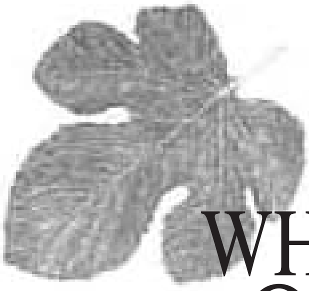
By *Irene Pollach*

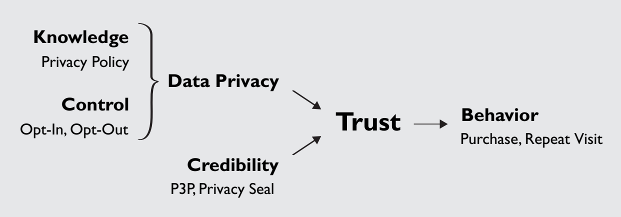


WHAT'S WRONG WITH ONLINE PRIVACY POLICIES?

Research has shown that privacy policies tend to intensify privacy concerns rather than engender trust. One way to combat this dichotomy is to redesign their content, language, and presentation format.

As technologies available for collection and analysis of Web data have become more elaborate, data privacy concerns among Internet users have grown. In particular, they worry that Web merchants sell customer data to third parties, clog their mailboxes with unsolicited email, place persistent cookies on their PCs or enable third parties to do so [10]. To protect their privacy, users abort transactions, falsify personal details, or maintain several email accounts. Such practices deprive Web merchants of information critical to meeting customer needs and sustaining a competitive advantage [3].

To encourage users to participate in online transactions, Web merchants must ease people's concerns about data misuse. To earn users' trust, a Web site should make it explicit that customer data is treated in a fair and responsible manner [4]. It has therefore become common practice for Web merchants to post privacy policies on their Web sites to inform users about data handling practices. Another factor conducive to user trust is the level of control users have over their data by means of opt-in or opt-out facilities [7].



Privacy and trust-building on the Internet.

For this knowledge and control to engender trust, it is crucial that users perceive an organization making commitments to user privacy as credible [11]. To achieve this, companies supplement their privacy policies with privacy seals or make their Web sites P3P-compliant [12]. The more trust users have in a Web site, the more likely they are to buy from the site, visit it again, or recommend it to others [8] (as illustrated in the accompanying figure).

Previous research has found that U.S. online privacy policies do not address those areas of data handling that concern users. Rather, these documents are written in a manner that protects companies against privacy lawsuits by integrating privacy legislation that regulates, for example, information gathered from children (Children's Online Privacy Protection Act), financial data (Gramm-Leach-Bliley Act), and medical records (Health Insurance Portability and Accountability Act), or state legislation such as California's Online Privacy Protection Act [5]. In addition, Internet users have been found not to read online privacy policies because they find them too legalistic and therefore difficult to understand [9]. Another study has assessed privacy policies by means of readability formulae and found that readers would require at least some college education to understand the complex words and sentence structures in these texts [1]. By discouraging users from reading policies, companies forego the opportunity to ease privacy concerns and build trust.

Data Collection	Collection and storage of personally identifiable information (PII); collection of aggregate information; users' ability to view and update data profiles; collection of user data via surveys; sweepstakes used to gather customer data; obtaining user information from other sources; storage and usage of email addresses from inquiries; cookies; information on disablement of cookies; information on consequences of disabling cookies; Web beacons;
Third-Party Data Collection	Types of data collected by third parties; third-party cookies or Web beacons; privacy agreement with third parties collecting data; opt-out of third-party data collection;
Data Storage	Measures taken to ensure secure offline storage of data; measures taken to prevent unauthorized employee access; users' ability to delete PII; records of PII kept after user deletes PII;
Data Sharing	Privacy agreements with business agents receiving PII; sharing of aggregate information with affiliates; sharing of PII with affiliates; sharing of aggregate information with third parties other than business agents; sharing of PII with third parties other than business agents; selling of data; sharing of email addresses; sharing of data obtained in sweepstakes/surveys;
Marketing Communication	Unsolicited email; unsolicited email from third parties;

Table 1. Summary of questions.

