Models of data protection (1)

- Comprehensive laws:
  - general laws governing collection, use and dissemination of data and an oversight body (adopted by EU)
  - Variation: coregulatory model (Canada, Australia). Industry develops rules.
- Sectorial laws:
  - different rules for different sectors (e.g. Financial information, medical records) (adopted in USA)

Models of data protection (2)

- Self-regulation:
  - Code of practice
  - Code of conduct
- Technology for privacy:
  - Exploitation of technology for privacy protection: for example encryption.

(incomplete chronological list of) Major Privacy Legislation/Regulation

- 4th Amendment
  - Perhaps the oldest that is currently in effect/direct impact
- EC 95/46
  - Widespread/international impact
- U.S. agency rules
  - Specific discussions of location aggregation
- HIPAA
  - Technically meaningful guidelines for anonymity
- EC 2002/58
  - Explicitly discusses location

4th Amendment to the U.S. Constitution

- Protects against government intrusion into private life
  - Goal: Prevent suppression of peaceful political dissent
- The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.
- Protects activities that a reasonable person would not expect to be visible to other than known observers
  - Most location data likely to fail this test

EC95/46: European Directive on Privacy

http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm

- Passed European Parliament 24 October 1995
- Goal is to ensure free flow of information
  - Must preserve privacy needs of member states
- Effective October 1998
- Effect
  - Provides guidelines for member state legislation
    - Not directly enforceable
  - Forbids sharing data with states that don't protect privacy
    - Non-member state must provide adequate protection,
    - Sharing must be for "allowed use", or
    - Contracts ensure adequate protection
  - US "Safe Harbor" rules provide means of sharing (July 2000)
    - Adequate protection
    - But voluntary compliance
- Enforcement is happening
  - Microsoft under investigation for Passport (May 2002)
  - Already fined by Spanish Authorities (2001)

The EU data protection directive (95/46/EC)

- The right to know where the data originated
- The right to have inaccurate data rectified
- The right of recourse in the event of unlawful processing
- The right to withhold permission to use data in certain circumstances

EU 95/46/EC: Meeting the Rules

- Personal data is any information that can be traced directly or indirectly to a specific person
- Use allowed if:
  - Unambiguous consent given
  - Required to perform contract with subject
  - Legally required
  - Necessary to protect vital interests of subject
  - In the public interest, or
  - Necessary for legitimate interests of processor and doesn’t violate privacy

The relevant (to technolgy) definitions (1)

- Definition (a):

  “personal data” shall mean any information relating an identified or identifiable natural person (“data subject”); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;

The relevant (to technolgy) definitions (2)

- Definition (b)

  “processing of personal data” (“processing”) shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.

The relevant (to technolgy) definitions (3)

- Premise 2:

  Data-processing systems are designed to serve man; they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, and contribute to economic and social progress, trade expansion and the well-being of individuals.

EU 95/46/EC: Meeting the Rules

- Some uses specifically proscribed
  - Can't reveal racial/ethnic origin, political/religious beliefs, trade union membership, health/sex life
- Must make data available to subject
  - Allowed to object to such use
  - Must give advance notice / right to refuse direct marketing use
- Limits use for automated decisions (e.g., creditworthiness)
  - Person can opt-out of automated decision making
  - Onus on processor to show use is legitimate and safeguards in place to protect person's interests
  - Logic involved in decisions must be available to affected person
- europa.eu.int/comm/internal_market/privacy/index_en.htm

US Health Insurance Portability and Accountability Act (HIPAA)

- Governs use of patient information
  - Goal is to protect the patient
  - Basic idea: Disclosure okay if anonymity preserved
- Regulations focus on outcome
  - A covered entity may not use or disclose protected health information, except as permitted or required…
    - To individual
    - For treatment (generally requires consent)
    - To public health / legal authorities
  - Use permitted where “there is no reasonable basis to believe that the information can be used to identify an individual”
- Safe Harbor Rules
  - Data presumed not identifiable if 19 identifiers removed (§ 164.514(b)(2)), e.g.:
    - Name, location smaller than 3 digit postal code, dates finer than year, identifying numbers
  - Shown not to be sufficient (Sweeney)
  - Also not necessary
- http://www.hhs.gov/ocr/hipaa/finalreg.html

EC 2002/58

http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm

- Extends EC 95/46 to electronic communications
- Explicitly mentions location
  - *Location data may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded*.
  - *Where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service.*
- EC 2006/24: Requires retention of communication metadata
  - For law enforcement / investigation (e.g., anti-terrorism) only
  - Location of service activation if sender/recipient unknown
  - Retain 6 months to 2 years

