

CSE 303: Concepts and Tools for Software Development

Hal Perkins

Autumn 2008

Lecture S2— Societal Implications: Data Ownership & Privacy

A Day in the Life

- You buy groceries and use your “club card” to get the “discount”
 - What you bought, and when, is logged with your “club” id
- You log on to Amazon; they suggest things you might want based on your previous history, you buy something
- You search for “amazons” on Google
 - Your IP address, the time, and the query are logged
- You use your new ORCA pass to ride the bus, light rail, ferry, or
- You use the HOT toll lanes or the new Tacoma narrows bridge . . .
 - The time and location of your trip is logged
- You use your bankcard to buy something
 - The details are logged and used to debit your account

Theme: Trading information (& privacy) for convenience

Your Online Life

- You post information on your UW webpage
 - The university (e.g., Wa. state) owns the computers; is it their data?
- You post all sorts of information on Facebook
 - Other people post comments on your pages
- You post a video clip to YouTube
 - Someone claims that it's theirs
 - YouTube takes it down for copyright infringement
- The things you post are scanned and indexed by Google, others
 - Everything on the web lives *forever*

Theme: Who's data is it? Who controls it?

Simple (?) Questions

- What data should be publicly available?
- What should be stored about where users web-surf?
- Who should be to blame when the wrong data / surf-histories become known?
- Can data be used for unanticipated purposes once it's been collected? (e.g., data mining)
- Can someone associated with data require that it be corrected/deleted? Who decides?
- Can an organization operate effectively if it does or doesn't control data it collects or uses?

Data Content

- Is it clear “who controls” a web-site? Should it be?
- My homepage, but on a government site.
- Search engines copy other possibly illegal (in what country?) sites.
- YouTube, Facebook, Flickr, ... post strangers' content.
- What if gmail or Google Calendar “got hacked”?

A universal dilemma: replication increases recovery (backups) and performance, and decreases security (privacy).

Anachronistic laws? Is “who owns the hard-drive” the point? If not, then what is?

Surfing / Search-query privacy

- Search for “drug rehab”, “domestic-violence shelter” or “how to make a bomb”, “child pornography”
- Jealous ex-S.O. snooping IM or parent snooping 8-year-old’s IM
- Lawyers can subpoena highway toll records in divorce cases
- Amazon suggesting “similar products” or totalitarian regime finding buyers of “banned books”
- Employers have right to ensure productivity and network-security or employees have right to “close the door”

A universal dilemma: data can improve user-experience, but once collected can be used for other purposes.

Is there a “best answer” for all/most web-sites or networks? How can a user / site protect him/her/itself?

More Questions

1. How responsible is the “equipment owner” for content?
2. Does a “big site” have greater obligations than a “small site”?
3. What bounds should there be on “terms of service” for web sites? Should you read them?
4. What control should you have over data about you?
5. Do technical solutions (passwords, log-deletion, ...) help or is this “just” a legal/ethical issue?
6. How long should data be kept? Who should decide that?
7. How much privacy would you give up for “something really cool”?
8. How have data privacy issues changed in the last 5 years? How much is technical? Social?
9. ...