Identifying Weaknesses in Privacy Policies

Since the manner in which a company communicates its data handling practices to Internet users has a bearing on its success in e-commerce, this article sheds light on why privacy policies fail to communicate data handling practices effectively. I conducted two separate studies, one examining *what* these documents say or do not say about data handling practices, and the other focusing on *how* companies describe their data handling practices in these documents. The purpose of these studies was not to describe the current state of privacy policies but to identify weaknesses and make suggestions for improvements.

The sample chosen was therefore not intended to be representative of commercial Web sites. Rather, 50 Web sites covering a broad spectrum of business models were chosen on the basis of their commercial success, since successful e-commerce sites may serve as lead innovators for other Web sites. Their success was determined using Alexa.com traffic rankings, a ranking of online retailers in *Store* magazine, and articles from the business press. The privacy policies were collected from the following commercial Web sites:

- ***Retailers:*** 1-800-flowers, Amazon, Apple Store, Barnes & Noble, Best Buy, BMG Music, Buy.com, Circuit City, Cyberian Outpost, Dell, eBay, eToys, Gap, Gateway, Home Depot, JC Penney, Lands' End, L.L. Bean, Office Depot, QVC, Sears, Staples, Target, Ticketmaster, uBid, Wal-Mart
- ***Internet service providers:*** About.com, AlltheWeb.com, AOL, Earthlink, Excite, Hotmail, Lycos, Netscape, Prodigy, USA.net, Yahoo
- ***News sites:*** *Economist, Fortune, Investor's Business Daily, Los Angeles Times, New York Times, Wall Street Journal, Washington Post*
- ***Travel agents:*** American Express, Expedia, Hotels.com, Orbitz, Priceline, Travelocity

2a.

Category	Questions	All Companies	Seal Companies
Data Collection	11	27.6%	24.9%
Third-Party Data Collection	4	48.0%	47.4%
Data Storage	4	75.0%	68.4%
Data Sharing	8	37.5%	27.6%
Marketing Communication	2	24.0%	18.4%
Total	29	39.4%	34.3%

2b.

	Aggregate Data		Personal Data	
	Affiliates	Third parties	Affiliates	Third parties
yes	34%	62%	42%	6%
no	-	-	10%	42%
if authorized	-	-	-	38%
No answer	66%	38%	48%	12%

Table 2a. Proportion of questions unanswered.
Table 2b. Data sharing with affiliates and third parties.

Of the 50 companies, 19 displayed at least one privacy seal at the time of data collection. Varying in length from 575 to 6,139 words, the privacy policies resulted in a text corpus of 108,570 words.

Coverage of Privacy Issues

The scope and depth of content was assessed by trying to answer 29 questions on corporate data handling using the information provided in each privacy policy (see Table 1). These questions pertained to the key privacy concerns among Internet users: data collection, data storage, data sharing, and unsolicited marketing communications [5, 10]. The questions resulted from an inductive pilot coding of the five longest policies in the sample. They were then tested on the five longest policies in the remaining sample and amended to ensure the questions encompass all data handling practices commonly engaged by companies. These questions were factual codes, intended to condense precisely defined facts rather than represent the entire content of a document. When answering these questions, the "at-least-some" rule was applied, which considers a practice true even if it is carried out only occasionally. To ensure the reliability of the results, the 50 privacy policies were coded twice, resulting in an intra-coder reliability of 98.84%.

