Usability of privacy policies: Notice and choice

Michelle Mazurek

With material from Lorrie Cranor, Florian Schaub, and Carman Neustaedter
Logistics

- HW3 grades are out
  - See detailed comments in ELMS
- Today: Privacy policies, notice and choice
- Today/Thurs: Coding qualitative data
  - Important for HW4
Review: Privacy self-regulation
Notice and choice

Protect privacy by giving people control over their information

Notice about data collection and use

Choices about allowing their data to be collected and used in that way
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits

*We know this does not occur in practice!*
Approaches to improvement

• Better labels and icons
  – Nutrition labels
  – Privacy icons

• Automated policy processing
  – P3P
  – Do Not Track
  – Crowdsourcing
  – NLP
Towards a privacy “nutrition label”

• Standardized format
  – People learn where to find answers
  – Facilitates policy comparisons

• Standardized language
  – People learn terminology

• Brief
  – People find info quickly

• Linked to extended view
  – Get more details if needed
Iterative design process

• Focus groups, lab studies, online studies
• Comparison to text, standardized text, etc.
• Metrics
  – Reading comprehension (accuracy)
  – Time to find information
  – Ease of comparison between policies
  – Subjective opinion (easy, fun, trustworthy)

<table>
<thead>
<tr>
<th>Information we collect</th>
<th>Ways we use your information</th>
<th>Information sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>provide service and maintain site</td>
<td>marketing</td>
</tr>
<tr>
<td>contact information</td>
<td>opt out</td>
<td>opt out</td>
</tr>
<tr>
<td>cookies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>demographic information</td>
<td>opt out</td>
<td>opt out</td>
</tr>
<tr>
<td>financial information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>health information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>preferences</td>
<td>opt out</td>
<td>opt out</td>
</tr>
<tr>
<td>purchasing information</td>
<td>opt out</td>
<td>opt out</td>
</tr>
<tr>
<td>social security number &amp; gov't ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>your activity on this site</td>
<td>opt out</td>
<td>opt out</td>
</tr>
<tr>
<td>your location</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Privacy label for Android

**Description**
Like word games? Like weasels? This is the game for you. Word Weasel is a fast word game where you find as many words as you can from 9 letters in 50 seconds. Compete with everyone else playing at the same time, a new game starts every minute!

"The most fun you can have on your own with 9 random letters. Brilliantly simple, devilishly addictive." —Kim, Android Market comment

**Privacy Facts**
This app collects your:
- Personal information
- Location
- Calendars
- Photos

This app uses:
- Credit card / financial
- Diet / nutrition
- Health / medical
- Photos
- Analytics

**Reviews**
Average 4.6

<table>
<thead>
<tr>
<th>Stars</th>
<th>Rating</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>364</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>210</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

Ad supported, with ability to upgrade to ad-free full version in game.

Uses the SOWPODS word list containing 100,000+ words. They may be obscure, but yes those are words!
Role play studies

• Select apps for friend with new Android phone
  – Choose from 2 similar apps in each of 6 categories
  – Click on app name to visit download screens
  – Different permissions per app

• Post-task questionnaire

• Participants who saw Privacy Facts more likely to select apps that requested fewer permissions
  – Other factors such as brand and rating reduce effect

P.G. Kelley, L.F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. CHI 2013.
Your Data is Used Only for the Intended Use

Your Data May be Used for Purposes You Do Not Intend
In groups: Design icons and tag lines for smartphone app privacy

1. App only collects the information it needs to work and only uses and shares information as necessary to provide the service you requested

2. Same as 1 but app also collects information about your location and use of apps and provides it to advertising companies to target ads to you

3. App may collect any information and use or share it for any purpose
P3P Overview (Review)

- W3C specification for XML privacy policies
  - Proposed 1996
  - Adopted 2002
- Optional P3P compact policy HTTP headers to accompany cookies
- Goal: Your agent enforces your preferences
- Lacks incentives for adoption
Too much is not enough?

• 17 data categories, 12 possible collection purposes, 6 possible recipients, 5 retention policies
  – Annotations: description, opt-in/out, etc.

• Too much detail? Insufficiently expressive?
  – Both!
Why provide more detail?

- Companies’ actions are nuanced
- What is important may change over time
- Broad categories may make things look worse
  - Compact P3P policies
- Provide all info and let user agent sort it out
Why is this too much detail?

• Difficult to author a policy accurately
  – Ambiguous, redundant categories
• Bugs in user agent parsing/display
• Different agents may abstract differently
  – Hard for users to compare across tools
  – Companies must test different views
Do Not Track

• An HTTP header sent by your browser
  – Websites and services can promise to respect it
  – No client-side enforcement

• What does tracking mean?

