Machine Readability and P3P The Basics

Lorrie Faith Cranor

P3P Specification Working Group Chair Carnegie Mellon University

January 2004

http://lorrie.cranor.org/

Privacy policies

- Policies let visitors know about site's privacy practices
- Visitors can then decide whether or not practices are acceptable, when to opt-in or opt-out, and who to do business with
- The presence or privacy policies increases trust

Privacy policy problems

BUT policies are often

- *difficult to understand
- ★hard to find
- ★ take a long time to read
- ★ change without notice

Platform for Privacy Preferences Project (P3P)

- Developed by the World Wide Web Consortium (W3C) http://www.w3.org/p3p/
 - ★ Final P3P1.0 Recommendation issued 16 April 2002
- Offers an easy way for web sites to communicate about their privacy policies in a standard machine-readable format
 - ★ Can be deployed using existing web servers
- Enables the development of tools (built into browsers or separate applications) that
 - **★** Summarize privacy policies
 - ★ Compare policies with user preferences
 - ★ Alert and advise users

Basic components

- P3P provides a standard XML format that web sites use to encode their privacy policies
- Sites also provide XML "policy reference files" to indicate which policy applies to which part of the site
- Sites can optionally provide a "compact policy" by configuring their servers to issue a special P3P header when cookies are set
- No special server software required
- User software to read P3P policies called a "P3P user agent"

What's in a P3P policy?

- Name and contact information for site
- The kind of access provided
- Mechanisms for resolving privacy disputes
- 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
- Data retention policy

Compact policies

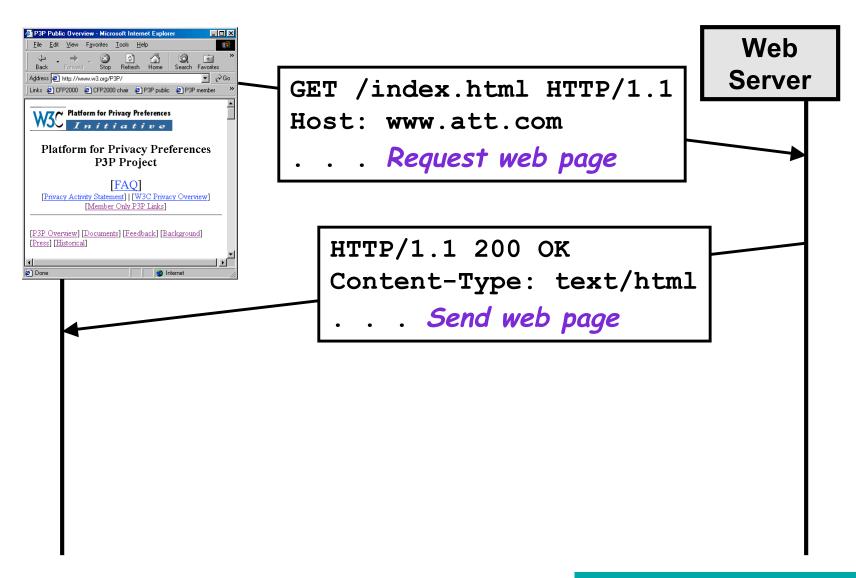
- Provide very short summary of full P3P policy for cookies
- Not required
- Must be used in addition to full policy
- May only be used with cookies
- Must commit to following policy for lifetime of cookies
- May over simplify site's policy
- IE6 relies heavily on compact policies for cookie filtering especially an issue for third-party cookies

