An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites as of May 2003

Lorrie Faith Cranor, Simon Byers, and David Kormann
AT&T Labs-Research
Florham Park, NJ

Technical Report prepared for the 14 May 2003
Federal Trade Commission Workshop on
Technologies for Protecting Personal Information

http://www.research.att.com/projects/p3p/

EXECUTIVE SUMMARY

The Platform for Privacy Preferences (P3P) provides a standard computer-readable format for privacy policies and a protocol that enables web browsers to read and process these policies automatically. We developed software to query a set of web sites for P3P policies, check the validity of each policy, and analyze the information practices it describes. We used this software to analyze 538 P3P-enabled web sites found by checking for P3P policies on 5,856 web sites on 6 May 2003. The sites we checked for P3P policies were taken from several lists of popular web sites, as well as from "crawling" indexes of shopping, news, children's and government web sites. We present the first major analysis of the data practices of P3P-enabled web sites.

Our system used the P3P evaluation engine built into the AT&T Privacy Bird P3P user agent to analyze the P3P policies we discovered. We checked these policies against Privacy Bird's standard "high," "medium," and "low" settings as well as against 62 other "rule sets" that we developed.

A comparison of our results with previous studies indicates that P3P adoption is increasing over time [1][14]. Adoption remains highest for the most popular web sites.

We found a large number of errors in the P3P policies of the sites we evaluated. About one third of the P3P-enabled sites had technical errors. In many cases these errors were due to use of syntax from a draft version of the P3P specification that is not permitted by the final P3P 1.0 Recommendation [6]. However, 7% of the P3P-enabled sites had critical errors that prevented their evaluation by our Privacy Bird evaluation engine. We also found 74 sites that violated the P3P specification by posting P3P compact policies without their corresponding full P3P policies.

Our analysis of data collection at P3P-enabled web sites indicates that most sites collect computer information, click stream information, and demographic data. Almost as many sites also collect online contact information, physical contact information, interactive data, and unique identifiers. The majority of sites also collect preference information, purchase information, and state management information (cookies). However, fewer sites collect financial information (which excludes information such as credit card numbers used only as part of a purchase). The least collected information is content (email messages, bulletin board postings, etc.), government-issued identifiers, health information, political information, location information (for example GPS positioning data), and information not falling into any of the pre-defined categories.

Almost all web sites reported using data for completion and support of the activity for which data was provided, web site and system administration, and research and development. The majority of sites also reported using data for email and postal mail marketing, one-time tailoring of the site content, and pseudonymous profiling. Substantially fewer sites reported using data for telemarketing or profiling in which individuals are identified. Very few sites reported using data for historical preservation or other purposes that do not fall into these categories.

About half the web sites we studied indicated that they share personally identifiable data with parties other than agents who use data for the purpose for which it was provided. 46% of these sites indicate that they offer third-party choice (opt-in or opt-out to this sharing).

Most web sites reported providing some access provisions for individuals wishing to find out what data of theirs was in a web site's records as well as some dispute resolution option for disputes related to their privacy policy. However, most sites reported that they did not have a data retention policy covering all of the data they collected at their site.

As debates continue about the need for further privacy legislation and the effectiveness of industry self-regulation in the privacy area, automated analysis of the data practices of P3P-enabled web sites can provide valuable information. Furthermore, as US government web sites begin posting P3P policies to comply with the privacy requirements of section 208 of the E-Government Act of 2002 [16], we can monitor compliance with these requirements. We plan to repeat our experiments on a regular basis to allow for longitudinal analysis of P3P policies. In the future we may also expand the list of web sites we analyze, develop additional rule sets to facilitate more detailed analysis, and expand our analysis to include P3P compact policies.

1. INTRODUCTION

The posting of privacy policies on commercial web sites has been a key component of the US self-regulatory approach to online privacy protection. In addition, privacy regulations in the US, Europe, and elsewhere include provisions that require regulated companies to provide notice of their data practices through a privacy policy. Whether posted voluntarily or to comply with regulations, privacy policies serve to increase transparency about data practices and support the "notice" or "openness" fair information practice principle [2][17][22].

Over the past several years a number of studies have been undertaken to measure the percentage of US commercial web sites that have posted privacy policies and to assess the types of practices disclosed in these policies [1][9][10][11][17][22]. The US Federal Trade Commission has taken these studies into consideration in their evaluation of the extent to which privacy self-regulation is working [17][18][19].

Each of the privacy policy assessment studies has required considerable effort to carry out, as the process of reading privacy policies is time consuming and error-prone. However, web sites are increasingly making their privacy policies available in a computer-readable format called P3P, thus making automated privacy policy assessments possible.

1.1 The Platform for Privacy Preferences

The Platform for Privacy Preferences (P3P) [3] provides a standard computer-readable way for web sites to communicate about their privacy policies. Privacy policies are intended to describe a company's *data practices*—what information they collect from individuals and what they do with it. The P3P 1.0 Specification [6] defines an XML encoded language for creating a P3P policy that describes a site's data practices. A P3P policy is composed essentially of the answers to a number of multiple-choice questions, and thus does not always contain as much detailed information as a human-readable privacy policy (i.e., a policy that is intended for people, rather than computers, to read). However, sites can provide additional detail through the use of human-readable fields within the P3P policy. Furthermore, because most of the fields in a P3P policy are required, P3P policies often include information about aspects of privacy that sites have chosen not to cover in their human-readable policies.

The P3P specification includes a protocol, built on the HTTP protocol, for requesting and transmitting P3P policies. P3P user agents—software tools, typically built into web browsers, that fetch P3P policies and process them on a user's behalf—use standard HTTP requests to fetch a P3P policy reference file from a well-known location on the web site to which a user is making a request. The policy reference file indicates the location of the P3P policy file that applies to each part of the web site. There might be one policy for the entire site, or several policies that each cover a different part of the site. A P3P user agent can then fetch the appropriate policy, parse it, and take action according to the user's preferences.

P3P also allows sites to place policy reference files in locations other than the well-known location. In these cases, the site must declare the location of the policy reference file using a special HTTP header or by embedding a link tag in the HTML files to which the P3P policies apply. Special HTTP headers are also used to transmit an optional P3P compact policy whenever cookies are set. Compact policies are very short summaries of full P3P policies that describe only the data practices related to cookies. They do not have the full expressive capabilities of P3P policies. A site that uses a P3P compact policy is also required to post a corresponding full P3P policy.

The P3P specification was developed by a working group of the World Wide Web Consortium (W3C). The work took place over a period of approximately five years and involved representatives from industry, academia, non-profits, and government from around the world [4]. P3P became an official W3C "Recommendation" just over a year ago on April 16, 2002. P3P user agents are already built into the Microsoft Internet Explorer 6 [20] and Netscape Navigator 7 [12] web browsers. Other P3P user agents are available as browser add-ons or proxies [3]. The AT&T Privacy Bird is a P3P user agent implemented as a "browser helper object" [15] that works with the Microsoft Internet Explorer 5.01, 5.5, and 6.0 web browsers on Microsoft Windows 98/2000/ME/NT/XP operating systems. It displays a bird icon in the browser title bar that changes color and shape to indicate whether or not a web site's P3P policy matches a user's privacy preferences [5].

Privacy Bird makes use of a rule-based XML language called APPEL [7] for storing user privacy preferences. Users can export their preferences as APPEL rule sets or import APPEL rule sets they created themselves or obtained from other sources.

1.2 Web Sweeps

From 1998 to 2001, the US Federal Trade Commission conducted or solicited annual surveys of commercial web site privacy policies, dubbed "web sweeps." The methodology used to conduct each of these surveys varied slightly, but generally involved having a team of "web surfers" visit several hundred web sites and look for the presence of privacy disclosures and systematically analyze them for mention of specific elements related to fair information practice principles. The survey reports indicate that surfers spent up to 35 minutes searching for and analyzing the privacy disclosure at each site they visited, and in most studies each surfer's findings were verified by a second surfer [22].

Due to differences in methodology used each year, especially in the approach to selecting the sites to survey, not all the web sweep results are directly comparable. Milne and Culnan extracted more comparable results from the original data by accounting for methodological and sampling differences [22]. They report that the percentage of the 100 "most popular" web sites posting privacy policies was 45% in 1998, 85% in 1999, 97% in 2000, and 99% in 2001. The percentages for random samples of web sites (drawn from various lists of frequently-visited web sites) was significantly lower each year. For example, in 2001, 77% of a random sample of the sites with more than 39,000 unique visitors each month (as reported by Nielsen/NetRatings) had posted privacy policies.

