IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4928

Online Privacy Statements: Are they Worth Reading?

Irene Pollach
Vienna University of Economics and Business Administration, Augasse 9, A-1090 Vienna, Australia, irene.pollach@wu-wien.ac.at

INTRODUCTION

Internet users have become increasingly concerned about data collection and data handling practices of Web sites. This issue has been explored in academic studies, focusing on the nature of users' privacy concerns (Hoffman, Novak, & Peralta, 1999; Cranor, Reagle, & Ackerman, 1999; Sheehan & Hoy, 2000), their awareness of privacy issues (Dommeyer & Gross, 2003), their willingness to provide information (Phelps, Nowak, & Ferrell, 2000), and the determinants of consumer trust (Schoenbachler & Gordon, 2002). To ease users' fears about data misuse, privacy policies have become *de rigueur* among U.S. commercial Web sites over the past couple of years (Messmer, 1997; Federal Trade Commission, 1998; Culnan, 1999; Federal Trade Commission, 2000; Liu & Arnett, 2002). Culnan and Milne (2001) found that the majority of Internet users do not read online privacy policies, primarily because they are too long, difficult to understand, and "all the same", as one survey respondent put it. The Platform for Privacy Preferences (P3P) and its XML-based, machine-readable privacy specifications may be a good alternative to traditional privacy policies, but as of 2003 only 10% of Web sites have P3P-based privacy policies (Cranor, Byers, & Kormann, 2003). Reasons for its slow adoption include unresolved legal issues and the fact that companies find it hard to squeeze their complicated privacy policies into the more straightforward P3P scheme (Thibodeau, 2002). Thus, privacy policies still have to be read by humans rather than Web browsers, and so their quality is critical to users' trust in Web sites. The present study examines 50 privacy statements from commercial Web sites in greater detail than the previous studies (Liu & Arnett, 2002; Miyazaki & Fernandez, 2000; Johnson-Page & Thatcher, 2001), focusing not just on the issues they address but also on the specific practices companies have adopted to collect, use and share customer data. The objective of this study is to determine how accurately privacy statements communicate data-handling practices, to what extent these practices respect user privacy, and whether companies displaying privacy seals handle user data more responsibly than companies without such a seal.

METHODOLOGY

Sampling

Since small, systematically selected samples are considered more interesting for content analyses than larger convenience samples (Bauer, 2000), only 50 Web sites were selected for the analysis (see Appendix). They were chosen on the basis of their commercial success, as successful Web sites were assumed to serve as lead innovators for other Web sites. All Web sites included in the sample disclose their privacy policies and collect personally identifiable information (PII) when users register with the site or place orders. The sample companies fall into four different categories (retail, Internet services, news, travel) and reflect a variety of online business models, including pure e-brands and offline brands with online outlets; companies selling physical goods and companies selling digital goods and services; and B2C stores and C2C auctions.

Methods of Analysis

To systematically analyze how data handling practices are described in privacy policies, a content analysis of the 50 privacy statements was conducted. The corpus of the 50 documents contains 108,570 words, with document lengths varying from 575 to 6,139 words. In a pilot study of the four longest privacy policies in the sample, 35 coding categories were developed based on an inductive strategy. Due to lack of space, the coding sheet is not reproduced in this publication. After a pilot coding of 5 other privacy statements and subsequent amendments to the coding sheet, all 50 privacy statements were coded by the author in July 2003. The coding was based on the "at-least-some" rule, which considers practices true even if they are carried out only occasionally. As a check on intra-coder reliability, all texts were recoded in August 2003. The agreement between the two codings was 98.71%.

Since content analysis should not only be concerned with what a text is about but also with its vocabulary (Bauer, 2000), a word frequency list of the total text corpus was created to identify words of potential interest (Krippendorff, 1980). The initial list of 4,016 different words was lemmatized automatically to remove inflectional suffixes (e.g. plural endings, *-ing* forms) and contracted verb forms (Stubbs, 2001). Further, grammatical words (e.g. auxiliary verbs, articles, prepositions), numbers, and names (e.g. for cities, companies, months, domains) were excluded. The resultant list of 1,637 content words, representing 52.15% of the running words, was used for the analysis.