Legal issues

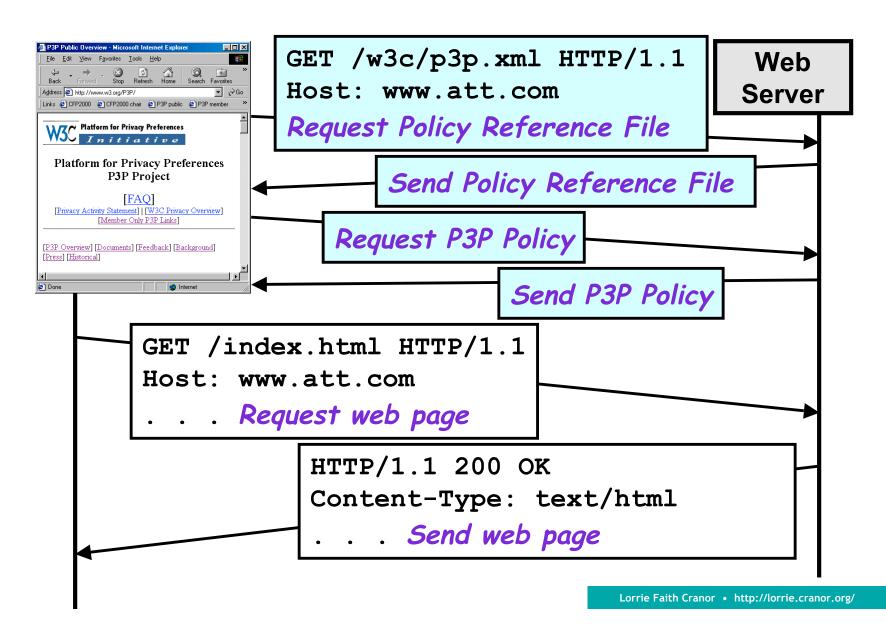
- P3P specification does not address legal standing of P3P policies or include enforcement mechanisms
- P3P specification requires P3P policies to be consistent with natural-language policies
 - ★P3P policies and natural-language policies are not required to contain the same level of detail
 - ★ Typically natural-language policies contain more detailed explanations of specific practices
- The same attorneys and policy makers involved in drafting natural-language privacy policy should be involved in creating P3P policy

Privacy policy	P3P policy
Designed to be read by a human	Designed to be read by a computer
Can contain fuzzy language with "wiggle room"	Mostly multiple choice - sites must place themselves in one "bucket" or another
Can include as much or as little information as a site wants	Must include disclosures in every required area
Easy to provide detailed explanations	Limited ability to provide detailed explanations
Sometimes difficult for users to determine boundaries of what it applies to and when it might change	Precisely scoped
Web site controls presentation	User agent controls presentation

A simple HTTP transaction



... with P3P 1.0 added



P3P increases transparency

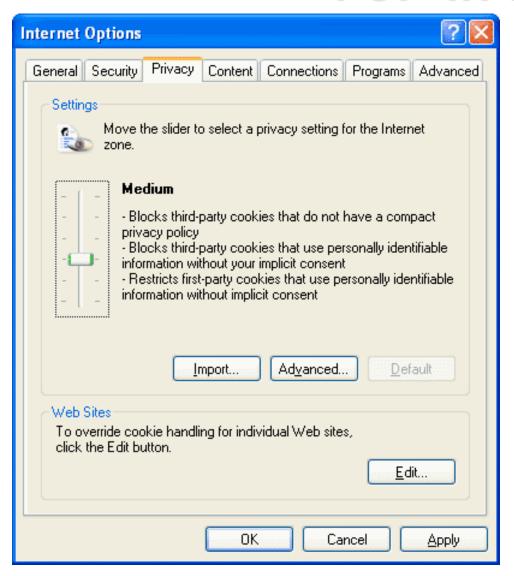
- P3P clients can check a privacy policy each time it changes
- P3P clients can check privacy policies on all objects in a web page, including ads and invisible images

http://www.att.com/accessatt/



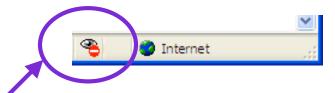
http://adforce.imgis.com/?adlink|2|68523|1|146|ADFORCE