The web sweeps produced statistics on the percentage of sites collecting certain types of information, using cookies and third-party cookies, and supporting various fair information practice principles. Furthermore, the 2001 sweeps included statistics on P3P adoption. Separately, the consulting firm Ernst & Young (which conducted the 2001 sweeps) has been issuing periodic reports on P3P adoption since August 2002 [13].

All of the web sweeps measured the extent to which web sites offered visitors choices about marketing uses and sharing of their data. The 2000 and 2001 web sweeps differentiated between *internal choice*—"the use of personal information by the site to send communications (other than those related to processing an order or responding to a consumer's question) to the consumer"—and *third-party choice*—"the disclosure of PII [personally identifying information] to entities other than the domain" [11][1]. Sites were further classified according to whether they provided opt-in or opt-out choice options. Unfortunately, surfers were not always able to distinguish between opt-in and opt-out practices, and thus a substantial number of sites are reported to be "unclear."

2. SYSTEM DESIGN AND IMPLEMENTATION

We developed a system for automating the process of measuring P3P adoption and gathering data from P3P-enabled web sites that allows for analyses similar to those conducted as part of the various web sweeps. Our system includes five major groups of components: a URL collection mechanism, a P3P policy retriever, a scripted interface to the W3C P3P Validator, a P3P policy evaluator, and some generic data analysis tools.

2.1 URL Collector

The first stage in studying web site adoption of P3P is to identify sets of sites of interest. We use some existing lists of URLs, however, we also use lists we constructed ourselves that focus on particular types of web sites. We use web spidering techniques to gather information from web directories and other sources. This URL collection process may involve some substantial work and time to yield a set of URLs that meet a particular selection criterion.

2.2 P3P Policy Retriever

We developed a Perl script for retrieving as much P3P information as possible from websites, including all policies, policy reference files and compact header polices. Our script first makes an HTTP request to the root page of the web site (http://*host.domain*/) and checks the response. We distinguish between two types of HTTP success responses: successful retrieval of the root page or retrieval of an HTML page informing us indirectly of an HTTP 404 Not Found. Secondly we attempt to

establish whether the web server has redirected us and if so we note the new location along with the initial requested URL. In order to allow for sites that use redirects to set and check cookies we enable cookies in our user agent. We note any sites that appear to be unreachable (fail to respond to our requests within 10 seconds), and we revisit these sites after checking all of the other sites on our list. We classify any sites that fail to respond to both our first and second attempts to contact them as unreachable.

We begin searching for P3P data by checking the headers of the response to our request for the root page of the web site. We record any references to policy reference files and P3P compact policy headers present. Second, we analyze the returned HTML file for a P3P link tag that provides the location of a P3P policy reference file. Finally, we make another HTTP request in an attempt to obtain a P3P policy reference file from the well-known location on the web site (http://*host.domain*/w3c/p3p.xml). At sites where a redirect response has been detected, we check for a policy reference file at the well-known location on both the original site and at the site to which we were redirected.[1] Most websites use the well-known location method; however we check for policy reference files at the other locations as well to obtain full and accurate data. Those sites that do not appear to have a policy reference file at any of these locations are deemed to not be P3P enabled.

Given a list of policy reference files we then proceed to download each one in turn and make rudimentary checks that they are indeed P3P policy reference files, however, we do not validate these files. For each policy reference file we parse it and determine the P3P policy files it references.

Now given a list of unique P3P policy files we fetch each one in turn and store them for further processing along with the name of each policy and the location where it was found. As with policy reference files we do not validate the policy files; however, we eliminate those that appear to be HTML files rather than P3P policy files.

Our policy retrieving system results in the retrieval of almost all valid P3P policies from web sites as well as some policies that are technically invalid, but may be readable by some popular P3P user agents. Our system also yields other data such as information about P3P compact policies and web server data. We intend to do further analysis of compact policies in the future.

2.3 P3P Validator

Initially, rather than developing our own scripts for retrieving P3P policies, we had developed a scripted interface to the W3C P3P Validator.[2] The Validator fetches P3P policy reference files, policy files, and compact policies and checks them for compliance with the P3P 1.0 specification. However, the Validator stops validation upon encountering an error. Thus, we were unable to use this tool to retrieve P3P policies from sites with errors in their policy reference files. We discovered that quite a few sites that we surveyed had errors in their policy reference files. However, most of these errors did not prevent IE6, Netscape 7, or AT&T Privacy Bird from accessing a site's P3P policy. We decided to write our own scripts for retrieving P3P policies, but continue to use the W3C P3P Validator to gather statistics on P3P compliance. Using the Validator we can derive information easily about the frequency and type of compliance errors, as well as the use of P3P compact policies. We store a file containing Validator output for each P3P policy for further analysis.

---

[1] If the well-known location method is used, only the policy reference file at the well-known location on the site from which a page is actually served can be used to determine the applicable policy. Thus, if a server issues a redirect response to a page request but serves a policy reference file in response to a request for the file at the well-known location, that policy reference file cannot be used for determining the policy that applies to the redirected page. However, we have found some sites that use multiple redirects—sometime through an authentication server—and eventually do serve the requested page from the server to which the request was originally made. Without the appropriate passwords and scripting for our automated system to authenticate itself we cannot verify that this is indeed what happens and thus we cannot determine automatically which policy reference file is actually applicable. Therefore we check both the original site and the redirect site for policy reference files at the well-known location. Thus, there is a possibility that we may be counting a small number of sites as P3P enabled that actually do not have their P3P files at the appropriate location. However, we believe this occurrence to be quite rare.

[2] The W3C P3P Validator is available as a free service at http://www.w3.org/P3P/validator. It was implemented in Perl by Yuichi Koike and Shojima Taiki.

2.4 P3P Policy Evaluator

The AT&T Privacy Bird user agent[3] includes a P3P policy evaluator engine that compares a web site's policy with a user's privacy preferences encoded as an APPEL rule set. We extracted the C++ code for the evaluator engine from the Privacy Bird code and removed the Microsoft Windows-specific code so that it would compile on a Linux system and be used independently from the Privacy Bird graphical user interface. We developed a command-line front-end for the evaluator and a Perl interface that calls the C++ module with an APPEL rule set and a locally- or remotely-stored P3P policy file. The evaluator returns an integer that specifies the number of APPEL "limited" rules that fired. Each of these rules indicates a mismatch between the P3P policy and the privacy preferences encoded in the rule set.

2.5 Data Analysis

The outputs of the many policy evaluations are gathered into a rectangular matrix, with each row corresponding to a policy from a web site, and each column an APPEL rule set file. Other attributes of the web sites can be included such as the type of web site. We then run a Perl script over the matrix to produce various tabulations such as the number of sites that returned response values greater than zero for each APPEL rule set. Further analysis can be done with other scripts or interactively.

3. METHODOLOGY

We assembled 10 lists of web sites and merged them to produce a set of 5,856 unique web sites. We used our P3P policy retriever script to check for P3P policies at each site and the W3C P3P Validator to check the validity of each P3P-enabled site. We then ran our P3P policy evaluator over each P3P policy using 64 APPEL rule sets plus a test APPEL rule set used for sanity checking. At sites that reference multiple policies in their policy reference file, we evaluated only the first policy referenced.[4] We used Perl scripts to analyze the output of the policy evaluator and tabulate our results.

3.1 Web Site Selection

It is not feasible to check every web site in existence to see whether it is P3P enabled. Even if this was practical, it is not clear that statistics derived from this study would be meaningful as there is considerable variation in the frequency with which web sites get visited. If we are interested in determining the extent of P3P enabled sites from a user's perspective, we need to focus our study on the sites frequently visited by users. There are several lists compiled on a regular basis of top web domains. There is some overlap between the lists; however, there is considerable variation in the methodology used to compile them. There are also a number of popular web site indexes. It is likely that sites listed in indexes are also among the more frequently visited web sites on the Internet. In this study we examined ten lists of web sites:

- **PFF Most Popular.** This list was used by the Progress and Freedom Foundation for the 2001 web sweeps [1]. It contains 85 of the 100 busiest sites as determined by the October 2001 Nielsen/NetRatings ranking of sites with the most unique visitors per month. PFF excluded adult sites, children's sites, business-to-business sites, and sites not in the .com top level domain in order to focus their study on US consumer web sites.
- **PFF Random.** Also used in the 2001 web sweeps [1], this list contains a random sample of 302 of the 7,821 domains with at lest 39,000 unique monthly visitors in October 2001, as estimated by Nielsen/NetRatings. Adult sites, children's sites, business-to-business sites, and non-dot-coms were also excluded.
- **PFF Refined Random.** Also used in the 2001 web sweeps [1], this list contains the 209 domains from the PFF Random list that were in the top 5,625 domains in October 2001, as estimated by Nielsen/NetRatings. PFF used this sample in addition to the random sample in order to make their results more comparable with previous web sweeps.
- **Netscore Top 500.** Used by Ernst & Young in their P3P Dashboard Reports [13][14], this list includes the 500 domains with the most unique visitors during July 2002 according to the comScore Media

---

[3] The AT&T Privacy Bird software was implemented in C++ by Praveen Guduru and Manjula Arjula.

