Privacy Policy: Alignment with Systems, Legislation and People

Julie Earp
College of Management
North Carolina State University

WISE 2010

Sponsored by TRUST
June 21-24, 2010: Vanderbilt University

Research Motivations

- Consumers are increasingly concerned about privacy violations
- Consumers don't have a good understanding of what happens to their information
- Companies are increasingly being held accountable for their privacy practices
- Privacy laws require companies to enforce their policies

...how can companies ensure that consumers understand what will happen with their data?

Disclosure of Privacy Practices

But really… who cares?

- Consumers
  - Seeking protection
  - Awareness
  - Will they keep their promises?

- Businesses
  - Must comply with legislation
  - FTC Act

- Software Engineers
  - Legally compliant software
  - Policy compliant systems

Who has a privacy policy?

Privacy Policy Content

- **2002 and 2008:** Consumers are most concerned with (in order):
  - information transfer
  - notice/awareness
  - information storage
- **2002:** Privacy policies emphasize (in order)
  - data integrity/security
  - information collection
  - user choice/consent

Are privacy policies readable?

the privacyplace.org

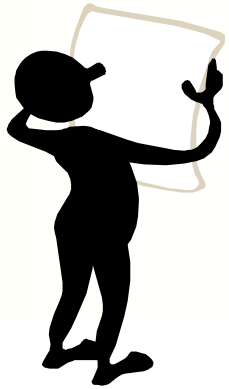
Understanding "well-written"

Education and Internet Use

[IEEE S&P, 2004]

Educational Level	General Population (GP)		%GP Online	Internet Population	
	# People (in millions)	%of Total Population		# People (in millions)	%of Internet Population
Less Than High School	27.5	15.5	12.8	3.5	3.8
High School Diploma / GED	57.4	32.4	39.8	22.8	24.5
Some College / Associate D.	45.4	25.6	62.4	28.3	30.5
Bachelors Degree	30.6	17.7	80.8	24.7	26.6
Beyond Bachelors	16.3	9.2	83.7	13.6	14.6

Calculating Privacy Policy Readability



- **Flesch Reading Ease Score (FRES):**
  - 206.835 - 84.6 * (total syllables/ total words) – 1.015 * (total words/total sentences)

- **Flesch Grade Level (FGL):**
  - (0.39 * Average sentence length (in words)) + (11.8 * Average number of syllables per word) - 15.59

The Lack of Clarity in Financial Privacy Policies and the Need for Standardization, Antón, Earp, Bolchini, He, Jensen and Stufflebeam. "*IEEE Security & Privacy*, 2(2), pp.36-45, March/April 2004.

Readability Impact

[IEEE Security & Privacy, 2004 and 2006]

- Only 52% of general population has obtained at least 2 years of college
- Financial Industry (2004)
  - 40 online privacy documents at 9 financial websites
  - Flesch readability range = 10.42 – 18.72
  - Avg. Flesch readability level = 14.1 (2 years of college)
- Healthcare Industry (2006)
  - 24 online privacy documents at 9 healthcare websites
  - Avg. Flesch readability level = 14.2 (2 years of college)
- Difficult policies are time consuming to read

Is the Clarity Requirement Satisfied?

- A breakdown of those 40 financial policies:
  - 8 require equivalent of a high school education or less
  - 13 require equivalent of some college education
  - 12 require 14-16 years of schooling
  - 7 require the equivalent of a postgraduate education (> 16 years).
  - 2/3 institutions had at least one policy document requiring the equivalent of a postgraduate education

- A full understanding of what 2/3 of these organizations are promising is only available to <15% of the adult Internet population

User perception vs. comprehension of policies…

t h e p r i v a c y p l a c e .o r g

Privacy Policy Experiment

[IEEE Trans. On Engineering Management, 2007]

- Experiment
  - Investigate user comprehension and perception of privacy policy expressions
  - Compare user perception with user comprehension in order to determine whether they are in alignment with one another
  - Theoretical framework based on the Privacy Taxonomy
  - 993 responses

Experimental Design

- Factor 1: Three Website Policies
  - Drugstore.com
  - Healthcentral.com
  - Novartis.com
- Factor 2: Four Variants (or treatments)
  - Original natural language policy
  - List of privacy goals and privacy vulnerabilities
  - Categorical representation based on the taxonomy
  - Original natural language policy supplemented with highlighted privacy goals and vulnerabilities

Variant #1:
Original NL privacy policy

When you place an order, we will ask you to set up "your account," which includes your name, e-mail address, mailing address, credit card number and expiration date, as well as certain other information when you order prescriptions. Using your account information, we will send you communications that we believe are relevant to you, including eMedalert, prescription refill and renewal reminders, newsletters or emails. If you prefer not to receive optional email or other communication from us, you may adjust your account to prevent such communications. If we receive updated account information from our shippers or other third parties, we may revise your account for you so that we can efficiently process your orders, deliver your packages or otherwise communicate with you. If you would like to review or revise the information we have in your account, you may access such information by clicking on the "your account" tab on any screen.