P3P in IE6



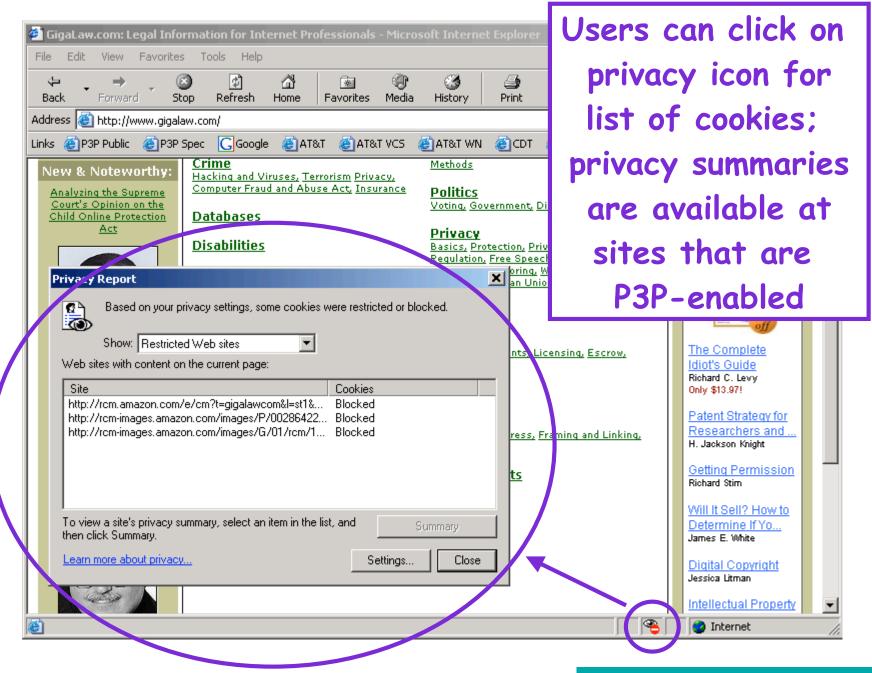
Automatic processing of compact policies only;

third-party cookies without compact policies blocked by default



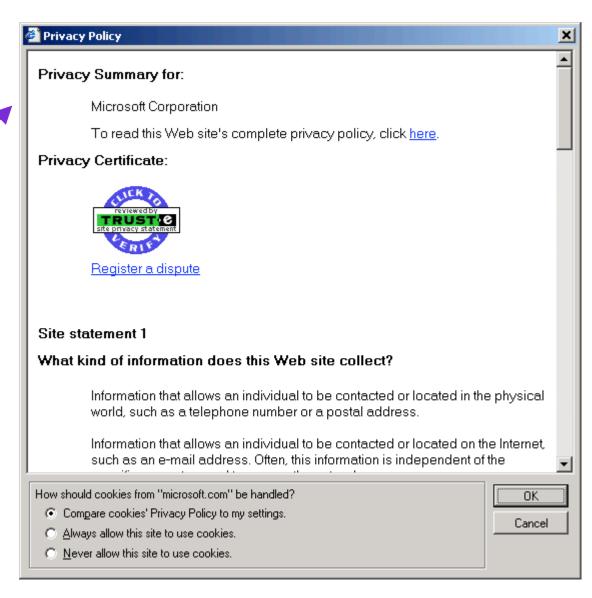
Privacy icon on status bar indicates that a cookie has been blocked - pop-up appears the first time the privacy icon appears



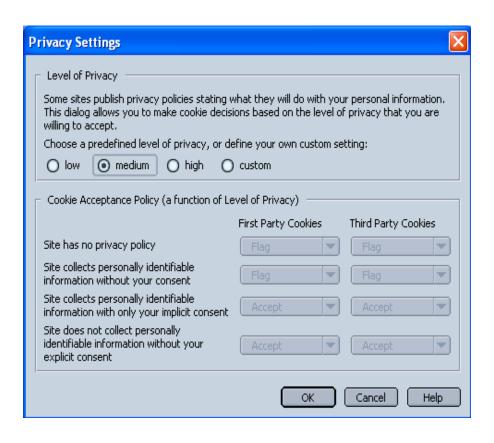




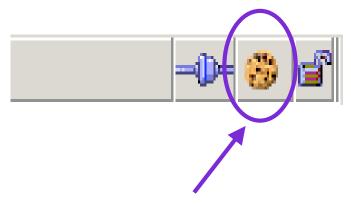
Privacy summary report is generated automatically from full P3P policy