[4] In most cases this is the policy that covers the site's homepage, but for some sites this is not the case.

Metrix netScore Standard Traffic Measurement report. We also report statistics for the top 100 sites on this list.

- **Key Measures.** This list includes the top 500 domains with the most unique visitors during July 2002 according to the comScore Media Metrix Key Measures report. This list includes "third-party" sites, such as advertising networks, that don't appear in the other samples.
- **Alexa.** This list includes the top 500 domains according to the Alexa Traffic Ranking on February 4, 2003. The traffic rank[5] is a combined measure of page views and unique visitors based on three months of aggregated traffic data from Alexa Toolbar users. This list includes many non-US domains and adult sites that don't appear in the other samples.
- **Froogle.** This list includes 1017 sites obtained by crawling the www.froogle.com web site in April 2003. Froogle indexes sites that offer products for sale.
- **Yahooligans.** This list includes 900 sites obtained by crawling www.yahooligans.com in April 2003. Yahooligans indexes sites geared towards children ages 7-12. We included any link to offsite content or advertising obtained by our crawler.
- **Firstgov.** This list includes 344 government sites indexed at www.firstgov.gov in April 2003. These include US federal government sites as well as some US state government sites and sites for some quasi-government organizations.
- **News.** This list includes 2,429 sites obtained by crawling news.google.com in April 2003. These include a variety of news-reporting organizations from the US and other countries.

We selected the PFF lists in order to draw comparisons with the 2001 web sweeps. We selected the Netscore list in order to draw comparisons with the Ernst & Young Dashboard Reports. We selected the Key Measures list in order to compare the variation in results from two lists derived using differing methodologies during the same time period. We selected the Alexa list because of the large number of non-US domains it contains. We selected the Froogle, Yahooligans, Firstgov, and News lists to get larger samples of shopping, children's, government, and news web sites respectively.

For the PFF, Netscore, and Key Measures lists we checked only for a P3P policy covering the www host in each domain. For the domains we obtained from our web crawls we checked the specific hosts retrieved as part of the spidering process. Because some of the lists we used had been generated some time before this study was conducted, a number of sites were no longer reachable when we conducted our study.[6]

3.2 Privacy Bird Evaluation

The AT&T Privacy Bird user agent comes with three standard settings: high, medium, and low. We ran our P3P policy evaluator over APPEL rule sets representing each of these three settings.[7] A policy that matches the preferences expressed in a rule set receives a "green bird" from the policy evaluator, while a policy that does not match the preferences expressed in a rule receives a "red bird." The three rule sets encode the following preferences:

- **Low.** Trigger a red bird at sites that collect health or medical information and share it with other companies or use it for analysis, marketing, or to make decisions that may affect what content or ads the user sees. Also trigger a red bird at sites that engage in marketing but do not provide a way to opt-out.
- **Medium.** Same as low, plus trigger a red bird at sites that share personally identifiable information, financial information, or purchase information with other companies. Also trigger a red bird at sites that collect personally identified data but provide no access provisions.

---

[5] Alexa Traffic Rank is described in more detail at http://pages.alexa.com/prod_serv/traffic_learn_more.html.

[6] If we were unable to connect to a site on our second attempt using a 10-second time out we classified a site as unreachable.

[7] The APPEL files we used were generated with the Beta 1.2 version of AT&T Privacy Bird.

- **High.** Same as medium, plus trigger a red bird at sites that share any personal information (including non-identified information) with other companies or use it to determine the user's habits, interests, or other characteristics. Also trigger a red bird at sites that may contact users for marketing or use financial or purchase information for analysis, marketing, or to make decisions that may affect what content or ads the user sees.

For all three settings (as well as the choice assessment below), a site is classified as not sharing data if it shares data only with agents that use it only to complete the transaction for which it was provided or with delivery companies (which may have unknown data practices). In addition, the Privacy Bird settings classify a site as not sharing data if data sharing occurs only under an opt-in policy.

For these three settings (as well as the assessments below) data from the following P3P categories are considered *personally identifiable information*: physical contact information, online contact information, and government issued identifiers.

3.3 Types of Data Collected

The P3P 1.0 specification enumerates 17 types of data a web site might collect. Sites use the <CATEGORIES> element in their P3P policies to disclose the types of data they collect. Sites may also list specific data elements they collect (for example, first name or last name). All data elements are assigned to one or more of the 17 data categories. We created APPEL rule sets to test for disclosures about each of the 17 data categories in a P3P policy. These rule sets identify a site as collecting data of a particular category if it explicitly references that category or if it references a data element assigned to that category.

3.4 Data Usage

The P3P 1.0 specification enumerates 12 purposes or uses of data. Sites use the <PURPOSE> element in their P3P policies to disclose their data usage. Most "primary" data uses are captured under a single purpose, while the remaining 11 purposes are usually considered "secondary" data uses. We created APPEL rule sets to test for disclosures about each of the 12 purposes in a P3P policy.

3.5 Data Recipients and Sharing

The P3P 1.0 specification enumerates six categories of data recipients. Sites use the <RECIPIENT> element in their P3P policies to disclose the potential recipients of user data. One of the recipients disclosures restricts data sharing to the web site and its agents, while the others permit broader data sharing under various conditions. We created five APPEL rule sets to test for disclosures in a P3P policy about the sharing of personally identifiable information. (Note, we did not test for disclosures about sharing data not included in the three categories we are considering to per personally identifiable.) We also created a rule set to test for the presence of any of the four recipients disclosures that would indicate sharing of personally identifiable information beyond the web site and its agents and delivery companies.

3.6 Choice Options

In our study we reproduced the choice assessment in the 2001 web sweeps [1] and made an additional distinction about types of internal choice—telemarketing and other types of marketing. We also report statistics on choice only for sites that market or share data rather than for all sites that collect personally identifying information (which fails to distinguish sites that don't market or share at all from those that provide choice about marketing or sharing).

We conducted the choice assessment through the use of six APPEL rule sets, which tested for the following conditions:

- Site engages in telemarketing but offers opt-in
- Site engages in telemarketing but offers opt-out (and not opt-in)
- Site engages in marketing (other than telemarketing) but offers opt-in
- Site engages in marketing (other than telemarketing) but offers opt-out (and not opt-in)
- Site shares personally identifiable information but offers opt-in

• Site shares personally identifiable information but offers opt out (and not opt-in)

Consistent with the 2000 and 2001 web sweeps, we gave a site credit for opt-in or opt-out if that choice was offered at all, even if it was not offered for all collected data.

3.7 Access Provisions

The P3P 1.0 specification enumerates six different provisions for providing individuals with access to the personally identified information a web site has collected about them (including no access, and no personally identified information collected). Sites use the <ACCESS> element in their P3P policies to disclose their access policies. We created APPEL rule sets to test for each of the access disclosures.

3.8 Dispute Resolution Options and Remedies

Web sites can disclose privacy-related dispute-resolution procedures in their P3P policies. The P3P 1.0 specification enumerates four categories of dispute resolution procedures. Sites use the <DISPUTES> element to disclose their dispute resolution procedures. In addition, they can use the <REMEDIES> element to disclose remedies that are available to individuals should the site fail to abide by its privacy policy. We created APPEL rule sets to test for each type of disputes disclosure as well as for the presence of a remedies disclosure.

3.9 Data Retention Policies

The P3P 1.0 specification enumerates five types of data retention policies (including no retention policy). Sites use the <RETENTION> element in their P3P policies to disclose the type of retention policy in effect. We created APPEL rule sets to test for each type of retention policy.

