The Platform for Privacy Preferences

Web Privacy *with* P3P

O'REILLY®

Lorrie Faith Cranor

Web Privacy with P3P

Lorrie Faith Cranor

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

CHAPTER 1

Introduction to P3P

Internet users are becoming increasingly concerned about what personal information they may reveal when they go online and where that information might end up. It's common to hear about companies that derive revenue from personal information collected on their web sites. Information you provide to register for a site might later be used for telemarketing or sold to another company. Seemingly anonymous information about your web-surfing habits might be merged with your personal information. Web sites use cookies to gather information about users, but disabling cookies prevents you from doing online banking or shopping at some web sites.

Web sites might email you to say that their privacy policies are changing, but most of us find it difficult and time-consuming to read and understand privacy policies or to figure out how to request that the use of our personal information be restricted. Privacy concerns are making consumers nervous about going online, but current privacy policies for web sites tend to be so long and difficult to understand that consumers rarely read them.*

The Platform for Privacy Preferences (P3P) project addresses this problem by providing both a standard, computer-readable format for privacy policies and a protocol that enables web browsers to read and process privacy policies automatically. The World Wide Web Consortium (W3C) developed P3P as a standard way for web sites to communicate about their privacy policies. P3P enables machine-readable privacy policies that can be retrieved automatically by web browsers and by other user agent tools that can display symbols, prompt users, or take other appropriate actions. Some of these tools can also compare each policy against the user's privacy preferences and assist the user in deciding when to exchange data with web sites.

Unlike anonymity tools, which seek to prevent any transfer of personally identifying information, the P3P project seeks to enable the development of tools for making informed decisions about when and if personal information should be revealed.

* Privacy Leadership Initiative, "Privacy Notices Research Final Results" (conducted by Harris Interactive, December 2001), *http://www.ftc.gov/bcp/workshops/glb/supporting/harris%20results.pdf*.

These tools may work hand-in-hand with anonymity software or filters that actually prevent the transmission of personal information in situations when users do not want their information revealed.

P3P tools are currently available that allow users to configure their web browsers with their personal privacy preferences. P3P-enabled web browsers check for P3P privacy policies at web sites and display symbols to alert users at sites that do not match their preferences. They can also provide summaries of web site privacy policies and use P3P policies to make decisions about cookies.[*]

How P3P Works

The Platform for Privacy Preferences 1.0 (P3P1.0) Specification is the authoritative source for information on the P3P protocol and vocabulary. Throughout this book, I generally refer to it simply as the "P3P specification." You can retrieve the specification from *http://www.w3.org/TR/P3P/*.

P3P was developed through a consensus process involving several dozen W3C working group members. Participants came from around the world and included representatives from industry, government, nonprofit organizations, and academia. In addition, public comments on the many P3P working drafts helped shape the final P3P specification. This section gives a brief summary of how P3P works.

Privacy policies are intended to describe a company's *data practices*—what they do with the information they collect from individuals (usually customers and potential customers, but sometimes also employees and others). The P3P specification includes a standard *vocabulary* for describing these data practices and a *base data schema* for describing the kinds of information collected. A P3P *policy* is a collection of vocabulary and data elements that describes the data practices of a particular web site (or section of a web site).