Cdn. Initiatives

- 1975: Quebec *Charter of Human Rights & Freedoms*
  - “every person has a right to respect for his private life”
- 1982: *Canadian Charter of Rights and Freedoms*
- 1980s: Public sector privacy laws
- 1990s: CSA Model Privacy Code
  - based on Fair Information Principles (FIPs)
  - adopted as formal standard in 1996
  - incorporated into federal law: PIPEDA
- 1994: Quebec private sector law
- 2001: Federal private sector law
- 2004: Alta, B.C. private sector laws

courtesy of Pippa Lawson, CIPPIC

Privacy Commissioners

- Federal + some provincial
  - Ontario, B.C., Alberta
- Public sector vs. private sector
- Ombuds vs. binding powers
- Role as educators, advocates, watchdogs, dispute resolvers, reporters…

courtesy of Pippa Lawson, CIPPIC

Charter of Rights

- s.7: "*Everyone has the right to life, liberty, and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice"*
  - emerging privacy right
- s.8: *"Everyone has the right to be secure against unreasonable search or seizure"*
  - protects an individual's "reasonable expectation of privacy" (usually in criminal law context)
- s.1: Rights are subject to "*such reasonable limits as can be justified in a free and democratic society*"

Public Sector legislation

- Federal: *Privacy Act*
- Provincial:
  - Ontario *Freedom of Information and Protection of Privacy Act* (“FIPPA”)
  - similar statutes in other provinces

Private Sector Legislation

- PIPEDA
  - federally regulated
  - interprovincial or international data flows
  - where no “substantially similar” provincial law
  - applies to “organizations” in the course of “commercial activities”
- Quebec, Alberta, B.C. laws
  - provincially regulated, in those provinces
  - cover non commercial activities as well

PIPEDA

- Purpose:
  - balancing individual's "right of privacy" with "[legitimate] need of organizations"
- Protects:
  - "personal information"

    = "information about an identifiable individual"

Canada: PIPEDA (2000)

- Personal Information Protection and Electronic Documents Act
- Support & promote e-commerce by “protecting personal information that is collected, used or disclosed in certain circumstances, by providing for the use of electronic means to communicate or record information or transactions”
- States that an organization “may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances”
- It does not apply to data “rendered anonymous” and no reasonable method of identification

PIPEDA: Principles

1. Accountability
2. Identifying Purposes
3. Consent
4. Limiting Collection
5. Limiting Use, Disclosure and Retention

6. Accuracy
7. Safeguards
8. Openness
9. Individual Access
10. Challenging Compliance
11. Limiting Purposes

courtesy of Pippa Lawson, CIPPIC

Other Initiatives

- Canadian *Principles for Electronic Authentication* (2004)
  - "….the collection, use and disclosure of personal information in the context of authentication should be minimized."
    - applies to designers as well as those using authentication mechanisms

courtesy of Pippa Lawson, CIPPIC

Canada: CIHR

- Canadian Institutes of Health Research
- Proposed clarification of PIPEDA that offers an interpretation of “reasonableness”
- “A reasonably foreseeable method” of identification or linking of data with a specific individual
- BUT, it also refers to “anonymized” data as information “permanently stripped” of ***all*** identifiers, such that the data has “no ***reasonable*** potential for any organization to make an identification”
- Finally, it states that reasonable foreseeability should “be assessed with regard to the circumstances prevailing at the time of the proposed collection, use, or disclosure.”

Canada: AHIA

- Alberta Health Information Act
- Takes a different perspective than CIHR
- Individually identifiable is defined as “can be readily ascertained from the information”
- Non-identifiable is defined as “cannot readily ascertained from the information

Regulatory Constraints: Use of Results

- US Telecom (Fraud, not marketing)
  - Federal Communications Commission rules
  - Rooted in antitrust law
- US Mortgage “redlining”
  - Financial regulations
  - Comes from civil rights legislation
- Spanish Case Law on EC 2002/58
  - Location data retained for law enforcement not useable for civil case (copyright)

What is “made anonymous”?

- *principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable* EC95/46 1(26)
  - *an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity* EC 95/46 2(2)
- HIPAA: *no reasonable basis to believe that the information can be used to identify an individual* 164.514(1)
  - First 3 digits of postal code (metropolitan area) (HIPAA 164.514(2)(i)(B)
    - Minimum 20,000 individuals

Data Protection Working Party 4/2007 Opinion (01248/07/EN WP 136)

- *In general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group.*
  - Is 2-anonymous anonymous?
  - Or does this mean every individual is 2-anonymous w.r.t. all subsets of the data?
  - Ex: *If a criterion appears to lead to identification in a given category of persons, however large (i.e. only one doctor operates in a town of 6000 inhabitants), this "discriminating" criterion should be dropped altogether or other criteria be added to "dilute" the results on a given person so as to allow for statistical secrecy.*
- Key-coded data (unique identifier) can be considered anonymous when the mapping of the identifier to the individual is not disclosed
  - Commission Decision 2000/520/EC of 26.7.2000 - O. J. L 215/7 of 25.8.2000

The relevant (to technolgy) definitions (4)

- Premise 26

  The principle of protection must apply to any information concerning an identified or identifiable person; whereas to determine whether a person is identifiable, account should be taken of all means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable

Data Protection Working Party 4/2007 Opinion (01248/07/EN WP 136)

- *this test is a dynamic one and should consider the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed. … If the data are intended to be stored for one month, identification may not be anticipated to be possible during the "lifetime" of the information, and they should not be considered as personal data. However, it they are intended to be kept for 10 years, the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment.*

The relevant (to technolgy) definitions (5)

- Premise 26 (cont.)

  whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible.