3.10 Other Assessments

Web sites can optionally use the <CONSEQUENCE> element to provide human-readable explanations about their data practices in their P3P policies. We created an APPEL rule set to test whether a site makes use of the <CONSEQUENCE> element.

Web sites can use the <NON-IDENTIFIABLE> element to indicate that they do not collect data or that all of the data they collect is immediately anonymized. The P3P specification has some fairly stringent restrictions on when a site can use the <NON-IDENTIFIABLE> element. We created an APPEL rule set to test whether a site makes use of this element.

We also created a test APPEL rule set that simply checks to see whether a policy contains a <POLICY> element (required for all policies). We use this as a sanity check for our system to make sure all policies are getting processed.

4. RESULTS

Once constructed, our system was able to retrieve and evaluate P3P policies quickly and without human intervention. Running on a 1.4 Ghz Pentium 4 computer connected to the Internet via a cable modem our system took approximately 4 hours to check 5,856 web sites for P3P policies and policy reference files, 3 hours to check those sites with P3P policies using the W3C P3P Validator, and 1 hour to evaluate 538 P3P policies against 65 APPEL rule sets. We gathered approximately 16 Mb of data. With further optimizations such as parallelizing the web requests, running the Validator code locally, and doing validation concurrently with policy retrieval, we expect we could improve our system performance considerably.

4.1 P3P Adoption

On 6 May 2003 we evaluated 5,856 web sites (of which 5,728 were reachable) and discovered 538 of them that had been P3P-enabled with P3P policies and policy reference files. Appendix A lists these sites. Table 1 summarizes our findings on P3P adoption. In the course of developing and testing our system we noticed some day-to-day fluctuations in our results, mostly due to web servers being temporarily unresponsive. We estimate that these fluctuations impact our adoption rate results by less than 1%.

Table 1. Web Site P3P Adoption

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Date	Oct. 2001	Oct. 2001	Oct. 2001	July 2002	July 2002	July 2002	Feb. 2002	April 2003	April 2003	April 2003	April 2003	--
Sites	85	209	302	100	500	500	500	1017	900	344	2429	5856
Sites reachable in May 2003	84	194	289	100	485	486	487	1013	874	343	2383	5728
P3P-enabled sites sites (with P3P PRFs and P3P policies)	25	33	27	29	105	110	83	128	25	7	195	538
% of reachable sites that are P3P-enabled	30% (22%)	17% (5%)	9% (4%)	29% [28%]	22% [18%]	23%	17%	13%	3%	2%	8%	9%
P3P-enabled sites readable by Privacy Bird	23	31	25	27	97	99	74	118	24	6	190	502
% of reachable sites that are readable by Privacy Bird	27%	16%	9%	27%	20%	20%	15%	12%	3%	2%	8%	9%
P3P-enabled sites unreadable by Privacy Bird	2	2	2	2	8	11	9	10	1	1	5	36
% of P3P-enabled sites unreadable by Privacy Bird	8%	6%	7%	7%	8%	10%	11%	8%	4%	14%	3%	7%

Note, percentages in parentheses represent comparable percentages from the 2001 web sweeps [1]; percentages in square brackets represent comparable percentages from the January 2003 P3P Dashboard report [14].

A comparison of our results with previous studies indicates that P3P adoption is increasing over time [1][14]. Adoption remains highest for the most popular web sites. In addition, a closer examination of the results for the Netscore and Key Measures site lists suggests that the slightly higher adoption among the Key Measures sites may be due to the presence of many "third-party" sites on that list. Third-party sites have been quick to adopt P3P in order to avoid having their cookies blocked by IE6. The Alexa top 500 list, resulted in the lowest adoption numbers of the three top 500 lists, probably due to its international nature and the large number of adult sites on that list.

We found a large number of errors in the P3P policies of the sites we evaluated. About one third of the P3P-enabled sites had errors flagged by the W3C P3P Validator. In many cases these errors were due to use of syntax from a draft version of the P3P specification that is not permitted by the final P3P 1.0 Recommendation [6]. However, 7% of the P3P-enabled sites had errors that prevented their evaluation by our Privacy Bird evaluation engine. These included omitting required components of a P3P policy and improperly referencing data elements.

Overall we found 321 web sites that used P3P compact policies. Of these, 247 also had full P3P policies and 74 did not have full P3P policies. Thus 46% of the P3P-enabled sites we examined used compact policies. Most of the 74 sites with P3P compact policies but no full P3P policies probably created their compact policies to prevent the IE 6 web browser from blocking their cookies (by default IE6 blocks cookies used in a third-party context that do not have compact policies). However, it is a

violation of the P3P specification to post a P3P compact policy without a corresponding full policy. Appendix B lists the sites with P3P compact policies but no full P3P policies.

The P3P specification allows web sites to declare multiple P3P policies covering different parts of their site. Of the 538 P3P-enabled sites we examined, only 27 of them had more than one policy. We found 17 sites with two policies each and four sites with three policies each. In addition we found five sites that had four to eight policies each, and one site with over 10 policies. 504 sites placed their policy reference files at the well-known location (94%), while 17 used HTML link tags and 177 referenced the location of their policy reference file in a header. (Some sites used more than one of these methods.)

4.2 Privacy Bird Evaluation

Table 2 summarizes our findings on how P3P-enabled web sites are evaluated under the three standard Privacy Bird settings. We report the number of sites that receive "red birds" under each setting, indicating that they do not match the user preferences specified by that setting. Not surprisingly, the number of sites receiving red birds on the low setting is about half the number receiving red birds on the medium setting and less than a third the number receiving red birds on the high setting. Only 24% of the sites we evaluated received a red bird on the low setting, in most cases because they did not offer users the ability to opt-out of marketing and/or telemarketing. The most popular sites were more likely than other sites to receive green birds on the low setting, probably due to a greater awareness of the importance of the "choice" principle among these sites. On the other hand, the most popular sites were also more likely than other sites to receive red birds on the high setting, probably due to the fact that most offer rich ecommerce environments that rely heavily on targeted marketing and profiling visitors. The Froogle and Yahooligans sites were the most likely to receive red birds on the low setting, probably because these sites were the most likely to collect health and medical information.

Table 2. Privacy Bird Evaluation of P3P-Enabled Web Sites: Percentage of Sites Receiving "Red Birds" Under High, Medium, & Low Settings

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
High setting	83	72	68	81	69	71	74	77	75	67	93	82
Medium setting	35	24	29	30	35	33	46	55	50	17	51	47
Low setting	13	20	23	11	24	16	18	43	42	0	15	24

4.3 Types of Data Collected

As shown in Table 3, nearly every web site disclosed collecting computer information (e.g. type of computer, operating system, and IP address) and click stream information. This is not surprising considering that this information is routinely transferred and recorded as part of the HTTP protocol used for retrieving content from web sites.

Most web sites also collected demographic data. Sites on the Froogle list and government web sites were less likely than other sites to collect this information than other sites. Almost as many sites collected online contact information, physical contact information, interactive data, and unique identifiers. News web sites were most likely to collect this information than other sites.

The majority of sites also collected preference information, purchase information, and state management information (cookies). Not surprisingly, fewer collected financial information (which excludes information that is used only to process a purchase).

The least collected information was content (email messages, bulletin board postings, etc.), government-issued identifiers, health information, political information, location information (for example GPS positioning data), and information not falling into any of the pre-defined categories. Most

of this information is unnecessary for typical e-commerce transactions. As more location-based services are offered via Internet-enabled hand-held devices, we would expect more sites to collect location information. Interestingly, none of the government web sites reported collecting government-issued identifiers. However, the number of government web sites with P3P policies is still very small and these sites are not necessarily representative of all government web sites.

Table 3. Types of Data Collected at P3P-Enabled Web Sites: Percentage of Sites Collecting Each P3P Data Type

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Computer information	100	100	100	96	95	94	96	90	100	100	98	95
Content	17	24	19	26	32	23	23	32	46	50	13	23
Demographic and socioeconomic data	96	100	100	100	92	90	89	87	96	67	99	94
Financial information	43	12	10	37	18	21	32	30	33	17	38	30
Government-issued identifiers	9	8	6	7	10	10	9	22	17	0	5	11
Health information	9	4	3	7	8	7	9	23	33	0	10	13
Interactive data	83	68	74	85	74	71	80	55	71	83	89	76
Location data	0	0	0	0	3	1	8	3	0	0	1	2
Navigation and click-stream data	96	100	100	93	94	93	97	90	100	100	98	96
Online contact information	70	68	71	89	73	71	70	79	75	83	87	78
Other	0	4	6	0	2	3	3	3	0	0	2	2
Physical contact information	70	68	71	89	73	71	73	80	67	83	87	78
Political information	9	4	3	7	5	5	5	21	33	0	9	11
Preference data	70	60	52	63	58	51	59	35	50	17	81	59
Purchase information	52	52	48	59	48	49	53	72	54	67	52	54
State management mechanisms	43	44	45	56	47	57	53	66	63	67	54	55
Unique identifiers	78	72	74	93	81	79	84	81	63	67	96	84

Note, definitions of data types can be found in the P3P 1.0 specification [6]. Sites may collect multiple types of data so percentages do no sum to 100.