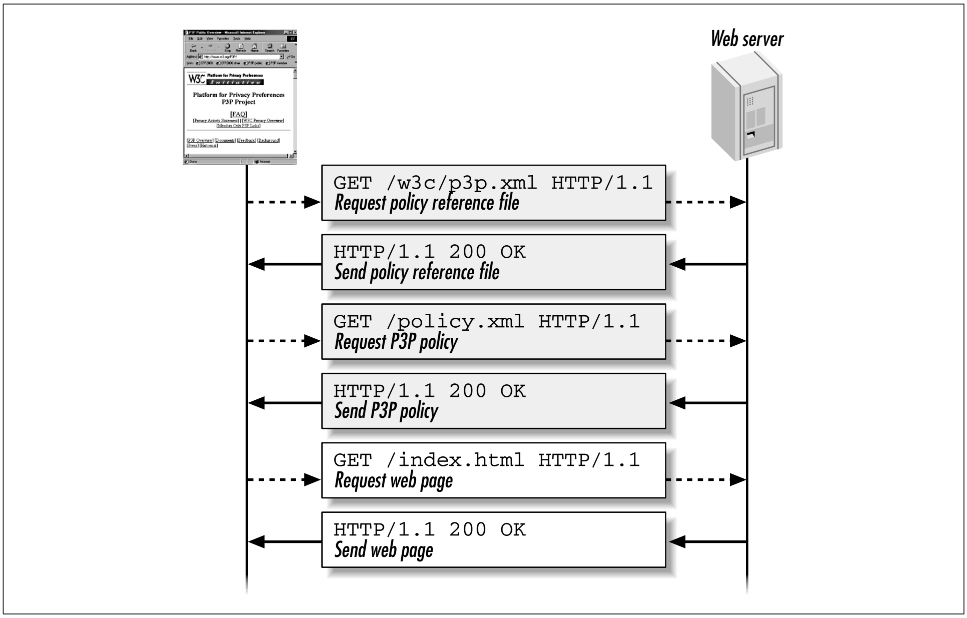
A P3P policy is essentially composed of the answers to a number of multiple-choice questions and thus does not always contain as much detailed information as a human-readable privacy policy (i.e., a policy written in English or another spoken language that is intended for people, rather than computers, to read). The standard format of a P3P policy allows it to be processed automatically.

The P3P specification also includes a protocol for requesting and transmitting P3P policies. The P3P protocol is built on the same HTTP protocol[†] that web browsers

[*] Cookies are bits of text that web sites can send in their HTTP headers and ask web browsers to send back to them on subsequent visits to the same web site. They help enable features such as electronic shopping carts and logging into a web site without a password. Cookies are discussed in more detail in Chapter 2.

[†] HTTP is short for HyperText Transfer Protocol (*http://www.ietf.org/rfc/rfc2616.txt*). For in-depth information on how HTTP and related protocols work, see Balachander Krishnamurthy and Jenifer Rexford, *Web Protocols and Practice: HTTP/1.1, Networking Protocols, Caching, and Traffic Measurement* (Boston: Addison Wesley, 2001).

use to communicate with web servers. As shown in Figure 1-1, P3P user agents 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. The user agent can then fetch the appropriate policy, parse it, and take action according to the user's preferences.



`GET /w3c/p3p.xml HTTP/1.1`
Request policy reference file

`HTTP/1.1 200 OK`
Send policy reference file

`GET /policy.xml HTTP/1.1`
Request P3P policy

`HTTP/1.1 200 OK`
Send P3P policy

`GET /index.html HTTP/1.1`
Request web page

`HTTP/1.1 200 OK`
Send web page

Figure 1-1. The basic protocol for fetching a P3P policy

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 by 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.

Here's a plain English example of the kind of disclosure a web site might make in a P3P policy:

> We do not currently collect any information from visitors to this site except the information contained in standard web server logs (your IP address, referer, information about your web browser, information about your HTTP requests, etc.). The information in these logs will be used only by us and the server administrators for web site and system administration, and for improving this site. It will not be disclosed unless required by law. We may retain these log files indefinitely. Please direct questions about this privacy policy to *privacy@p3pbook.com*.

And here's what this policy would look like using the P3P syntax and encoding:

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
<POLICY discuri="http://p3pbook.com/privacy.html"
        name="policy">
  <ENTITY>
  <DATA-GROUP>
    <DATA
      ref="#business.contact-info.online.email">privacy@p3pbook.com
    </DATA>
    <DATA
      ref="#business.contact-info.online.uri">http://p3pbook.com/
    </DATA>
    <DATA ref="#business.name">Web Privacy With P3P</DATA>
  </DATA-GROUP>
  </ENTITY>
  <ACCESS><nonident/></ACCESS>
  <STATEMENT>
    <CONSEQUENCE>We keep standard web server logs.</CONSEQUENCE>
    <PURPOSE><admin/><current/><develop/></PURPOSE>
    <RECIPIENT><ours/></RECIPIENT>
    <RETENTION><indefinitely/></RETENTION>
    <DATA-GROUP>
       <DATA ref="#dynamic.clickstream"/>
       <DATA ref="#dynamic.http"/>
    </DATA-GROUP>
  </STATEMENT>
</POLICY>
</POLICIES>
```

If you are familiar with the eXtensible Markup Language (XML), this encoding may look familiar to you. But if not, don't worry! I'll get to the details of writing policies in later chapters and introduce you to some tools you can use to create policies without having to write any XML yourself. It is also important to note that P3P policies are designed to be read by computers, not people. User agents will interpret these policies on a user's behalf. In addition, every policy should contain the URL of the web site's human-readable privacy policy, so that users have someplace to turn for more detailed information.

The example policy above is fairly brief, because this web site does not collect much information from visitors. Commercial web sites typically have lengthier policies that describe their more complicated data practices.

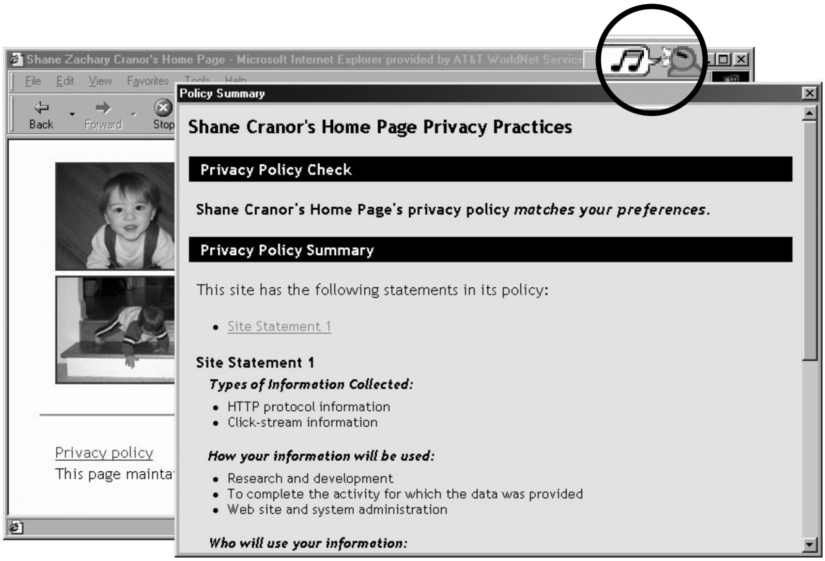
P3P user agents typically allow users to specify their privacy preferences so that they can automatically compare a web site's policies to these preferences. P3P user agents can also provide tools that make it easier for users to quickly assess a site's privacy practices for themselves. Some user agents display symbols that summarize a site's privacy policy or indicate that it has a privacy seal (a certification that the site follows its stated privacy policy and/or complies with some set of privacy standards) or is bound by certain privacy laws. Some user agents also include buttons that load a site's human-readable privacy policy without the users having to search for it on the site.

User Agents

The P3P specification rarely uses the terms *browser* or *client*; instead the term *user agent* is used. Although end-user P3P implementations might naturally be built into web browsers, P3P implementations can also be built into electronic wallets, standalone applications, ISP software, or other tools. Thus, the more general term "user agent" is used in the specification and in many places throughout this book.

The P3P specification places few requirements on user agents, so what P3P user agents do varies considerably. This book contains descriptions of several P3P user agents and a variety of possible user agent functions.

Figure 1-2 shows an example of the kind of information displayed by one P3P user agent, the AT&T Privacy Bird beta 1.1 (*http://privacybird.com)*. The AT&T Privacy Bird displays a green, "happy" bird icon at sites with P3P policies that match a user's privacy preferences and a red, "angry" bird icon at sites with P3P policies that do not match a user's preferences. Users can click on the bird icon to view a summary of the site's privacy policy that is generated automatically from the site's P3P policy. At sites that do not match a user's preferences, the policy summary also explains where the policy differs from the user's preferences.


Shane Zachary Cranor's Home Page - Microsoft Internet Explorer provided by AT&T WorldNet Service
File
Edit
View
Favorites
Tools
Help
Back
Forward
Stop
Privacy policy
This page mainta
Policy Summary

Shane Cranor's Home Page Privacy Practices

Privacy Policy Check

Shane Cranor's Home Page's privacy policy *matches your preferences*.

Privacy Policy Summary

This site has the following statements in its policy:

- Site Statement 1

Site Statement 1

Types of Information Collected:

- HTTP protocol information
- Click-stream information

How your information will be used:

- Research and development
- To complete the activity for which the data was provided
- Web site and system administration

Who will use your information:

Figure 1-2. The AT&T Privacy Bird displays a green bird icon at sites that match a user's privacy preferences; users can click on the bird to obtain a summary of the site's privacy policy

