Data Privacy

INFX 551
Winter 2018

Course Outline

Week 7

Data	Data Systems	Policy, Privacy, & Ethics
Types & Roles	Repositories	Policy
Documentation	Preservation	Privacy
Standardization	Cost Models	Ethics

Agenda

- Legal Frameworks for intersection of data and privacy
- Conceptualizations of Privacy
- Methods of Protecting Data Privacy
- Data Curation and Data Privacy

“Data are information objects playing the role of evidence…”

“Privacy is…” <—- ?

Privacy roots

- 1890 Warren and Brandies ‘Right to Privacy’ - “right to be let alone”
    - Daniele Solove (1980s) information privacy is not well served by appealing to definitions privacy.
        - ICTs move faster than the law (that is a feature not a bug)
        - Focus on problems introduced by emerging technologies - and by nature of defining a problem norms will emerge
            - Taxonomy of Privacy: Information collection, processing, dissemination, and invasion

Legal Context in USA

- 4th Amendment (Common Law Privacy Rights) - “right to be secure” and entitlements to privacy of self, possessions, and proceedings.

- Privacy Act of 1974 - “any item, collection, or grouping of information about an individual” by a government agency - IF and ONLY IF that info includes PII

- eGovernment Act of 2002 - broad framework to “enhance citizen access to Government information and services.” - provisions include what gov knows and stores about public.

USA vs EU

1. Privacy laws change with each administration

2. Individuals have little ownership of their online data, which allows large businesses can monetize consumer behavior and habits.

3. Privacy laws are often a messy combination of public regulation, private self-regulation, and legislation which varies by state.

4. Enforcement of privacy laws is carried out by several different government organizations, e.g. Federal Communications Commission (FCC) and Health Insurance Portability and Accountability Act (HIPAA).

5. Numerous privacy organizations exist to provide legal framework, which ensure digital privacy to Americans. Ex: American Civil Liberties Union (ACLU) and the Electronic Frontier Foundation (EFF).

6. Companies can keep data indefinitely, depending on their own Terms of Service

1. Privacy laws have less turnover when administrations change because most EU member states aren't as polarized as the US.

2. EU laws respect "private and family life" and allow citizens to delete their data.

3. Privacy laws are generally more comprehensive and geared towards consumers.

4. Enforcement of privacy laws is carried out by one authority, equally for all 28 member states.

5. Due to the nature of EU rights, fewer privacy organizations exist but there are: The European Digital Rights (EDRi) and The European Privacy Association (EPA.)

6. EU citizens have the "right to be forgotten," meaning that search results can be removed if they are irrelevant or inadequate

Sources:

https://www.marketplace.org/2017/04/20/tech/make-me-smart-kai-and-molly/blog-main-difference-between-internet-privacy-us-and-eu

http://politicsandpolicy.org/article/european-union-and-internet-data-privacy

HIPPA - Health Insurance Portability and Accountability Act of 1996.

- Protected Health Information (PHI) = “any information about health status, provision of health care, or payment for health care”

- Title 2 establishes the “Privacy Rule”

- Regulatory framework for protecting PHI, and establishes sanctions for violation of privacy rule

    - 45 CFR 164.510(b) – the requirements to obtain a patient’s agreement to speak with family members or friends involved in the patient’s care.

    - 45 CFR 164.520 – the requirement to distribute a notice of privacy practice

    - 45 CFR 164.522(a) – the patient’s right to request privacy restrictions.

    - 45 CFR 164.522(b) – the patient’s right to request confidential communications.

* FERPA - Family Educational Rights and Privacy Act contains similar provisions (but more comprehensive privacy regulations) for student education records.

For the sake of general data curation, we care about data privacy in three contexts:

- Storage
- Access
- Use (or analytics)

Note how these map on to Solove's Taxonomy of Privacy: Information collection, processing, dissemination, and invasion

Approaches to Data Privacy through **Storage**

- **Security:** encryption (multiple factor authentication, key generation, etc.)

- **What to store**…

    - Data Minimization - Storing only minimal information for minimal periods - the idea being risk of breach is far greater than profit of storage (and reuse). (Not a great option for research data)

    - Data Sovereignty - Data should be stored geographically proximate to subjects. Data storage media treated to same laws / regulations, etc. as locale under which it was gathered.

Approaches to Data Privacy through **Access** Restrictions

- **Keep people out:**
    - Sanctioned Access (credentialed access) - a secondary institution creates system of credentialing for verified users of data

- **Obscure stored data**:
    - Data Salting - adding known errors to data in order to reduce potential that it can be meaningfully reused - storing the algorithm to re-normalize the data separately.

Approaches to Data Privacy through **Analytic Restrictions**

- **Differential Privacy** - Introduced errors that do not impact statistical significance. Imagine (simply) introducing enough “noise” into data that individuals can’t be recognized, but demographic summaries can be created and verified.

- **Honest Broker systems** - provide framework that separates PHI from individual specimens. Hospitals, clinics, and research labs often provide honest broker systems for biological research on tissue.

Curation + Privacy by Design (PbD)

- Privacy by Design is the idea that protecting sensitive information, and privileging privacy SHOULD be a part of all information systems / data collection projects

- PbD is becoming an increasing concern in Data Management Plans - And therefore, an important component of Data Curation.

- Curators role in DbP is to help various stakeholders understand the potential harms and risks

    - Privacy Risk Assessments.