4.4 Data Usage

Not surprisingly, almost all web sites reported using data for completion and support of the activity for which data was provided, web site and system administration, and research and development. The majority of sites also reported using data for email and postal mail marketing, one-time tailoring of the site content, and two forms of pseudonymous profiling. Substantially fewer sites reported using data for

telemarketing or profiling in which individuals are identified by name or other personally identifiable data. Very few sites reported using data for historical preservation or other purposes that do not fall into these categories. Interestingly, while the historical preservation purpose is applicable primarily to government web sites, no government sites claimed to use data for this purpose. News web sites were more likely than other sites to use data for almost every purpose. The percentage of sites using data for each purpose are shown in Table 4.

Table 4. Data Usage at P3P-Enabled Web Sites: Percentage of Sites Using Data for Each P3P Purpose

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Web site and system administration	96	100	94	85	89	83	88	81	92	83	95	88
Contacting visitors for marketing services or products	70	64	58	81	59	56	61	59	63	67	82	66
Completion and support of activity for which data was provided	96	92	94	96	95	90	93	83	92	100	95	91
Research and Development	91	100	97	81	80	80	78	75	75	67	92	83
Historical preservation	9	12	13	7	10	7	7	14	4	0	2	8
Individual analysis	22	28	26	37	34	32	30	34	46	17	24	30
Individual decision	35	32	29	44	35	38	43	36	46	17	54	44
Other	9	4	6	7	5	4	5	25	17	0	2	9
Pseudonymous analysis	65	48	48	74	53	53	61	54	63	67	86	67
Pseudonymous decision	48	44	45	59	46	48	58	50	63	67	85	65
One-time tailoring	65	64	68	67	69	66	70	67	58	50	88	74
Contacting visitors for marketing services or products via telephone	30	32	26	22	25	23	38	31	38	33	42	35

Note, definitions of data use purposes can be found in the P3P 1.0 specification [6]. Sites may use data for multiple purposes so percentages do no sum to 100.

4.5 Data Recipients and Sharing

About half the web sites we studied indicated that they share personally identifiable data with parties other than agents who use data for the purpose for which it was provided. News web sites were most likely to share data and government web sites were least likely to share data. Sites on the Froogle list were most likely to share data with a delivery company, which is not surprising considering that most of

these sites sell physical goods. The percentage of sites sharing data with each type of data recipient is shown in Table 5.

Table 5. Data Recipients and Sharing of PII at P3P-Enabled Web Sites: Percentage of Sites Sharing Data with Each Type of P3P Data Recipient

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Delivery services possibly following different practices	17	20	23	19	28	28	22	61	33	33	40	38
Legal entities following different practices	22	16	13	26	19	19	19	29	38	0	14	18
Public for a	9	8	6	7	4	3	5	22	21	0	4	9
Legal entities following our practices	26	24	23	26	18	18	38	30	33	17	67	42
Unrelated third parties	13	8	10	15	8	7	5	25	29	0	11	13
Sharing with parties other delivery services and agents	43	36	35	44	29	31	46	35	46	17	74	49

Note, definitions of data recipients can be found in the P3P 1.0 specification [6]. These statistics are for sharing of PII only. Sites may share data with multiple types of recipients or they may not share data at all, so percentages do no sum to 100.

4.6 Choice Options

Table 6 summarizes our findings on the choice options offered by P3P-enabled web sites. Note that our assessment of choice options is not directly comparable to the statistics presented in the web sweeps due to the fact that the web sweeps data includes sites that do not report whether or not they share data or use it for marketing, and sites that indicate that they offer choice without explaining whether they offer opt-in or opt-out. P3P-enabled web sites must make concrete disclosures on these points so there is much less room for ambiguity. Furthermore, the 2001 web sweeps [1] reported choice as the percentage of domains collecting PII that offered choice, rather than as the percentage of domains engaging in marketing (or sharing) that offered choice. None-the-less, our observations about choice are similar to those reported in the web sweeps.

The top sites were more likely to engage in marketing than less popular sites, but also more likely to offer choice. Internal choice was more often offered using opt-out than opt-in, except at children's sites and news sites. Third-party choice was more often offered using opt-in than opt-out.

Sites were less likely to share PII with third parties than they were to use data for marketing. However sites that did share data were less likely to offer third-party choice than internal choice. While others have found similar results [1], the magnitude of the differences previously reported has been smaller. We suspect that a significant fraction of the sites with P3P policies indicating that they share data but do not offer third-party choice do in fact offer this choice (and we confirmed this for several sites by reading their privacy policies). Earlier versions of the P3P specification did not permit sites to indicate that they offered third-party choice. As indicated by our W3C Validator results, a large fraction of the sites we evaluated were compliant with early versions of the P3P specification rather than the P3P 1.0 Recommendation [6].

Table 6. Choice Options at P3P-Enabled Web Sites

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
% of sites telemarketing that offer opt-in	0	25	25	0	21	13	4	6	33	0	18	13
% of sites telemarketing that offer opt-out	71	50	50	67	33	61	75	11	11	100	70	53
% of sites marketing that offer opt-in	19	25	22	36	30	27	18	14	47	50	54	35
% of sites marketing that offer opt-out	63	44	44	50	37	49	62	17	13	50	37	36
% of sites marketing or telemarketing that offer opt-in	19	31	28	36	32	29	18	16	47	50	54	37
% of sites marketing or telemarketing that offer opt-out	63	44	44	50	37	49	64	17	13	50	37	36
% of sites marketing or telemarketing that offer internal choice (opt-in or opt-out)	81	75	72	86	68	76	80	31	60	100	91	72
% of sites sharing PII that offer opt-in	30	56	45	42	39	45	24	15	45	0	51	40
% of sites sharing PII that offer opt-out	0	11	9	0	7	6	12	15	0	100	1	5
% of sites that share PII offering third-party choice (opt-in or opt-out for sharing PII)	30	67	55	42	47	52	35	29	45	100	52	46

Note, definitions of data use purposes can be found in the P3P 1.0 specification [6]. Sites may use data for multiple purposes so percentages do no sum to 100.

4.7 Access Provisions

As shown in Table 7, most web sites reported providing some access provisions for individuals wishing to find out what data of theirs was in a web site's records. 92% of sites collecting identified data reported providing some access provisions. Most sites reported providing access to both contact information as well as some other data. A smaller number reported providing access to only contact information or to all identified data. Very few of the web sites that reported collecting identified data indicated that they provided no access, and none indicated that they provided access only to non-contact information.

Table 7. Access Provisions at P3P-Enabled Web Sites: Percentage of Sites Offering Each Type of P3P Access Provision

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
All identified data	9	16	19	15	26	15	18	12	21	17	5	13
Identified contact information and other identified data	70	64	61	70	53	57	53	55	46	50	55	52
Identified contact information	13	4	3	11	7	5	3	9	4	0	25	14
None	0	0	0	0	2	6	5	11	8	17	6	7
Web site does not collect identified data	9	16	16	4	12	16	22	13	21	17	9	14
Other identified data	0	0	0	0	0	0	0	0	0	0	0	0

Note, definitions of access types can be found in the P3P 1.0 specification [6].

4.8 Dispute Resolution Options and Remedies

As shown in Table 8, most web sites reported offering some dispute resolution option for disputes related to their privacy policy. In addition, most offered some sort of remedies. Most of these sites indicated that individuals could contact customer service to resolve their disputes. About one-third also offered to resolve the dispute via an independent organization such as a privacy seal provider. Very few indicated that disputes could be resolved under an applicable law and almost none indicated that they could be resolved in court. The most popular sites were most likely than other sites to offer to resolve disputes via an independent organization.