Variant #2:
List of goals and vulnerabilities

- COLLECT PII when placing an order
- USE PII to offer products/services
- OPT-OUT from receiving emails from our company
- UPDATE PII automatically using information received from 3rd parties
- ALLOW customer to modify/remove their PII

Variant #3: Categorical list

Privacy Policy

Access/Participation

This category contains policies relevant to denying access to pages or services if customers do not provide their PII

Choice/Consent

This category outlines ways users have control over how what information is collected from them and whether the information can be transferred to others.

Contact

This category outlines how and for what purposes organizations use customer PII to contact them.

Variant #3: Categorical list

Privacy Policy

Access/Par

This categor
services if cu

Choice/Con

This categor
is collected f
others.

Contact

This categor
customer PII

Choice/Consent

Definitions:

PHI - PHI stands for Personal Health Information. This includes any information that is related to one's medical history such as prescriptions, family illnesses, past treatments, current treatments, etc.

BrandX's Choice/Consent policies:

- We will disclose PHI at request of patient
- Allow consumers to opt-out from receiving emails from our company
- Allow customers to opt-out from sharing website usage information with 3rd parties
- Allow customers to opt-out of sharing information with 3rd parties

Back to the Categories

Variant #4:
Policy with G/V highlights

When you place an order, we will ask you to set up "your account," which includes your name, e-mail address, mailing address, credit card number and expiration date, as well as certain other information when you order prescriptions. Using your account information, we will send you communications that we believe are relevant to you, including eMedalert[illegible] reminders, newsletters or emails. If you do not wish to receive optional email or other communication from us, you may adjust your account to prevent such communications. If we receive updated account information from our shippers or other third parties, we may revise your account for you so that we can efficiently process your orders, deliver your packages or otherwise communicate with you. If you would like to review or revise the information we have in your account, you may access such information by clicking on the "your account" tab on any screen.

USE PII to offer products/ services

“I read the entire set of privacy policies of the website”

- Categories 62%
- Policy 56%
- Goals/Vulnerabilities & Policy 50%
- Goals/Vulnerabilities 44%

Average Comprehension Score
(only respondents who read the entire policy)

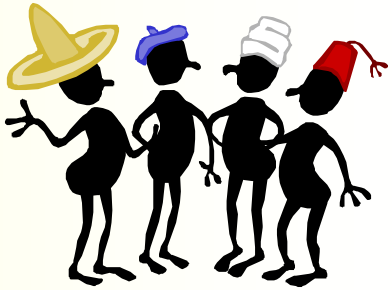
Variant	Average
Categorical	**64.22**
Goals/Vulnerabilities	**55.46**
Goals/Vulnerabilities & Original Policy	49.38
Original Policy	40.00
Average	52.88

Finding: User perception and comprehension are misaligned.

Summary of User Perceptions

- Users believe:
  - they are more secure sharing PII with websites that display NL policies that highlight the goals and vulnerabilities
  - the companies that display NL policies that highlight goals and vulnerabilities will protect their information the most
  - the two NL variants are explained more thoroughly than alternative expressions
- User perception and comprehension are misaligned!!
  - Users feel most secure and protected by natural language policies, but comprehend them the worst

What about demographics?



- No correlation between demographic factors and comprehension/perception exists.

- Exception
  - Respondents age 57 and higher scored lower on comprehension questions

Problems with Privacy Policies

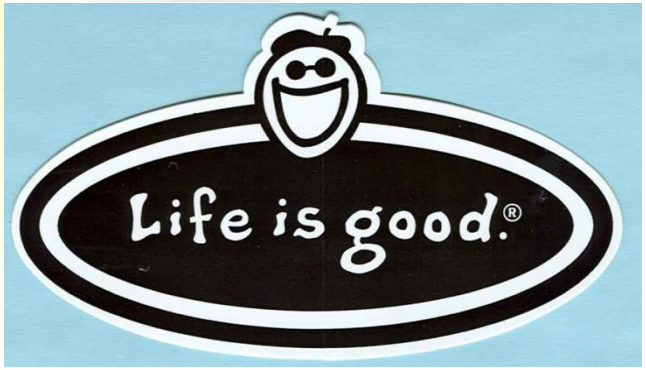
- Difficult to understand
- Information of little interest
- Ambiguous and “warm & fuzzy” phrases
- What exactly does the policy apply to?
- Website controlled appearance
- Too much information
- Too little information
- Organizations don’t always comply with relevant legislation
- **Organizations don’t always keep their stated promises!**

