.g. cellphones)
- DBs, visualization tools, data-mining destroy "practical obscurity"
- Repurposes data collected for different purpose

Electronic Frontier Foundation

Federal laws behind the curve?

- Privacy Act great innovation in 1974
- Big weakness now: "system of records"
- any group of records where information is *retrieved* by person's name or by individual identifier *assigned to* the person
- www.usdoj.gov/oip/1974definitions.htm
- Think about GIS DBs indexed by location

Electronic Frontier Foundation

Spatial data rules and privacy?

- Much state, local data as well
- NSDI (Clinton, 1994) aims at sharing
- FGDC rules balance public right to know (toxic spills, other environmental issues) and critical infrastructure security
- Seems like little concern for privacy

Electronic Frontier Foundation

Conclusion

- Search engines: intermediary problem
- Mapping/tracking: “privacy in public”?
- Mass collection of data w/o any target in mind, but individuals can be “found” later
- Business & government incentives huge
- Resolution, data mining will only get better
- Laws haven’t adapted

Electronic Frontier Foundation