Overall, 39.4% of these questions could not be answered because the policies did not contain sufficient information. Table 2a gives a breakdown of these questions by category, showing the results obtained for all 50 companies and for those 19 companies among them that display privacy seals. The high proportion of unanswered questions pertaining to third-party data collection can be attributed to the fact that not every company allows third parties to collect data and therefore does not mention it in its policy. However, the high proportion of companies disclosing insufficient information about data collection, storage, sharing, and spam email clearly shows that privacy policies do not cover data handling practices in a satisfactory manner. Users cannot be sure whether companies do not engage in such practices or simply fail to mention that they do. It is also worth noting that the results for those companies displaying a privacy seal are only slightly better than those of the total sample, suggesting that privacy seals are no guarantee for comprehensive privacy policies.

Data storage stands out as one area of insufficient disclosure among both the total sample and those companies displaying privacy seals. While no company said it did not take steps to ensure secure offline storage and prevent unauthorized employee access or that users cannot delete their personal information, the level of disclosure on these aspects of data storage was never higher than 32% for the total sample. These results call for more detailed disclosure of data storage practices and users' control over data stored about them.

Since data sharing is one of the most prevalent concerns among Internet users [10], the coverage of this practice was examined in more detail. Table 2b indicates whether or not the companies share aggregate or personally identifiable information (PII) with either affiliated companies or third parties other than business agents. The proportion of companies providing no relevant information is alarming, particularly regarding the sharing of aggregate and PII with affiliates. The high percentage of companies admitting to sharing personal data with affiliates is also worth noting, considering that affiliates may maintain completely different privacy policies.

Companies also admit to practices such as selling user data, sharing data obtained through sweepstakes and contests, allowing third parties to collect data, sending unsolicited communications to registered users, or sharing email addresses with third parties. This makes it all the more important that users read privacy policies to become aware of what can happen to their data and to be able to make an informed decision as to whether or not they want to disclose personal information on a Web site.

Language Use

The analysis of the language of privacy policies was based on critical linguistics [6], a method that seeks to uncover how authors of texts use language to construct their own versions of reality. In the context of privacy policies, this "version of reality" refers to how companies present their data handling practices to their readers. The goal of this analysis was to determine why privacy policies are difficult to understand and why readers do not consider them worth reading [1, 9]. Among the parameters put forward in critical linguistics, the following four were suitable for the analysis of privacy policies:

- *Lexical Choice.* Looking at the systematic use or avoidance of words;
- *Syntactical Transformation.* Exploring the use of passive voice and nominalizations;
- *Negation.* Examining which issues are denied; and
- *Modality.* Assessing the certainty of the speaker about the content of an utterance.

Lexical Choice. The analysis of the vocabulary has revealed that companies sugar-coat data handling practices by foregrounding positive aspects and backgrounding privacy invasions. These enhancements of data handling practices occur, for example, when companies claim the email messages they send to registered users are of "interest to them" or that the parties they share information with are "carefully selected." Also, companies choose verbs that exclude themselves in order to remove themselves from statements disclosing unethical practices. For example, they state that *you receive* unsolicited email messages instead of *we send* them.

Lexical choice also plays a role when companies talk about opt-in/opt-out facilities for certain practices. The framing of opt-in or opt-out messages has been found to influence people's privacy preferences [2]. In the policies examined, companies use phrases such as o*nly when authorized, if you authorize us,* or *not without your permission* to describe practices relating to unsolicited commercial email. However, these lexical choices do not make it clear whether this authorization is the result of opting-in or not opting-out. Users may thus not be aware they have given authorization to a company by not opting out.

Syntactical Transformations. These were found in connection with data sharing, when companies avoid using *we share* but use nouns (*the sharing*) or switch to passive structures instead in order to distance themselves from such practices. For example: "We want you to know about the personal information *we collect*, how *we use* that information and with whom *it* may *be shared*." Since such passive structures do not make it explicit who is responsible for an action, it seems that companies try to de-emphasize the fact they share information with third parties.