Table 8. Dispute Resolution Options and Remedies Offered by P3P-Enabled Web Sites: Percentage of Sites Offering Each Type of P3P Dispute Resolution Option

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Court	0	0	0	0	0	0	0	2	0	0	0	0
Independent organization	57	28	23	63	38	42	46	36	42	17	32	34
Applicable law	0	8	6	0	3	4	5	6	4	17	2	4
Customer service	87	88	87	70	72	73	78	78	71	83	71	75
Remedies offered	91	76	77	89	77	78	76	73	75	67	89	81

Note, definitions of dispute resolution options can be found in the P3P 1.0 specification [6]. Sites may offer multiple options so percentages do no sum to 100.

4.9 Data Retention Policies

As shown in Table 9, the majority of web sites reported that they did not have a data retention policy for all of the data they collected. Those that reported that they had a data retention policy were most likely to cite a policy based on their business practices, however, some indicated that they retained data only as long as necessary for the stated purpose or as required by law. A small number of sites indicated that they did not retain information. Government web sites were more likely than other sites to have a policy of not retaining information or to have a retention policy based on a legal requirement.

Table 9. Data Retention Policies at P3P-Enabled Web Sites: Percentage of Sites Offering Each Type of P3P Data Retention Policy

	PFF Most Popular	PFF Refined Random	PFF Random	Netscore Top 100	Netscore Top 500	Key Measures	Alexa	Froogle	Yahooligans	Firstgtov	News	Combined
Determined by the service provider's business practices	35	40	45	41	42	40	43	35	17	17	67	49
Indefinitely	65	48	48	70	72	70	64	71	83	33	63	65
As required by law or liability under applicable law	0	8	6	0	3	4	3	1	4	50	0	2
Information is not retained	4	0	0	0	7	5	4	11	4	33	3	6
For the stated purpose	13	28	26	4	5	6	5	14	17	0	2	8

Note, definitions of data retention policy types can be found in the P3P 1.0 specification [6]. Sites may have different policies for different types of data so percentages do no sum to 100.

4.10 Other Assessments

The human-readable <CONSEQUENCE> element is an optional component of a P3P policy that allows sites to provide further explanations about their data practices. Three-quarters of the sties we studied included this element at least once in their P3P policies. Some P3P user agents, such as AT&T Privacy Bird, display this element to users while others, such as IE6, do not.

Six percent of the P3P-enabled web sites we studied used the P3P <NON-IDENTIFIABLE> element to indicate that they do not collect data or that some or all of the data they collect is immediately anonymized. Given the stringent requirements for use of this element, we suspect that some of these sites may be using it incorrectly. It would be useful to check the human-readable privacy policies at each of these sites to see if they are consistent with the use of the <NON-IDENTIFIABLE> element.

5. DISCUSSION AND FUTURE WORK

Our study has demonstrated the feasibility of automated analysis of P3P-enabled web sites and presented the first major analysis of P3P adoption. We plan to repeat our experiments on a regular basis to allow for longitudinal analysis of P3P policies. In the future we may also expand the list of web sites we analyze, develop additional APPEL rule sets to facilitate more detailed analysis, and expand our analysis to include P3P compact policies.

Due to the differences in methodology between our study and the web sweeps, as well as the fact that the last web sweeps were conducted over a year before our study [1], we cannot directly compare our results to determine whether our sample of P3P-enabled web sites have policies that are representative of the policies that would be found in a sample of both P3P-enabled and non-P3P-

enabled sites. It would be interesting to supplement our study with a manual analysis of a small sample of non-P3P-enabled sites to see how the policies of P3P-enabled sites and non-P3P-enabled sites compare.

As debates continue about the need for further privacy legislation and the effectiveness of industry self-regulation in the privacy area, it is essential to have good statistics about privacy policies. As more web sites adopt P3P, it will be possible to increasingly automate the process of collecting these statistics, making more frequent and detailed "web sweeps" studies feasible. Furthermore, as US government web sites begin posting P3P policies to comply with the privacy requirements of section 208 of the E-Government Act of 2002 [16], we can continue to conduct sweeps of government web sites to monitor compliance with these requirements.

One interesting finding of our study was that the sites referenced by Yahooligans (and therefore recommended for children) were more likely to offer opt-in policies than other sites, but otherwise did not appear overall to have better privacy practices than other sites. It is not clear whether or not the P3P-enabled sites on the Yahooligans list are representative of sites designed for children. Indeed many of the sites on this list are not designed specifically as children's sites, although they do contain some content appealing to children. In addition, only 3% of the sites on the Yahooligans list had P3P policies, and these sites are not necessarily representative of the entire list. Nonetheless, given the US regulatory requirements for children's web sites, we believe it would be useful to examine the privacy practices of children's web sites in more detail in future studies.

One of the more surprising results of our study was the large number of web sites with technical errors in their P3P policies. This study highlights the need for site administrators to validate their P3P policies and keep them up to date. In the future we may study the types of errors in more detail. It would also be useful to perform a manual analysis of a small sample of P3P-enabled web sites, comparing human-readable policies with P3P policies to determine whether sites are making substantive errors in their P3P policies. It should be noted that errors in the implementation of web-related standards are fairly common. For example, over a year after the release of HTTP/1.1, a study found that a large number of web servers failed various compliance tests [21]. None-the-less, P3P errors arguably have more severe legal and policy-related consequence than errors in the implementation of the HTTP standard. While the former may result in less efficient web transactions and even occasional server crashing, the later may result in privacy policies being misrepresented and users being mislead [8]. If substantive errors are found or the error rate does not improve significantly over time, it may be necessary to explore the possibilities of third-party P3P policy certification, auditing, or other measures to ensure that P3P policies are trustworthy.

REFERENCES

[1] Adkinson, W.F., Eisenach, J.A., and Lenard, T.M. *Privacy online: A report on the information practices and policies of commercial web sites.* Progress & Freedom Foundation, Washington, DC, 2002. http://www.pff.org/publications/privacyonlinefinalael.pdf

[2] Cavoukian, A., and Hamilton, T.J. *The Privacy Payoff: How Successful Businesses Build Customer Trust.* McGraw-Hill Ryerson, Toronto, Ontario, 2002.

[3] Cranor, L. *Web Privacy with P3P.* O'Reilly & Associates, Sebastopol CA, 2002.

[4] Cranor, L. The role of privacy advocates and data protection authorities in the design and deployment of the platform for privacy preferences. In *Proceedings of the Twelth Conference on Computers, Freedom and Privacy* (San Francisco, CA, April 16-19, 2002) ACM Press. http://doi.acm.org/10.1145/543482.543506

[5] Cranor, L., Arjula, M., and Guduru, P. Use of a P3P User Agent by Early Adopters. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, (Washington, DC, November 2002) ACM Press.

[6] Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., and Reagle, J. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification.* World Wide Web Consortium Recommendation, April 2002. http://www.w3.org/TR/P3P/.

[7] Cranor, L., Langheinrich, M., and Marchiori, M. *A P3P Preference Exchange Language 1.0 (APPEL1.0).* World Wide Web Consortium Working Draft, April 2002. http://www.w3.org/TR/WD-P3P-Preferences.

[8] Cranor, L. and Reidenberg, J. Can user agents accurately represent privacy notices?. *TPRC 2002* (September 2002). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=328860

[9] Culnan, M.J. *The Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission.* Georgetwon University, Washington, DC, June 1999. http://www.msb.edu/faculty/culnanm/gippshome.html

[10] Culnan, M.J. *Privacy and the top 100 web sites: Report to the Federal Trade Commission.* Georgetown University, Washington, DC, June 1999. http://www.msb.edu/faculty/culnanm/gippshome.html

[11] Culnan, M.J. and Milne, G.R. *The Culnan-Milne survey of consumers and online privacy notices.* December 2001. http://intra.som.umass.edu/georgemilne/PDF_Files/culnan-milne.pdf

[12] Dhurvasula, H., Barrowman, D., and Morse, S. *Technical Issues in Implementing P3P in Netscape 7.0.* November 2002. http://www.w3.org/2002/p3p-ws/pp/netscape.html

[13] Ernst & Young. P3P Dashboard Report, August 2002. http://www.ey.com/global/download.nsf/US/P3P_Dashboard_-__August_2002/$file/P3PDashboardAugust2002.pdf

[14] Ernst & Young. P3P Dashboard Report, January 2003. http://www.ey.com/global/download.nsf/US/P3P_Dashboard_-_January_2003/$file/E&YP3PDashboardJan2003.pdf

[15] Esposito, D. Browser Helper Objects: The Browser the Way You Want It, MSDN Library, January 1999. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwebgen/html/bho.asp.