RESULTS OF THE CONTENT ANALYSIS

General Characteristics

On average, a privacy policy consists of 1.12 Web pages and 2,157 words. 90% of the privacy policies are accessible with just one click on a link on the bottom of the Web site's home page. 86% of the companies point out that their privacy policies are subject to change, but only 20% also indicate that they will post a notice on the Web site prior to the change, and 14% will also notify users by e-mail, if users have opted to receive such a notification or if the changes to the policy are significant. The majority of companies (62%) do not display or refer to any privacy seal, 30% have one seal, and 8% have two seals. The seals used are TRUSTe (n=11), BBB Online (n=8), and AOL (n=4).

Data Collection

Clearly, the collection of personally identifiable information is necessary to complete business transactions and is thus gathered by all companies. All companies state explicitly or implicitly that they collect aggregate user data: 92% admit to using cookies and collecting aggregate user data; 6% use cookies but do not point out that they collect aggregate data with them, and 2% collect aggregate user information but fail to mention that they use cookies; 36% of the companies use Web beacons in addition to cookies. Of those companies mentioning cookies (n=49), 71.43% point out that cookies could be disabled. On 5.71% of these sites, users can enjoy complete access to all areas of the Web site without accepting cookies; on 45.71% users will not be able to use certain features of the site; on 20% users can still shop; and on 17.14% users cannot shop without cookies. 22% of the companies use online surveys, 12% use sweepstakes or contests and 40% use both in order to gather additional customer data, especially demographic information. In addition 36% of the companies obtain customer information from unspecified "other sources" and match them with the data they have collected themselves. Another way of collecting customer data is storing e-mail addresses from customer inquiries: 18% of the companies admit to doing so, while 12% say explicitly that they do not do so, and the remaining 70% provide no information on this.

Third-Party Data Collection

The content analysis further included four codes on third-party collection, focusing on what kinds of data third parties are allowed to collect, what collection methods they use, whether they are bound to a privacy agreement, and whether users can opt out of third-party data collection. 64% of the companies say that they let third parties, e.g. advertising networks, collect either aggregate information (55.88%) or both aggregate and personally identifiable information (5.88%). 32% let third parties collect data by means of cookies and Web beacons, 26% only through cookies, and 6% only through Web beacons. Of those companies enabling third-party collection (n=34), 5.88% say that the third parties are bound to a privacy agreement, 44.12% said that they are not bound, and the remaining 50% do not provide any information on this. 61.67% point to the possibility that users may opt out of third-party data collection, but typically referring users only to the Web site of the third party collecting the data. The results for the sub-sample of companies which display at least one privacy seal (n=19), henceforth "seal companies", echo those for the total sample, with 13 (68.42%) companies enabling third-party collection. Of these 13 companies, 53.85% provide information on opting out of third-party collection. In addition, 7.69% promise that these third parties are bound, 46.15% state that they are not bound, 46.15% provide no information as to whether these third parties were bound.

Data Use and Data Sharing

Sending commercial e-mails to customers is common practice among the sample companies: 14% of the companies send unsolicited e-mails such as special offers, product updates or newsletters to their customers, 82% do so but offer opt-in or opt-out facilities, and 4% do not disclose any information on this. It was often not clear from the wording of the privacy policies whether opt-in or opt-out is offered, e.g. in phrases like "only when authorized", "with your permission", or "not without your consent". In other cases, companies offer opt-in only for certain types of communications and opt-out for others. Therefore, opt-in and opt-out were treated as one category in the analysis. All seal companies (n=19) provide at least some information on unsolicited marketing communications. 78.95% provide opt-in or opt-out facilities, and the remaining 21.05% send unsolicited e-mails without opt-out facilities.

Most companies (84%) point out that they share data with agents who either assist in completing orders, e.g. delivery companies, or perform other business services, e.g. customer communications or data analysis. While 28% of those companies which address data sharing with business agents (n = 42) do not specify whether these agents are bound, 50% state that they are bound, 4% state that only some agents are bound, and 2% say that they are not bound.