PATTERN	PURPOSE	TEXTUAL REALIZATIONS	SEAL	NON-SEAL
Mitigation	Downplaying Frequency	*occasional(ly), from time to time, sometimes, at times*	2.11	2.29
Enhancement	Emphasizing Qualities	data sharing: *trustworthy, reputable, carefully selected/screened* (third parties)	0.58	0.71
		spam: *of interest/value to you*	1.05	0.90
Obfuscation	Hedging Claims	*may, might, perhaps, in/at our discretion, except as, on a limited basis, we reserve the right to, including but not limited to*	22.37	19.68
	Obscuring Causality	*if you authorize us, only when authorized, (not) with(out) your consent/permission/knowledge, when we have your permission, not … unless you give us the permission to do so*	1.89	1.16
Omission	Removing Agents	*the sharing of, is shared* (rather than *we share*), *you receive* (rather than *we send*)	0.74	0.58

Table 3. Ambiguity in privacy policies.

Negation. To deny certain practices, negations were used frequently throughout the corpus, but not frequently enough, as the content analysis here has revealed. *Not*, for example, is the ninth most frequent word in the corpus, not counting grammatical words such as articles, prepositions, and pronouns. Although negative statements are generally more difficult to process for humans than positive ones, explicitly stating that a certain practice is not carried out is indispensable in easing users' privacy concerns. Negation is also used together with rhetoric hedges as in *except as otherwise stated we do not [...]*. Such phrases give carte blanche to the company to engage in any practice not expressly ruled out but provide little information about what actually happens with user data.

Modality. Essentially, modal verbs and adverbs make sentences vague. The corpus of privacy policies contains 948 instances of *may* and 123 instances of *might, perhaps, sometimes, occasional(ly), and from time to time*, all of which are instances of modality. May is, in fact, the fourth most frequent non-grammatical word in the corpus topped only by *information, use,* and *site*. These modality markers downplay the frequency with which companies carry out certain data handling practices. In addition, legal expressions such as *we reserve the right to* are reflections of modality, allowing several interpretations as to whether these practices are carried out or not.

The rhetorical features that emerged from this

THESE FINDINGS SUGGEST THAT ONLINE PRIVACY POLICIES HAVE BEEN DRAFTED WITH THE THREAT OF PRIVACY LITIGATIONS IN MIND RATHER THAN COMMITMENT TO FAIR DATA HANDLING PRACTICES.

analysis were grouped into four broad patterns (see Table 3), including mitigation, enhancement, obfuscation, and omission. Two sample sentences containing these patterns illustrate how companies use language to construct their own privacy realities:

- Circuit City Stores, Inc. (including its subsidiaries) "*from time to time may* also provide names, addresses or email addresses to strategic partners who *have* information, products or services that *may* be of *interest to you*." (Mitigation, Obfuscation, Enhancement)
- "Established members will *occasionally receive* information on products, services, special deals, and a newsletter." (Mitigation, Omission)

The main goal of the linguistic analysis was to identify realities created through language rather than to produce quantitative indices of language. However, these are necessary to examine differences in communicative quality between companies with privacy seals and those without seals. Table 3 compares the average number of instances of each pattern in the privacy policies of both types of companies, capturing occurrences of the textual realizations listed in the table as well as similarly worded phrases. The results indicate that the privacy policies of companies with seals are by no means less ambiguous than those of companies without seals, suggesting that compliance with a seal program impacts content but not language.

Implications

The findings noted here suggest that online privacy policies have been drafted with the threat of privacy litigations in mind rather than commitment to fair data handling practices. The content analysis has revealed that these documents fail to address important areas of user concern. We do not know whether companies simply do not mention practices they do not engage in or whether they abuse their knowledge about and control over user data by deliberately withholding information. Users would have more trust in a company's Web site if they can learn from a privacy statement not only what the company does with user data but also what it does not do. Thus, when drafting privacy policies, companies should focus not just on their own practices but also take into account the wider context of data handling on the Internet and address practices they do not engage in as well.