P3P in Netscape 7



Preview version similar to IE6, focusing, on cookies; cookies without compact policies (both first-party and third-party) are "flagged" rather than blocked by default



Indicates flagged cookie

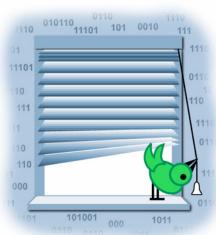
AT&T Privacy Bird

- Free download of beta from http://privacybird.com/
- "Browser helper object" for IE 5.01/5.5/6.0





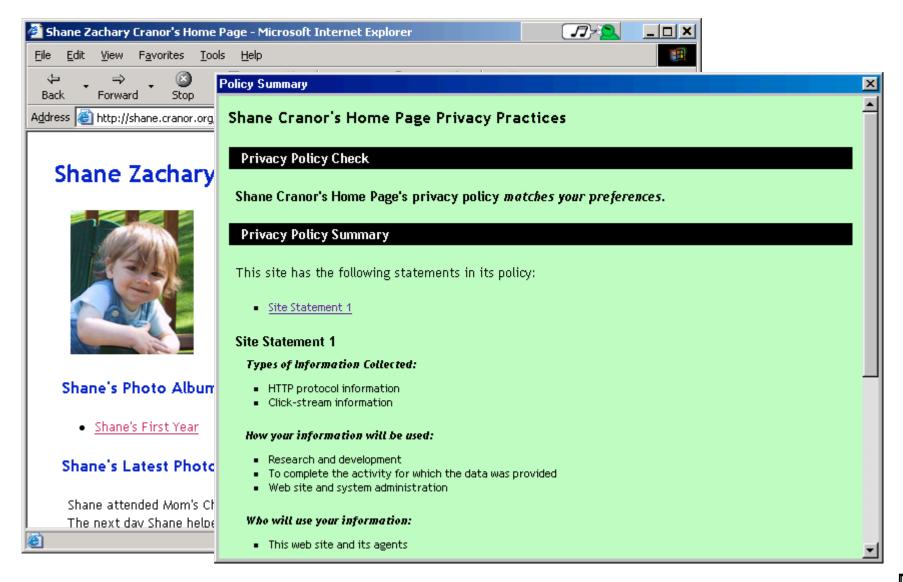
- Clicking on bird icon gives more information
- Current version is information only no cookie blocking