- Codes of conduct are issued by member states (e.g. Italy issued a code of conduct for journalists)

The relevant (to technolgy) definitions (6)

- Premise 29

  The further processing of personal data for historical, statistical or scientific purposes is not generally to be considered incompatible with the purposes for which the data was originally collected provided that member states furnish suitable safeguards; whereas these safeguards must in particular rule out the use of the data in support of measure or decisions regarding any particular individual

US National Center for Health Statistics (1978)

- “[All] micro data which are released outside of the NCHS, geographic identification must be deleted for all areas below the State level which contain fewer than 250,000* inhabitants in the most recent official population Census”

*Numbers vary across countries and the agencies within them

The relevant (to technolgy) definitions (7)

- Premise 40

  It is not necessary to impose this obligation [...] if would involve disproportionate efforts, which could be the case where processing is for historical, statistic or scientific purposes; whereas in this regard the number of data subjects, the age of the data, and any compensatory measures adopted may be taken into consideration.

US National Institutes of Health (2003 & 2006)

- 2003 Data Sharing Policy
  - Philosophy: Promote openness; share datasets that are costly to generate
  - Who: Any investigator receiving at least $500,000 in any year of an NIH-supported Study must have a data sharing plan for the data upon which their findings are based
  - Privacy: Data must be shared in a de-identified manner. No explicit directions are given, but the policy refers to HIPAA
  - IMPORTANT NOTE: An investigator may withhold data if they can explain why (e.g., privacy protection destroys data usefulness or scientific validity)!
- 2006 Data Sharing Policy for Genome Wide Association Studies
  - Extends the who to any investigator receiving NIH funds

Census & Policy

- Goal of the Census: Collect detailed survey information on a representative population.
- Dissemination of results to the public
  - Microdata: Can not disseminate information with a Census block smaller than X
  - Tabular (Summary Data): Can not share information with multi-dimensional cell counts below a small number
    - Informally defined as four or five
  - In practice, Census uses many protection mechanisms beyond “aggregation”, including “swapping”, “simulation”, “perturbation”, and “suppression”

What is Private? *US Case: Forest Guardians vs. FEMA (2005)*

- Forest Guardians asked for
  - electronic copies of GIS maps with federal data on Hurricane Katrina (home-level information, including financial information; e.g. compensations)
  - data without names & addresses
- FEMA (Federal Emergency Management Agency) refused
- Court ruled FEMA **did not have to disclose** the information because there was a chance the data could be linked to individuals

What is Private? *US Case: Multi Ag Media vs. USDA (2008)*

- United States Dept. of Agriculture (USDA) withheld
  - crop data provided by agricultural producers
  - GIS data including “information on farm, tract, and boundary identification, calculated average, and characteristics of the land”
    - Common Land Unit (CLU: now a standardized GIS field)
  - Multi Ag Media filed a Freedom of Information Act request for “public spatial data”
  - Portion of the request held back for “privacy reasons”
  - Court ruled in favor of disclosure, but stated there existed a privacy interest
  - Court agreed with USDA that data could crop finance information could be analyzed to reveal farm’s worth

What is Private? *US Food, Conservation, & Energy Act of 2008*

- CLU GIS data was made “private” – information no longer releasable
  - to the general public, as well as
  - most government agencies

- “In general, NRCS (National Resource Conservation Service) technical and financial information is **not releasable** to the public, and It cannot be released to any person, Federal agency, local agency, or Indian tribe outside of USDA”

Mobility data retention: 3 schemes

- Personal mobility/location data collected by service providers (telecom, navigation, …) are subject to (at least) 3 different data retention schemes:
  - Retention for service-related operations (delivery, billing, network optimiztion, …)
  - Retention for law enforcement (inspection by police or judge for investigation, …)
  - Retention for analytical purposes (mobility data mining, GeoPKDD, …)
- The 3 schemes are subject to
  - 3 different regulatory contexts
  - 3 different possibilities of exploiting anonymity

Laws for the 3 retention schemes (in Europe)

- Retention for service-related operations (delivery, billing)
  - The EU privacy framework speficies the responsabilities of the data custodian (i.e., the service provider)
- Retention for law enforcement (inspection by police or judge for investigation)
  - The EU Data Retention Directive 2006/24/CE for electronic communications specifies times and procedures (period of retention: at least 6 months, at most 2 years).
- Retention for analytical purposes (mobility data mining, etc.)
  - Anonymous data are outside the above two directives, but no regulatory framework for anonymity for mobility data has been set.

Anonymity in the 3 retention schemes

- Retention for service-related operations (delivery, billing)
  - Level of protection is decided in the contract between service providers and customer – LBS privacy
- Retention for law enforcement (inspection by police or judge for investigation)
  - No room for anonymity
- Retention for analytical purposes (mobility data mining, etc.)
  - A regulatory framework for anonymity would free data for scientific/social/commercial purposes