[16] Frank, D. OMB Honing Privacy Guidance. *Federal Computer Week* (14 March 2003). http://www.fcw.com/fcw/articles/2003/0310/web-guide-03-13-03.asp

[17] Federal Trade Commission. *Privacy online: A report to Congress.* Federal Trade Commission, Washington DC, June 1998. http://www.ftc.gov/reports/privacy3/index.htm

[18] Federal Trade Commission. *Self-regulation and privacy online: A report to Congress.* Federal Trade Commission, Washington DC, July 1999. http://www.ftc.gov/os/1999/9907/index.htm#13

[19] Federal Trade Commission. *Privacy online: Fair information practices in the elctronic marketplace: A report to Congress.* Federal Trade Commission, Washington DC, May 2000. http://www.ftc.gov/os/2000/05/index.htm#22

[20] Goldfeder, A. and Leibfried, L. Privacy in Internet Explorer 6. MSDN Library, October 2001. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnpriv/html/ie6privacyfeature.asp.

[21] Krishnamurthy, B. and Arlitt, M. PRO-COW: Protocol Compliance on the Web—A Longitudinal Study. In *Proceedings of Usenix Symposium on Internet Technologies and Systems, USITS 2001,* (March 2001) p. 109-122. http://www.usenix.org/events/usits01/krishnamurthy.html

[22] Milne, G.R. and Culnan, M.J. Using the Content of Online Privacy Notices to Inform Public Policy: A Longitudinal Analysis of the 1998-2002 U.S. Web Surveys. *The Information Society 18,* 5 (October 2002), 345-359.

APPENDIX A
P3P-Enabled Sites as of 6 May 2003

The following is the list of 538 P3P-enabled web sites found by checking our list of 5,856 sites on 6 May 2003. Those sites that had errors that caused the Privacy Bird evaluation engine to be unable to process them are annotated with an asterisk (*).

101weddingbands.com
abcnews.go.com
ads.addesktop.com
afr.com
amos.catalogcity.com
asia.cnet.com
augustachronicle.com
biz.yahoo.com
builder.com.com
channels.gay.com
checkout.verisign.com
click.atdmt.com
clickserve.cc-dt.com
disney.go.com
download.cnet.com
dreamscenes.net
espn.go.com
fox17.trb.com
fox40.trb.com
gifts.uget.us
* images.thebabyoutlet.com
launch.yahoo.com
littlemachineshop.com
massbaytrading.com
moneycentral.msn.com
msnbc.com
news.com.com
news.ft.com
news.ninemsn.com.au
newsobserver.com
* pc.ign.com
prints.artselect.com
salisburypost.townnews.com
sg.biz.yahoo.com
shop.avon.com
shopping.discovery.com
sports.espn.go.com
startheatres.moviefone.com
story.news.yahoo.com
the.standard.net.au
thepittsburghchannel.com
tomswiftpage.tripod.com
uk.towerrecords.com
valkyriepub.tripod.com
* wire.ign.com
www.00fun.com
www.1800flowers.com
* www.1800ussearch.com
www.1stblaze.com
www.2beadornot2bead.com
www.4outdoorfun.com
www.aaa.com
www.aardbargain.com
www.abcnews.com
www.abcnews.go.com
www.about.com
www.accountingweb.com
www.aclens.com
www.acop.com
www.adweek.com
www.alexblake.com
www.allergybuyersclubshopping.com
www.allmyhome.com
www.allposters.com
www.allslots.com
www.allyoucanink.com
* www.altavista.com
www.altrec.com
* www.aluminumbats.com
www.amnews.com
www.ananova.com
www.ancestry.com
www.aneki.com
www.angelfire.com
www.anywho.com
www.atdiscount.com
www.atdmt.com
www.att.com
www.att.net
www.attbi.com
www.attws.com
www.augustachronicle.com
www.autotrader.com
www.az.gov
www.babyage.com
www.babyuniverse.com
www.backstage.com
www.baldmountaincoffee.com
www.bbbonline.com
www.bcentral.com
www.beastwars.com
www.beatricedailysun.com
www.benicianews.com
www.bhg.com
www.bidforassets.com
www.bigbeargrizzly.net
www.biglobe.ne.jp
www.bigspringherald.com
www.biomedcentral.com
www.bismarcktribune.com
www.bits.com
* www.bizrate.com
www.blackamateurpages.com
www.bladenjournal.com
www.blair.com
www.blesstheday.com
www.blocket.se
www.bluenile.com
www.boatersworld.com
www.bozemandailychronicle.com
www.brainbashers.com
* www.brainpop.com
www.bravenet.com
* www.britishimports.com
www.btnmag.com
www.buddyblankies.com
www.budplant.com
* www.buffalonews.com
www.burstnet.com
www.businessweek.com
www.cadillacnews.com
www.calendarlive.com
www.callawaygolfpreowned.com
www.campingworld.com
www.campmor.com
www.careerbuilder.com
www.carprices.com
www.catholicstore.com
* www.celebrateexpress.com
www.celebritywonder.com
www.cellular-news.com
www.centennialcard.com
www.channel3000.com
www.channel4000.com
www.channelcincinnati.com
www.channeloklahoma.com
www.chicagotribune.com
www.chpower.com
www.christianitytoday.com
www.chtah.com
www.circuitcity.com
www.click10.com
www.click2houston.com
www.clickbank.com
www.clickondetroit.com
www.clickonsa.com
www.clickzs.com
www.cnet.com
www.coach.com
* www.coastalcontacts.com
www.coloradodaily.com
www.columbustelegram.com
www.com.com
www.compusa.com
www.computers4sure.com
www.connectionzone.com
www.cooking.com
www.cordeledispatch.com
www.corel.com
www.crateandbarrel.com
www.ctnow.com
www.cumberlink.com

www.cupviews.com
www.cyclestuffusa.com
www.daedalus-books.com
www.dailyillini.com
www.dailypress.com
www.dailyrepublic.com
www.dans.com
www.daum.net
www.davisenterprise.com
www.dell.com
www.democratherald.com
www.dickssportinggoods.com
www.digitalriver.com
www.directron.com
www.discoverthis.com
www.disney.com
www.doubleclick.net
www.drugstore.com
www.dunnconnect.com
www.dvdadvantage.com
www.eangler.com
www.ebags.com
www.ecommercetimes.com
www.economist.com
www.edailynews.info
www.ediets.com
www.ehobbies.com
www.elmundo.es
www.eluxury.com
www.elynews.com
www.emode.com
www.enterprise.com
www.entertainment.com
www.espn.com
www.ethnicgrocer.com
www.etoys.com
www.etronics.com
www.eveningtimes.com
www.everyone.net
www.exitexchange.com
www.exodustrading.com
www.expedia.com
www.fastclick.net
* www.fdic.gov
www.fidelity.com
www.figleaves.com
www.flowgo.com
www.focalex.com
www.focuspools.com
www.ford.com
www.forddirect.com
www.fortunecity.com
www.franklincovey.com
www.fredericks.com
www.freeserve.com
www.fremontneb.com
www.friendsreunited.co.uk
www.frii.com
www.frontiersman.com
www.ftc.gov
www.fye.com
www.gaiam.com
* www.gameshark.com
www.gap.com
www.gardeners.com
www.geocities.com
www.gift-clocks.com
www.giftstrain.com
www.gifttree.com
www.girlpower.gov
www.globes.co.il
www.globetechnology.com
www.goanacortes.com
www.goclick.com
www.godiva.com
www.gotlaughs.com
www.governmentjobs.com
www.greatoutdoorsdepot.com
www.greenwichtime.com
www.handango.com
www.havredailynews.com
www.hbo.com
www.headandshoulders.com
* www.headshop.com
www.heraldandnews.com
www.hickoryfarms.com
* www.highschoolalumni.com
www.hollywoodreporter.com
www.hotjobs.com
www.hotmail.com
www.hp.com
www.ibm.com
www.ibsys.com
www.ice.com
* www.ign.com
www.illawarramercury.com.au
www.incredimail.com
www.infomaster.co.kr
www.inphonic.com
* www.insightexpress.com
www.intelligentx.com
www.isize.com
www.jackpot.com
www.jackpotmadness.com
www.jackpotsinaflash.com
www.jacksonandperkins.com
www.jacksonvilleprogress.com
www.jandr.com
www.jessicalondon.com
www.journaltimes.com
* www.jubii.dk
www.juliantrubin.com
www.justmysize.com
www.justsaywow.com
www.kbtoys.com
www.keen.com
www.kennedy-center.org
www.kickosama.com
www.kiss.com
www.kleptomaniac.com
www.ktul.com
www.lacrossetribune.com
www.landware.com
www.latimes.com
www.lebanonenterprise.com
www.lgeshop.com
www.llbean.com
www.local6.com
www.lodinews.com
www.losaltosonline.com
www.lotte.com
www.lowcostprints.com
* www.lycos.co.uk
www.lycos.com
* www.lycos.de
* www.lycos.fr
www.m0.net
www.macombjournal.com
www.madblast.com
www.magazineline.com
www.mailbits.net
www.mapblast.com
www.marriott.com
www.match.com
www.matchmaker.com
www.mcafee.com
www.mediainfo.com
www.memoryessentials.com
* www.metareward.com
www.microsoft.com
* www.military.com
www.missoulian.com
www.misweb.com
www.moberlymonitor.com
www.moendepot.com
www.moneycentral.com
www.motherjones.com
www.moultrieobserver.com
www.moviemom.com
www.msn.be
www.msn.co.kr
www.msn.com.br
www.msn.com
www.msn.fr
www.msnbc.com
www.msnusers.com
* www.musicstack.com
www.mydjconnection.com
www.myfree.com
www.mysurvey.com
www.namaste.com
www.nationwide.co.uk
www.natlallergy.com
www.naturalreflections.com
www.nbc13.com
www.nbc30.com
www.nbc4.com
www.nbc4.tv
www.nbc4columbus.com
www.nbc5.com
www.nbc5i.com
www.nbc6.net
www.nbcsandiego.com
www.netflip.com
www.netflix.com
www.netgrocer.com
www.netimperative.com
www.newalbanygazette.com
www.newbalancewebexpress.com
www.news-observer.com
www.news4jax.com
www.newsday.com