Table 1 shows the companies' data sharing practices of aggregate and personally identifiable information with affiliates and third parties. The percentage of companies providing no information is alarming, particularly regarding the sharing of information with affiliates (66%). As for the sharing of aggregate data, the results obtained for the seal companies (36.8%) closely mirror those obtained for the total sample (34%). The high percentage of companies sharing PII with affiliates (42%) is also noteworthy, considering that affiliates may maintain completely different privacy policies. The fact that relatively more seal companies provide specific information on PII sharing obviously results in a relatively higher number of companies sharing (52.6%) and not sharing data with affiliates (15.8%), compared to the total sample. Similarly, more seal companies (94.7%) provide information on PII sharing with third parties than companies in the total sample (88%). Notably, no seal company admits to sharing PII with third parties without the user's permission.

Another code in the analysis addressed the selling of customer data. 50% of the companies say that they do not sell or rent customer data to third parties, and only one company admits to selling customer data. Interestingly, the percentage of companies promising not to sell PII is higher than that not sharing PII with third parties (42%). As for the seal companies, the percentage of those not selling and those not sharing PII is the same (57.89%).

Table 1: Data Sharing Practices

	AGGREGATE		PERSONAL	
	Affiliates	Third parties	Affiliates	Third parties
yes	34% (36.8%)	62% (63.2%)	42% (52.6%)	6% (0%)
no	-	-	10% (15.8%)	42% (57.9%)
if authorized	-	-	-	38% (36.8%)
not specified	66% (63.2%)	38% (36.8%)	48% (31.6%)	12% (5.3%)

Only 9 companies say that they do not share e-mail addresses with third parties, while 6 admit to doing so and 14 do so only with the user's permission. Of the 19 seal companies, 2 share e-mail addresses and 10 offer opt-in or opt-out facilities, but not a single one claims that it does not share e-mail addresses. As mentioned above, 52% of the companies use sweepstakes as a means of collecting customer data, but only one of them promises not to share these data, while 4 share them, 2 share only aggregate data, 2 share data but notify users, one shares data but gives users the option to opt-out, and one shares aggregate data but gives users the option to opt-out of sharing PII.

RESULTS OF THE COMPUTER-ASSISTED TEXTUAL ANALYSIS

To gain additional insights into how companies communicate privacy practices, the vocabulary used in the privacy statements was examined. The 10 most frequent content words and their counts are: *information* (3,023), *use* (1,339), *site* (1,174), *may* (948), *personal* (902), *privacy* (850), *service* (811), *email* (777), *not* (753), and *will* (740). The first 3 words on the list account for almost 10% of all content words and the first 10 words for as much as 20%. Also, the 70 most frequent content words represent 50.10% of all content words. This suggests that the vocabulary used in privacy statements is rather homogenous, which explains why people consider them to be "all the same" (Culnan & Milne, 2001). *May* and *not* were the only surprising words among the ten most frequent words and were thus examined in more detail.

The modal verb *may* expresses either permission (intrinsic modality) or possibility (extrinsic modality) (Greenbaum & Quirk, 1990). Extrinsic modality adds intermediate degrees to the choice between *yes* and *no* either in terms of probability ('maybe yes, maybe no') or in terms of frequency ('sometimes yes, sometimes no'). Thus, modal verbs like *may* cause a proposition to become arguable (Halliday, 1994), thereby lowering the degree of certainty of a text (Stillar, 1998). This is a strategy for speakers/writers to mitigate negative content in order to make it more acceptable to the audience (Callow, 1998). For a closer inspection of *may*, the words *might* and *will sometimes/occasionally* were also examined in view of their similar meanings. Table 2 shows all co-occurrences of *may*, *might* and *will sometimes*/*occasionally* with verbs related to data handling practices. The results indicate that the policies contain a large number of vague and to some extent ambiguous

Table 2: The Use of Modal Verbs in Privacy Statements

	may	**might**	**will s./o.**	**TOTAL**
collect	52	1	-	
gather	6	-	-	**59**
use	120	3	1	**124**
share	54	2	3	
disclose	27	-	-	
provide	24	-	-	
release	9	-	-	**119**
send	23	3	-	
receive	20	-	7	**53**
TOTAL	**335**	**9**	**11**	**355**

statements. They leave the reader in the unknown as to when and how data are collected, used or shared and whether users receive unsolicited marketing communication. This suggests that the companies use modal verbs strategically to downplay their questionable data handling practices and the frequency with which they occur.

