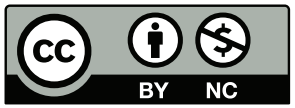
Databases and the Law



Historical Digression

- IBM mainframe computers still use EBCDIC (Extended Binary-Coded Decimal Interchange Code) instead of ASCII
- EBCDIC (vintage 1964) is derived from BCD; BCD is derived from punch card codes
- It is possible to do context-free translation from punch cards to BCD to EBCDIC; it is not possible with ASCII
- This was important because of the vast databases that had been compiled on *punch cards*
- (Historical note: the design for punch card processing equipment became an 1890 CU School of Mines (the ancestor of SEAS) PhD dissertation)

What's Different Now?

- Operations
- Scale
- Speed

Operations

- Punch card databases were limited in complexity: exceeding 80 bytes (plus 160 bits — don't ask. . .) per record was difficult
- It was easy to sort, count, and extract records from punch card databases
- It was *not* easy to do "join" operations — to match records from two or more databases

Scale

- Vastly more data is in machine-readable form
- A lot of data is generated by machine, which makes capture very easy
- The price of disk space has dropped even faster than CPU cycles, making long-term storage very cheap
- Data persists longer, creating a “data shadow” of our activities

The Data Shadow

- Term coined by Alan Westin (a CU prof) in the 1960s
- Many daily activities create a record somewhere
- Does your cell phone company routinely log your location when you're not talking? They certainly can
- Any credit card transaction does the same
- Your web site visits are quite trackable

A Visual Shadow?

- Face recognition is getting much better — think Google Picassa
- There are many surveillance cameras around — and some people want to link them
- Will a computer be able to track your movements?
- It's already very easy to do with license plates

Speed

- The speed of equipment and the programming techniques are vastly improved
- Operations that were always conceivable became economical
- It is (more or less) possible to process the huge amounts of data that are available

But. . .

- If this situation was so long in coming, why wasn't it anticipated?
- It was. . .

Ownership of Data

- Who owns data?
- More precisely, who owns data about *individuals*?
- Who regulates it?

American Attitudes

- Traditional American attitude: data belongs to the *compiler* of the data
- When you buy something and a store records it, the store owns the *record*
- Note: not the *fact* of the purchase, but the record of it
- But you own your image, though probably not the copyright on any pictures of you

More American Attitudes

- Free speech is a core value
- Recording and publishing of *truthful* information is acceptable
- No preemptive regulation of “speech”

Consequences

- Compiling and using data is accepted — the database is the property of the compiler
- As with other forms of property, it can be used, sold, etc., with few regulations
- Restrictions have been imposed after abuses

Example: Video Rental Records

"A video tape service provider who knowingly discloses, to any person, personally identifiable information concerning any consumer of such provider shall be liable to the aggrieved person." (18 USC 2710(b)(1))

But: "however, the subject matter of such materials may be disclosed if the disclosure is for the exclusive use of marketing goods and services directly to the consumer" (18 USC 2710(b)(2)(D)(ii) — but there is an opt-out provision)

Passed after a reporter obtained Judge Robert Bork's rental records, looking for evidence that he was viewing porn

Privacy Laws in America

- Very few, and mostly patchwork
- Major law: HIPAA, which protects medical information
- Credit records: primary issue is correctness
- FERPA: protects school records

HIPAA

- Intended to provide strong protections on disclosure
- But — given health payment system, doctors must disclose sensitive patient information to insurance companies
- Result: you're always asked to sign waivers
- Also — HIPAA applies to providers and insurance companies, not to third-party record providers like Google and Microsoft

The Medical Information Bureau (MIB)

- Fraud is a major concern for insurance companies
- They've organized the *Medical Information Bureau*, which is a clearinghouse of diagnosis codes
- Consulted by all life and health insurance companies, to look for patient omissions on applications
- Privacy waivers typically allow access to "all" records, not just health records

Electronic Health Records

- Promise: single database of all of your health information
- Available to all doctors, everywhere
- Who will have authorized access?
- ☞ The information is far more detailed than the MIB's data
- Who will have the ability to abuse authorized access?
- What are celebrity health records worth to the tabloids? What are your records worth to a snoopy neighbor?

But. . .

- Statistical information valuable to researchers
- Completeness and availability of records is a major health issue
- Very useful in case of emergencies
- Very useful when people switch doctors
- What about mentally incompetent patients?
- Single nationwide system currently stalled by cost and usability issues, as well as privacy concerns

Privacy and the Government

- Americans do not trust the government
- More restrictions on what the government can do, both in general and with data
- Major restriction: *Privacy Act* — establishes Fair Information Principles for the government
- Note well: does *not* apply to the private sector

The Privacy Act of 1974

- Individuals have the right to access and correct their records
- Records (generally) cannot be disclosed without consent
- Agencies must have legal authority to collect such data
- Agencies must publish *SORNs* (System of Records Notices)
- Agencies must carry out *PIAs* (Privacy Impact Assessments)

PIAs

- Nominally, honest descriptions of the privacy impact of a system
- (Can be partly or completely classified, for classified programs)
- Might be controversial. Example: DHS asserts that IP addresses are not personally identifiable information. The EU disagrees. Who's right?

The Private Sector

- Those rules do not apply to the private sector
- As noted, companies can generally do what they want with their data
- This is sometimes explicitly enshrined in the law

Phone Company Records

"A provider described in subsection (a) may divulge a record or other information pertaining to a subscriber to or customer of such service (not including the contents of communications . . . (6) to any person other than a governmental entity." (18 USC 2702(c))

That's right – phone companies and ISPs can sell your calling habits to anyone not in the government. (Effectively repealed in 1996, but not for privacy reasons.)

Can the government purchase such records from some third party? Yes. . .

More Casual Attitudes?

- Mark Zuckerberg of Facebook:

  "And then in the last 5 or 6 years, blogging has taken off in a huge way and all these different services that have people sharing all this information. People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm is just something that has evolved over time.

  "We view it as our role in the system to constantly be innovating and be updating what our system is to reflect what the current social norms are."

- Or — Google Buzz
- Abuse? Or reflection of societal change?

Government Abuses

- Government database systems are run by fallible, corruptible human beings
- Every few years, some IRS employees are caught looking at celebrity tax returns
- During the last presidential campaign, some people looked at candidates' passport applications
- Regular scandals involving abuse of law enforcement databases
- During the Nixon administration, the White House sought access to tax records of political enemies

Ourside the US

- Privacy laws outside the US are generally much stronger
- The EU and Canada have privacy laws that implement a variety of protections, including use limitations and restrictions on overcollecting
- In many countries, certain databases even need to be registered with a government privacy office

Privacy Laws and Security

- Privacy laws can protect personal security, too
- ☞ Data that isn't collected can't be stolen
- Example: TJX case, where many credit card numbers were stolen by hackers
- TJX's database included driver's license information, too, as protection against refund fraud — but this is sensitive from an identity theft perspective
- If they had stored the "hash" — a non-invertible function — of the license number, they could have achieved the same goal without the risk

But. . .

- Many such databases are useful to law enforcement, especially for anti-terrorism investigations
- Especially useful: communications records
- New EU law: data retention directive
- Requires phone companies and ISPs to retain records (including URL accesses) for 1-2 years
- The FBI wants a similar law here

International Issues

- With the Internet, offshore processing of data is very easy
- What if the receiving country has laxer data protection laws or practices?
- The TSA wants lots of information on EU airline passengers — but it's normally illegal to export that data from the EU to the US