www.newsnet5.com
www.newszap.com
www.newzcentral.com
www.nifty.com
www.nifty.ne.jp
www.nordictrack.com
www.novica.com
www.nupplegal.com
www.nwfusion.com
www.nynewsday.com
www.odordestroyer.com
www.officedepot.com
* www.officemax.com
www.offshoreclicks.com
www.okmulgeetimes.com
www.oldnavy.com
www.onehanesplace.com
www.onet.pl
www.onjava.com
www.openp2p.com
www.oriental.com
www.orlandosentinel.com
www.oshmans.com
www.otxresearch.com
www.outdoorsuperstore.com
www.overstock.com
www.oxygen.com
www.paloalto.com
www.palossports.com
* www.passport.com
* www.passport.net
www.payless.com
www.paypopup.com
www.pch.com
www.pdns.com
www.peelworld.com
www.petco.com
www.pilotonline.com
www.pioneerlocal.com
www.planetout.com
www.polo.com
www.portervillerecorder.com
* www.powells.com
www.pressdemocrat.com
www.pricegrabber.com
www.progressive.com
www.qksrv.net
www.questionmarket.com
www.qvc.com
www.register-herald.com
www.rei.com
www.saksfifthavenue.com
www.savannahbusiness.com
www.sayclub.com
www.search.com
www.shopathome.com
www.signonsandiego.com
www.simplycheap.com
www.skyauction.com
www.skynet.be
www.slegg-tools.com
www.smartbargains.com
* www.smarterliving.com
www.smartwareetc.com
www.smh.com.au
www.solutions4sure.com
www.soyunica.gov
www.sparco.com
* www.spray.se
www.stacksandstacks.com
www.stamfordadvocate.com
www.staples.com
www.startribune.com
www.state.va.us
www.successmtgs.com
www.sun-sentinel.com
www.sunspot.net
www.suntimes.com
www.superpages.com
www.suwanneedemocrat.com
www.sweepsclub.com
www.t1msn.com.mx
www.targetnet.com
www.tdn-net.com
www.techdepot.com
www.technologymarketing.com
www.techrepublic.com
www.teenhollywood.com
www.teenmusic.com
www.texarkanagazette.com
www.theage.com.au
* www.thebabyoutlet.com
www.thebakersfieldchannel.com
www.thebostonchannel.com
www.thecarolinachannel.com
www.thechamplainchannel.com
www.thedenverchannel.com
www.thehawaiichannel.com
www.thehomemarketplace.com
www.thehometownchannel.com
www.theindychannel.com
www.theiowachannel.com
www.thejacksonchannel.com
www.thekansascitychannel.com
www.thekcrachannel.com
www.theksbwchannel.com
www.thelouisvillechannel.com
www.themilwaukeechannel.com
www.thenewmexicochannel.com
www.theneworleanschannel.com
www.thenewsenterprise.com
www.theomahachannel.com
www.thepittsburghchannel.com
www.therecordherald.com
www.thesandiegochannel.com
www.thesportsauthority.com
www.thesupplenet.com
www.theuseful.com
www.thewbalchannel.com
www.thewgalchannel.com
www.thewmurchannel.com
www.theworldlink.com
www.thewpbfchannel.com
www.ticketmaster.com
www.tiftongazette.com
www.tigerdirect.com
www.timesunion.com
* www.tirerack.com
www.towerrecords.com
www.tradedoubler.com
www.tripod.com
www.troymall.com
www.truckstuffusa.com
www.truste.org
www.turnto10.com
www.udate.com
www.uhome.net
www.uline.com
www.usatoday.com
www.usps.com
www.usps.gov
* www.ussearch.com
www.valdostadailytimes.com
www.vallejonews.com
www.vipnet.org
www.vistaprint.com
www.vitacost.com
www.walla.co.il
www.walmart.com
www.warnerbros.com
www.webmd.com
www.webpower.com
www.webstat.com
www.whtm.com
www.wildjack.com
www.windows.com
www.windowsmedia.com
www.winonadailynews.com
www.wnbc.com
www.workingforchange.com
* www.worldwinner.com
www.wral.com
www.wset.com
www.x10.com
www.xml.com
www.xuppa.com
www.ya.com
www.yahoo.co.jp
www.yahoo.co.kr
www.yahoo.co.uk
www.yahoo.com
www.yellowpages.com
www.youwintrivia.com
www.yupimsn.com
* www.zap2it.com
www.zdnet.com.au
www.zdnet.com
www.zone.com
www1.internetwire.com
www1.ritzcamera.com
www1.storehost.com
www2.warnerbros.com
xtramsn.co.nz
zdnet.com.com

APPENDIX B
Sites with P3P Compact Policies and No Full P3P Polices as of 6 May 2003

The following is the list of 74 web sites with P3P compact policies but no full P3P policies found by checking our list of 5,856 sites on 6 May 2003.

abclocal.go.com
canada.com
english.aljazeera.net
home.netscape.com
home.nzcity.co.nz
shop.microsoft.com
slate.msn.com
soccernet.espn.go.com
tv.zap2it.com
www.bankofamerica.com
www.buy.com
www.canada.com
www.cjb.net
www.cjmall.com
www.dearyou.com
www.digitalcity.com
www.digitalstorefronts.com
www.edumoa.com
www.ew.com
www.fandango.com
www.fastmetasearch.com
www.fieldandstream.com
www.fileplanet.com
www.fortune.com
www.gamespy.com
www.gateway.com
www.genericgifts.com
www.giftcollector.com
www.hanafos.com
www.herald-sun.com
www.hmall.com
www.hotbot.com
www.hot.co.kr
www.icq.com
www.iloveschool.co.kr
www.interpark.com
www.jcpenney.com
www.lakecityreporter.com
www.localsponsors.com
www.mailcity.com
www.mapquest.com
www.mgame.com
www.mircx.com
www.moviefone.com
www.movies.com
www.msn.com.tw
www.netmarble.net
www.netscape.com
www.nexon.com
www.nexternal.com
www.ninemsn.com.au
www.ohiopurewaterco.com
www.okcashbag.com
www.overture.com
www.pier1.com
www.popupnation.com
www.poynter.org
www.qwestdex.com
www.real.com
www.realnetworks.com
www.rosiesbakery.com
www.siren24.com
www.skiingmag.com
www.superboard.com
www.themercury.com
www.time.com
www.timeforkids.com
www.transworldsnowboarding.com
www.twistedhumor.com
www.victoriassecret.com
www.vstore.com
www.wired.com
www.zoomerang.com
www.ztelligence.com