• The problem of defaults
CMU Usable Privacy Project

• Semi-automatically extract data practices from privacy policies
  – Crowdsourcing, machine learning, NLP
• Understanding and modeling user preferences
  – Focus on data practices users care about
• Provide effective for privacy notices
• Large-scale analysis of website privacy policies
  – To inform public policy
Which practices are relevant?

• From FTC enforcement, class action suits:
  – Unauthorized disclosure
  – Surreptitious collection
  – Unlawful retention
  – *Do you think this is the right approach?*

• Prior studies of privacy concerns:
  – Contact info, location, financials, health
Crowdsourcing policy extraction

- Does the site collect X information?
  - Yes, no, unclear; provide evidence
Crowdsourcing policy extraction

• Compare results: crowdworkers vs. experts
  – 76% of cases: crowdworkers agree w/ experts
  – 2%: agree with each other, but not experts
  – 22%: don’t reach consensus

User interface goals

• Use extracted data to inform consumers
  – With level of confidence
• Enable meaningful comparisons with similar sites
• Design and testing in progress

• What will it take for users to pay attention?
Summary: Privacy notice and choice

• Only works if understandable, actionable
• Incentives, enforcement are critical
• Better together: automated policy reading, usable notices and icons
  – Standardized, layered
• The problem of expressiveness
CODING QUALITATIVE DATA
Qualitative coding

• Today: Types of coding and methods
  – Open, axial, systematic

• Thursday:
  – Validating coded data
  – Reporting coded data
  – Hopefully: Try it!

• You may feel uncomfortable with this!
  – Work carefully, use established methods
Kinds of coding

• **Open coding (inductive)**
  – When you aren’t sure what you’re looking for
  – Fine-grained details

• **Axial coding (inductive)**
  – Draw connections and themes (from data or codes)
  – One option: Affinity diagrams

• **Systematic coding (deductive)**
  – When you start from a hypothesis or theory
Open coding

• Inductive: For generating theory
• Treat data as answers to open-ended questions
• Formulate questions (mostly) ahead
  – Go through transcript, asking the questions
  – Encounter a new possible answer, make a code
  – Record the participant, the code, and the evidence
Example: Access control in the home

• Questions:
  – What data should be protected?
  – How are physical files protected?
  – How are digital files protected?
  – Has the participant had a bad experience?

• Update and refine questions as you go
Example codes

• “If I didn’t want everyone to see them, I just had them for a little while and then I just deleted them.” (DEL)

• “If you name something ‘8F2R349,’ who’s going to look at that?” (HID)

• Using a laptop password “just in case … we have guests over” (PWD)
Keeping track

- Codebook: Questions, possible answers
- Excel, db software, expensive coding tools
- Track per question:
  - Participant
  - Code(s)
  - Where you found evidence, quotes
Axial coding – Finding themes

• Group low-level codes into categories
• One method: affinity diagramming
  – Write low-level data/codes on sticky notes
  – Group hierarchically
  – Update as themes emerge
Example: Calendar field study

- **Step 1: Affinity Notes**
  - Each note contains one idea / code
  - Place them on a flat surface / wall

It was really easy to check the calendar from work because of the web page.
The colors on the events made it really easy to see who had events.
I check my calendar on my cell phone while driving.
The size of the writing was too small to read.
We couldn't place the calendar in the spot we usually do in our home.

Neustaedter, 2007
Diagram building

- Collect related notes
Diagram building

The size of the writing was too small to read.

The colors on the events made it really easy to see who had events.

I check my calendar on my cell phone while driving.

It was really easy to check the calendar from work because of the web page.

we couldn’t place the calendar in the spot we usually do in our home.
Diagram building

- The size of the writing was too small to read.
- The colors on the events made it really easy to see who had events.
- We couldn't place the calendar in the spot we usually do in our home.
- I check my calendar on my cell phone while driving.
- It was really easy to check the calendar from work because of the web page.
Diagram building

- We couldn't place the calendar in the spot we usually do in our home.
- The colors on the events made it really easy to see who had events.
- The size of the writing was too small to read.
- I check my calendar on my cell phone while driving.
- It was really easy to check the calendar from work because of the web page.
Write affinity labels for each group

• Continue to further refine groupings

Calendar placement is a challenge
- We couldn’t place the calendar in the spot we usually do in our home.

Interface visuals affect usage
- The colors on the events made it really easy to see who had events.
- The size of the writing was too small to read.

People check the calendar when not at home
- I check my calendar on my cell phone while driving.
- It was really easy to check the calendar from work because of the web page.
Systematic coding

• For testing an existing hypothesis/theory
• Codes are created ahead of time
  – Before interviewing!
  – From existing literature/theory
  – From prior rounds of open coding
• Code just as before
Summary: Qualitative coding

• Generating or testing a theory?
  – Open, axial vs. systematic

• Short codes representing possible answers
  – If open coding, refine as you go

• Carefully track codebook and evidence