Not and *never* were found to occur most frequently in combination with the verbs *collect* (72 times), *share* (51 times), *use* (44 times), and *sell* (33 times). In general, negative statements suggest that the speaker/writer is "taking issue with the corresponding positive assertions" (Fairclough, 2001, pp. 128). Thus, when companies commit themselves to, for example, not selling or renting customer data, they implicitly contest the charge that they do. It thus seems that companies employ negations to raise the level of certainty and to dispel users' fears about privacy infringements.

IMPLICATIONS AND CONCLUSIONS

Even if privacy policies appear to be "all the same" (Culnan & Milne, 2001), they differ substantially in terms of content, scope and depth. They emphasize some issues, de-emphasize others and hide still others. Companies admit unethical data handling practices such as sharing e-mail addresses and personally identifiable information with third parties, sharing data obtained through sweepstakes, and even selling customer data. Disclosure of data sharing practices and third-party data collection was especially poor. The comparisons of the total sample and the seal companies have shown that the latter admit to the same questionable practices, such as unbound third-party collection or the sharing of PII, suggesting that a privacy seal is no guarantee that Web sites do not infringe upon user privacy. It also seems that sweepstakes are used as loopholes to obtain customer data that are not subject to the company's general privacy principles and may thus be shared or sold.

In view of the diversity of privacy practices identified, privacy statements are definitely worth reading, although they may not always tell readers what they want to know. In fact, when coding the 50 documents, 22.74% of all answers to the questions posed could not be answered. This level was 19.54% for the seal companies, which suggests that all privacy policies sometimes leave users in the unknown as to whether a certain practice is carried out or not. Although the computer-assisted textual analysis and the content analysis have shown that companies also mention things they do *not* do, more transparency is needed in communicating data handling practices. Companies probably do not see the need to mention certain practices if they do not engage in them, but users are more likely to have 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. Clearly, not mentioning something may also be a strategy for concealing unethical practices. The examination of the content words has shown that companies obscure privacy infringements by downplaying their frequency or probability. Thus, more exact lexical choice would also be desirable for privacy statements.

Ultimately, companies should look for more user-friendly alternatives to the narrative presentation format of privacy policies. eBay was the only company among the 50 sample companies that offered such an alternative. In addition to the long version of its privacy policy, the company posts a tabular version and a short version, both of which give users a much better idea of how their data are collected, used and shared. But also these formats give rise to the paradox situation that data may already have been collected before the user has viewed the privacy policy. This calls for a faster adoption of P3P-enabled privacy policies to eliminate conventional privacy policies altogether.

REFERENCES

Bauer, M.W. (2000). Classical content analysis: A review. In M.W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound* (pp. 131-151). London: Sage.

Callow, K. (1998). *Man and message. A guide to meaning-based text analysis*. Boston: University Press of America.

Cranor, L.F., Byers, S., & Kormann, D. (2003). *An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites as of May 2003*. Retrieved August 14, 2003, from http://www.research.att.com/projects/p3p/.

Cranor, L.F., Reagle, J., & Ackerman, M.S. (1999, April 14). Beyond concern: Understanding net users' attitudes about online privacy. *AT&T Labs-Research Technical Report TR 99.4.3*. Retrieved July 10, 2003, from http://www.research.att.com/projects/privacystudy/.

Culnan, M.J. (1999, June). *Georgetown Internet privacy policy survey: Report to the Federal Trade Commission*, Retrieved August 9, 2003, from http://www.msb.edu/faculty/culnanm/gippshome.html.

