Analyzing Privacy Policies using the Privacy by Design Framework

PRIVACYCON FEDERAL TRADE COMMISSION DC // 1.12.17

FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

Fer O'Neil
PhD student, Texas Tech
Technical Writer, ESET

https://www.linkedin.com/in/feroneil
fer.oneil@ttu.edu | fer.oneil@eset.com

TEXAS TECH UNIVERSITY

eset

Introduction

The results of this project provide a first look at applying a Privacy By Design (PbD) framework to analyze privacy policies. I examined the privacy policies of the top 10 most trusted companies for privacy to analyze whether the language they use to communicate their privacy policies conforms to the Foundational Principles of the PbD standard. The most important principle of PbD is "to keep it user-centric." That is, privacy policies exist to communicate how a person's information and data is collected, handled, and used by the companies that accumulate this information.

Methods

Using data analysis software (QDA Miner), I assigned codes to each privacy policy and then analyzed the frequency of categories and codes. The three steps included the following:

1. Categorization and Coding
2. Frequency Analysis
3. Cluster and Correspondence Analysis

The initial coding used the sections and headings present within the policies to create the initial categories and codes. Once words and phrases contained in the document were categorized and subjected to the consistency check, a frequency analysis was run at the Category level to determine how often policies referred to the coded topics. From the frequency analysis, I examined whether the most widely addressed category also contained the most words.

Category	PbD Principle
What Personal Information is Collected	PbD-1
How Information is Used	PbD-6 PbD-2
How Information is Kept Secure	-
What Choices Do I Have	PbD-2 PbD-7
Minors	-
Privacy Complaints	-

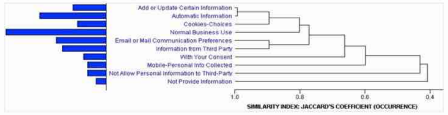
Privacy By Design (PbD)

PbD advances the view that we cannot assure the future of privacy solely by compliance with legislation and regulatory frameworks; rather, privacy assurance must become an organization's default mode of operation. This project focuses on the communicative and end-user perspective because it is more than just an engineering guideline, and we must make this approach to "avoid falling into techno-centric solutions to a socio-technical problem" (Gürses, Troncoso, and Diaz 2011, 5). The four PbD principles discussed in this project are displayed in the table to the right:

PbD Principle	Description of Principle
Proactive not Reactive; Preventative not Remedial (PbD-1)	The Privacy by Design approach is characterized by proactive rather than reactive measures. It anticipates and prevents privacy-invasive events before they happen.
Privacy as the Default Setting (PbD-2)	Privacy by Design seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice.
Visibility and Transparency – Keep it Open (PbD-6)	Privacy by Design seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to independent verification.
Respect for User Privacy – Keep it User-Centric (PbD-7)	Privacy by Design requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options.

Results and Analysis

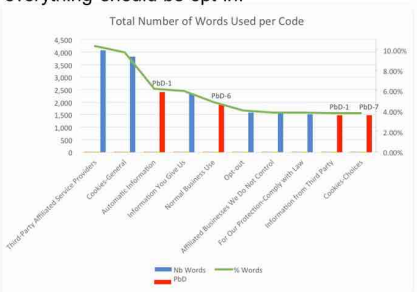
PbD Principle	Frequency of Occurrence	% of Total
PbD-1	62	12.90%
PbD-2	30	6.00%
PbD-6	48	10.00%
PbD-7	84	17.50%



PbD advocates for keeping the interests of the individual central. The following four codes are those that we can identify as conforming to the PbD principles for user-centric policies:

- Add or Update Certain Information (PbD-2, PbD-7)
- Email or Mail Communication Preferences (PbD-7)
- Not Allow Personal Information to Third-Party (PbD-2, PbD-7)
- Cookies-Choices (PbD-7)

PbD-7 may be the most applicable for performing a future content analysis on the specific categories and codes that apply to users. In this case (see the figure below), the highest percentage of words are within non-user-centric codes (Cookies-General). Cookies-Choices is the first user-centric principle, but this is the de-facto "token" user-centric principle in most privacy policies. Opt-out is not user-centric because PbD states that everything should be opt-in.



As the figure above displays, there are more words used in the business-focused codes (blue), which could indicate a preference for obtaining the data associated with the code, or a compliance reason for communicating about it.

Conclusion

This project contributes to a better understanding of what information is included in the privacy policies of the most trusted companies (Ponemon Institute 2015) because the results do suggest that the percent of codes closely correlates to the number of and percent of total cases that the code appears in. That is, the more times a coding category appears in total throughout all the policies, the more likely it is to appear in all policies. Proceeding with the preceding intimation, when combined with the PbD framework we can start to make recommendations for what is included, and what is missing as well.

Selected References

1. 7 Foundational Principles." 2015. Privacy By Design. https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf
2. Ponemon Institute Announces Results of 2014 Most Trusted Companies for Privacy Study http://www.ponemon.org/blog/ponemon-institute-announces-results-of-2014-most-trusted-companies-for-privacy-study
3. Gürses, Seda, Carmela Troncoso, and Claudia Diaz. 2011. "Engineering Privacy by Design." Computers, Privacy & Data Protection 14.