Supporting Policy Compliance

J.D. Young, "Commitment Analysis to Operationalize Software Requirements from Privacy Notices," to appear Requirements Engineering Journal, 2010.

J.D. Young and Annie I. Antón, "A Method for Identifying Software Requirements Based on Policy Commitments," 18th International IEEE Requirements Engineering Conference, 2010.

Do Their Policy Statements Reflect Actual Business Practices?



The Problem

- Organizations express ideas that can be legally-binding

- Requirements engineers must specify these as software requirements

- How can we obtain operationalizable and policy-compliant software requirements from documents such as a company privacy policy?

Goal Based Approach: Privacy Taxonomy

- **Privacy Protection Goals**
  - Access/Participation
  - Choice/Consent
  - Enforcement/Redress
  - Integrity/Security
  - Notice/Awareness

- **Privacy Vulnerabilities**
  - Information Aggregation
  - Information Collection
  - Information Monitoring
  - Personalization
  - Solicitation
  - Information Storage
  - Information Transfer

A Requirements Taxonomy to Reduce Website Privacy Vulnerabilities,
Annie I. Antón and Julia B. Earp.
Requirements Engineering Journal, 2004.

A New Approach

- Formative study
  - Four healthcare organizations
  - Seventeen policy documents

- Organizations express **commitments**, **privileges**, and **rights** that can be legally-binding

- Three Steps
  - Step 1: Parse
  - Step 2: Classify
  - Step 3: Operationalize

Step 1: Parse into Individual Statements

- Dossia Example Policy Statement:
  - Unless you explicitly and specifically consent, Dossia will not disclose your health information or contact information to third parties for them to use for marketing purposes.

Step 2: Classification

- 12 Classifications

- Procedural or Legal?
- External or Internal?
- Commitment, Privilege or Right?

Step 3: Operationalize Classified Statements into Requirements

- Commitment

  The system shall require the [actor] **to** [action] [object] **from** [object's source] **to/with** [target] **for/in order to** [purpose] **given/if** [conditions].

- Privilege/Right

  The system shall allow the [actor] **to** [action] [object] **from** [object's source] **to/with** [target] **for/in order to** [purpose] **given/if** [conditions].

Operationalizing Example from Dossia

The system shall require the organization to disclose user's (users') health information or contact information to third parties for third parties to use for marketing purposes only if user(s) explicitly and specifically consent(s).

Case Studies

- Health Insurance

- Online Drugstore

- Pharmaceuticals

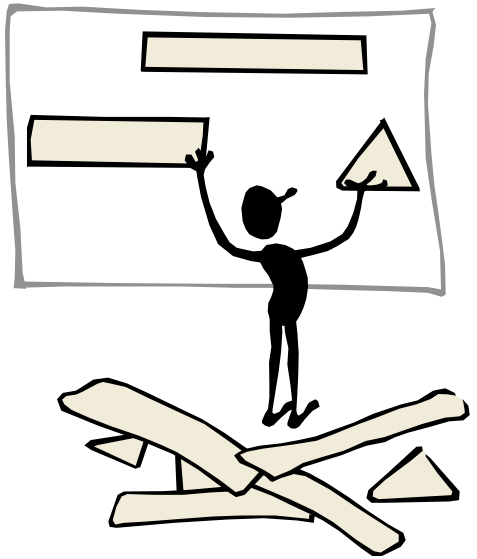
- Personal Health Record (PHR)

Case Studies

- 92.27% of the policy statements are *procedural*, whereas only 7.73% *are legal*
- Top two most used classifications
  - *procedural internal commitment* (36.3%)
  - *procedural internal privilege* (26.6%)

Still Working…

- Comparison with other approaches for extracting requirements from policy
- Generalizability beyond healthcare
- User study to determine effectiveness
- Tools to reduce the manual work



Summary

- Readability of privacy policies is poor
- Alignment of policies / practices to people is poor
- System compliance with legislation is not certain
- System compliance with policies is not certain