Another P3P user agent, the Microsoft Internet Explorer 6 web browser, automatically checks P3P compact policies at sites that set cookies. Users can configure IE6 to filter cookies that do not have compact policies or that have compact policies that do not match their preferences. IE6 displays an "eye" symbol in the bottom right corner of the browser window when cookies are blocked. Users can also select the "Privacy Report" option from the View menu to have IE6 fetch a site's P3P policy and generate and display a human-readable version.

In May 2002 Netscape released the preview version of its Netscape Navigator 7 software, which includes P3P functions similar to those found in IE6. Users can configure Netscape to filter cookies on the basis of their P3P compacy policies. They can also select Page Info from the View menu and go to the Privacy tab to have Netscape fetch a site's P3P policy and generate and display a human-readable version.

While the IE6 and Netscape P3P implementations are good first steps that are helping stimulate P3P adoption, they make cookie-filtering decisions based on compact policies only; they do not base these decisions on full P3P policies. Hopefully, in the future Microsoft and Netscape will offer configuration options that take advantage of full P3P policies.

Chapter 12 discusses a variety of ideas for P3P user agents. For example, a P3P user agent might be built into an electronic wallet or other software that includes a data repository that stores data users frequently exchange with web sites. The data in this repository might be identified by the standard names defined in the P3P base data schema. Before automatically filling out a form or submitting data on behalf of a user, a P3P-enabled electronic wallet might fetch the relevant P3P policy and compare it with the user's preferences. If a site does not have a P3P policy or has a policy that does not match the user's preferences, the wallet can alert the user. The wallet might also automatically create and fill out forms with requested data, annotating the forms with the site's data practices.

P3P also has a standard language for encoding a user's privacy preferences, called *A P3P Preference Exchange Language* (APPEL). APPEL files specify what actions the user agent should take, depending on the types of disclosures made by a web site. APPEL files are used by P3P user agents—they are not intended to be sent to web sites. APPEL is not designed to be read by end users either; it is useful mostly for organizations—such as privacy advocacy groups, privacy seal providers, or governmental privacy agencies—that don't like the default settings that come with P3P user agents and want to develop their own "canned" P3P configuration files to distribute to users. It also enables users who have found a configuration setting they like to export it from one user agent and import it into another. However, not all P3P user agents include the ability to import and export APPEL files. The APPEL files themselves are encoded in XML, just like P3P policies. The details of writing APPEL files are discussed in Chapter 13.

P3P-Enabling a Web Site

P3P-enabling a web site is usually a fairly easy process, from a technical standpoint. However, it may require web site operators to take a more detailed look at their data practices than they have done previously and to coordinate policies and practices across the hosts in their domains. Here is an overview of the steps required to P3P-enable a web site. Part II of this book details this entire process.

1. Create a privacy policy.
2. Analyze the use of cookies and third-party content on your site.
3. Determine whether you want to have one P3P policy for your entire site or different P3P policies for different parts of your site.
4. Create a P3P policy (or policies) for your site.
5. Create a policy reference file for your site.
6. Configure your server for P3P.
7. Test your site to make sure it is properly P3P-enabled.

Most P3P-enabled web sites end up with one P3P policy reference file on each of their servers and one or more P3P policies on a central server. They may also configure their servers to send a P3P compact policy whenever they set cookies.

P3P policies include the following information:

- Contact information for the business, organization, or person who owns the site
- Whether individuals can find out what personal data a site keeps about them in its databases
- How to resolve privacy-related disputes with the site (customer service desk, privacy seals, relevant privacy laws, etc.)
- The kinds of data collected
- How collected data is used, and whether individuals can opt-in or opt-out of any of these uses