Culnan, M.J., & Milne, G.R. (2001, December). *The Culnan-Milne survey on consumers & online privacy notices: A summary of responses*. Retrieved July 14, 2003, from http://intra.som.umass.edu/ georgemilne/PDF_Files/culnan-milne.pdf.

Dommeyer, C.J., & Gross, B.L. (2003). What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing, 17*(2), 34-51.

Fairclough, N. (2001). *Language and power*. 2nd ed. London: Longman.

Federal Trade Commission (1998, June). *Privacy online: A report to Congress*. Retrieved July 20, 2003, from http://www.ftc.gov/reports/privacy3/priv-23a.pdf.

Federal Trade Commission (2000, May). *Privacy online: Fair information practices in the electronic marketplace*. Retrieved July 20, 2003, from http://www.ftc.gov/reports/ privacy2000/privacy2000.pdf.

Greenbaum, S., & Quirk, R. (1990). *A student's grammar of the English language*. Harlow: Longman.

Halliday, M.A.K. (1994). *Introduction to functional grammar*. 2nd ed. London: Edward Arnold.

Hoffman, D.L., Novak, T.P., & Peralta, M. (1999, April). Building consumer trust online. *Communications of the ACM*, 80-85.

Johnson-Page, G.F., & Thatcher, R.S. (2001). B2C data privacy policies: Current trends. *Management Decision, 39*(4), 262-271.

Krippendorff, K. (1980). *Content analysis. An introduction to its methodology*. Beverly Hills, CA: Sage.

Liu, C., & Arnett, K.P. (2002). An examination of privacy policies in Fortune 500 Web sites. *Mid-American Journal of Business, 17*(1), 13-21.

Messmer, E. (1997, June 16). Group slams Web sites for lack of privacy policies. *Network World*, 41.

Miyazaki, A.D., & Fernandez, A. (2000). Internet privacy and security: An examination of online retailer disclosures. *Journal of Public Policy & Marketing, 19*(1), 54-61.

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing, 19*(1), 27-41.

Schoenbachler, D.D., & Gordon, G.L. (2002): Trust and customer willingness to provide information in database-driven relationship marketing. *Journal of Interactive Marketing, 16*(3), 2- 16.

Sheehan, K.B., & Hoy, M.G. (2000): Dimensions of online privacy concerns among online consumers. *Journal of Public Policy & Marketing, 19*(1), 62-73.

Stillar, G.F. (1998). *Analyzing everyday texts. Discourse, rhetoric, and social perspectives*. Thousand Oaks: Sage.

Stubbs, M. (2001). *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.

Thibodeau, P. (2002). P3P supporters struggle to increase adoption of data privacy standard. *Computerworld 36*(47), 20.

APPENDIX - SAMPLE COMPANIES

Retail
http://www.1-800-flowers.com/
http://www.amazon.com/
http://store.apple.com/
http://www.bn.com/
http://www.bestbuy.com/
http://www.bmgmusic.com/
http://www.buy.com/
http://www.circuitcity.com/
http://www.outpost.com/
http://www.dell.com/
http://www.ebay.com/
http://www.etoys.com/
http://www.gap.com/
http://www.gateway.com/
http://www.homedepot.com/
http://www.jcpenney.com/
http://www.officedepot.com/
http://www.landsend.com/
http://www.llbean.com/
http://www.qvc.com/
http://www.sears.com/
http://www.staples.com/
http://www.target.com/
http://www.ticketmaster.com/
http://www.ubid.com/
http://www.walmart.com/

Internet
http://www.about.com/
http://www.alltheweb.com/
http://www.aol.com/
http://www.earthlink.net /
http://www.excite.com/
http://www.hotmail.com/
http://www.lycos.com/
http://my.netscpae.com/
http://www.prodigy.net/
http://www.usa.net/
http://www.yahoo.com/

News
http://www.fortune.com/
http://www.investors.com/
http://www.latimes.com/
http://www.economist.com/
http://www.nytimes.com/
http://online.wsj.com/
http://www.washingtonpost.com/

Travel
http://www.itn.net/
http://www.expedia.com/
http://www.hotels.com/
http://www.orbitz.com/
http://www.priceline.com/
http://www.travelocity.com/