The linguistic analysis has shown that companies obscure privacy infringements by downplaying their frequency, mitigating or enhancing questionable practices, and omitting references to themselves when they talk about unethical data handling practices. One cannot safely say whether these rhetoric patterns are merely the chance product of poor writing skills or whether they are a manifestation of strategic ambiguity aimed at deceiving and confusing readers. At any rate, IS managers must be aware of the effects vague language has on readers and should tailor their privacy policies better to Internet users' information needs by representing data handling practices in a more accurate manner.

Changes are needed not only in the content and language of privacy policies but, most importantly, also in their presentation format. Tables would be a more suitable vehicle than narrative text, as they make content deficiencies evident right away and eliminate the problem of ambiguous language altogether. eBay, for example, posts a static chart summarizing parts of its text-based policy as an appendix, but does not exploit it to its fullest potential.

A more effective solution would be to present different types of data (for example, sales data, data from surveys, data from sweepstakes, click-stream data, and so on) and data handling methods (collecting, storing, sharing, selling, sending emails, and so on) in a matrix with each cell being clickable and leading the user to a plain-language explanation of when this data handling practice is carried out for this specific type of data. Chopping the information into manageable chunks and letting users decide which parts they want to read would make privacy policies more reader-friendly. This would, for example, spare users the need to read through the entire document just to check whether a company shares email addresses.

Certainly, the narrative privacy policies cannot be eliminated altogether, as they protect businesses if privacy litigations are brought against them, but more reader-friendly alternatives to conventional privacy policies should be offered to prevent poor writing skills or strategic ambiguity from undermining user trust. It is upon IS managers and system designers to take a proactive stance and let Internet users have the knowledge and control they need to make informed decisions about their personal data.

References

1. Antón, A.I., Earp, J.B., He, Q., Stufflebeam, W., Bolchini, D., and Jensen, C. The lack of clarity in financial privacy policies and the need for standardization. *IEEE Security and Privacy 2*, 2 (Mar. 2004), 36–45.
2. Bellman, S., Johnson, E.J., and Lohse, G.E. To opt-in or opt-out? It depends on the question. *Commun. ACM 44*, 2 (Feb. 2001), 25–27.
3. Brown, M., and Muchira, R. Investigating the relationship between Internet privacy concerns and online purchase behavior. *J. Electronic Commerce Research 5*, 1 (2004), 62–70.
4. Culnan, M.J., and Armstrong, P.K. Information privacy concerns, procedural fairness, and impersonal trust. An empirical investigation. *Organization Science 10*, 1 (Jan. 1999), 104–115.
5. Earp, J.B., Antón, A.I., Aiman-Smith, L., and Stufflebeam, W.H. Examining Internet privacy policies within the context of user privacy values. *IEEE Trans. Engineering Management 52,* 2 (May 2005), 227–237.
6. Fowler, R., Hodge B., Kress, G., and Trew, T. *Language and Control.* Routledge, London, 1979.
7. Hoffman, D.L., Novak, T.P., and Peralta, M. Building consumer trust online. *Commun. ACM 42*, 4 (Apr. 1999), 80–85.
8. Liu, C., Marchewka, J.T., Lu, J., and Yu, C.-S. Beyond concern: A privacy-trust-behavioral intention model of electronic commerce. *Information & Management 42*, 1 (Jan. 2004), 127–142.
9. Milne, G.R., and Culnan, M.J. Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *J. Interactive Marketing 18*, 3 (summer 2004), 15–29.
10. Miyazaki, A.D., and Fernandez, A. Internet privacy and security: An examination of online retailer disclosures. *J. Public Policy & Marketing 19*, 1 (spring 2000), 54–61.
11. Olivero, N., and Lunt, P. Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control. *J. Economic Psychology 25*, 2 (Apr. 2004), 243–262.
12. Turner, E.C., and Dasgupta, S. Privacy on the Web: An examination of user concerns, technology, and implications for business organizations and individuals. *Information Systems Management*, (winter 2003), 8–18.

Irene Pollach (ipollach@wu-wien.ac.at) is an assistant professor of business communication at the Vienna University of Economics and Business Administration in Austria.