- Whether/when data may be shared and whether there is opt-in or opt-out
- Policies for periodic purging of collected data

A variety of software tools are available to assist web site developers in P3P-enabling their sites. Some of these are described later in this book; however, for the most up-to-date lists of P3P tools, see *http://p3ptoolbox.org/tools/* and *http://www.w3.org/P3P/implementations/*.

Why Web Sites Adopt P3P

Since Microsoft released their P3P-enabled IE6 web browser in 2001, an increasing number of web sites have adopted P3P. A December 2001 survey by the Progress and

Freedom Foundation found that 23% of the most popular web sites and 5% of a random sample of the top 5,625 domains that collect personally identifiable data were P3P-enabled.[*] The authors of the report concluded: "This seems to be a fairly rapid rate of adoption, given the newness of the product and the fact that relatively few consumers have installed IE6."

By April 2002, about a third of the top 100 web sites had adopted P3P. Early adopters of P3P come from a variety of sectors and include:

- News and information sites, such as CNET and About.com
- Search engines, such as Yahoo! and Lycos
- Advertising networks, such as DoubleClick and Avenue A
- Telecommunications companies, such as AT&T
- Financial institutions, such as Fidelity
- Computer hardware and software vendors, such as IBM, Dell, Microsoft, and McAfee
- Retail stores, such as Fortunoff and Ritz Camera
- Government agencies, such as the U.S. Federal Trade Commission, the U.S. Department of Commerce, and the U.S. Postal Service
- Nonprofit organizations, such as the Center for Democracy and Technology
- Academic institutions, such as Vanderbilt University eLab

Sites outside the U.S. have also started adopting P3P, including commercial sites and the web sites of several data protection commissioners (for example, the Ontario Information and Privacy Commissioner and the Data Protection Commissioner of Bavaria, Germany).

Many early adopters P3P-enabled their web sites to show their support for the P3P effort and demonstrate their corporate leadership on privacy issues. They were motivated both by a desire to show customers that they respect their privacy and by a desire to demonstrate to regulators that the industry is taking voluntary steps to address consumer privacy concerns. While P3P addresses only a narrow set of privacy issues, it complements other efforts to improve privacy protections, including laws, technology tools, and privacy seal programs.

Some companies have started using privacy as a way of distinguishing their brand—they include privacy messages in their advertising and feature privacy-related aspects of their products. By adopting P3P, they further strengthen the message that they respect consumer privacy. In addition, by adopting P3P, they enable consumers to

[*] William F. Adkinson, Jr., Jeffrey A. Eisenach, and Thomas M. Lenard, "Privacy Online: A Report on the Information Practices and Policies of Commercial Websites" (Progress & Freedom Foundation, March 2002), *http://www.pff.org/publications/privacyonlinefinalael.pdf*. The web sites surveyed for this report were determined based on October 2001 Nielson/NetRatings data.

quickly locate and get a brief summary of their privacy policies, and to take advantage of any opportunities to remove themselves from marketing and mailing lists.

Some companies have adopted P3P in anticipation that it may soon become a standard that consumers look for at the web sites they visit. If consumers become accustomed to being able to request a privacy report from their web browser or to seeing a happy privacy-bird icon, they may grow suspicious of sites that are not P3P-enabled. In the future, P3P-enabled search engines may make it easy for consumers to identify P3P-enabled web sites.

Some companies have already found that their web sites do not function correctly when viewed using the latest web browsers if their sites are not P3P-enabled. By default, IE6 looks for P3P compact policies associated with third-party cookies (discussed in Chapter 2) on web sites. Third-party cookies are automatically blocked when they don't have compact policies. Thus, targeted advertising, page counters, and other features that rely on third-party cookies may not work unless companies P3P-enable their sites.

Finally, many web sites have adopted P3P because the individuals who run them value their personal privacy and want the companies they work for to take steps to give individuals more control over their personal information.