Chirping bird is privacy indicator

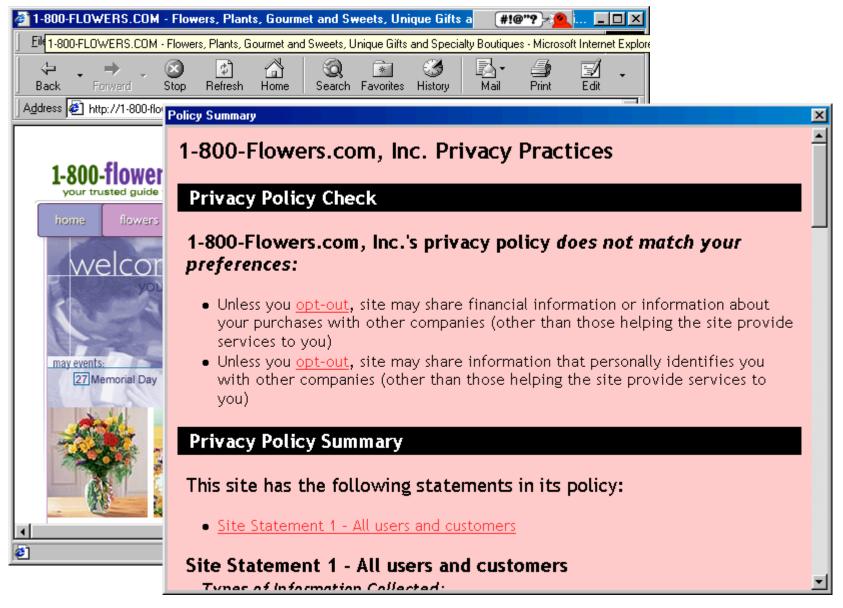


Click on the bird for more info

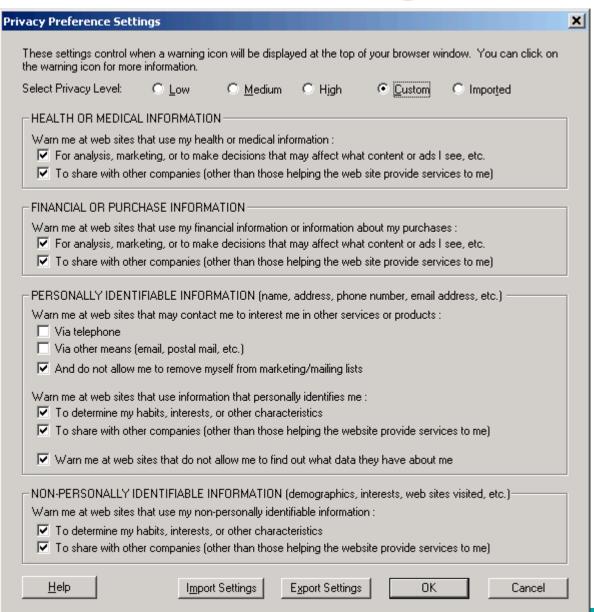




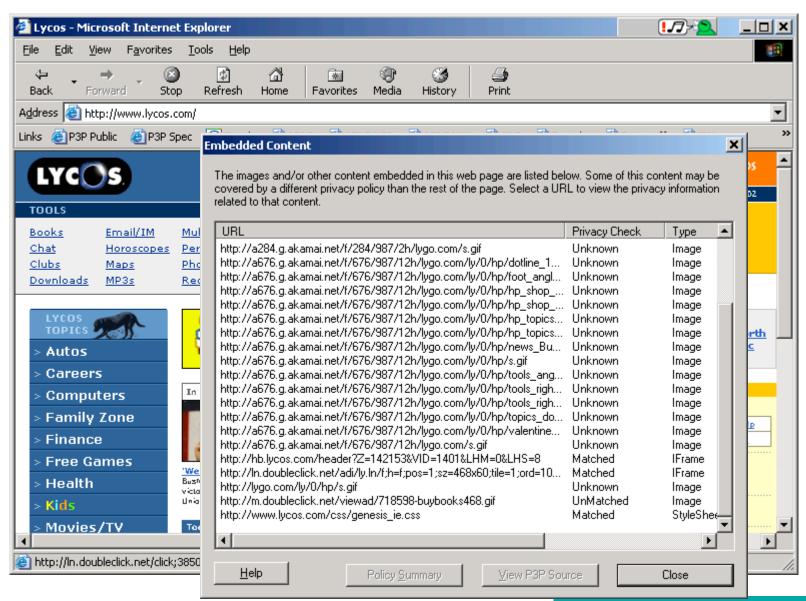
Privacy policy summary - mismatch



Users select warning conditions



Bird checks policies for embedded content



Questions?

How to make your Web site's privacy policy machine readable

Lorrie Faith Cranor

P3P Specification Working Group Chair Carnegie Mellon University

January 2004

http://lorrie.cranor.org/

P3P deployment overview

- 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
- Test your site to make sure it is properly P3P enabled

One policy or many?

- P3P allows policies to be specified for individual URLs or cookies
- One policy for entire web site (all URLs and cookies) is easiest to manage
- Multiple policies can allow more specific declarations about particular parts of the site
- Multiple policies may be needed if different parts of the site have different owners or responsible parties (for example, different congressional offices)

Policy reference files (PRF)

- Allows web sites to indicate which policy applies to each resource (URL or cookie)
 - ★Every resource (HTML page, image, sound, form action URL, etc.) can have its own policy
- User agents can cache PRFs (as long as permitted by EXPIRY) so they don't have to fetch a new PRF every time a user clicks

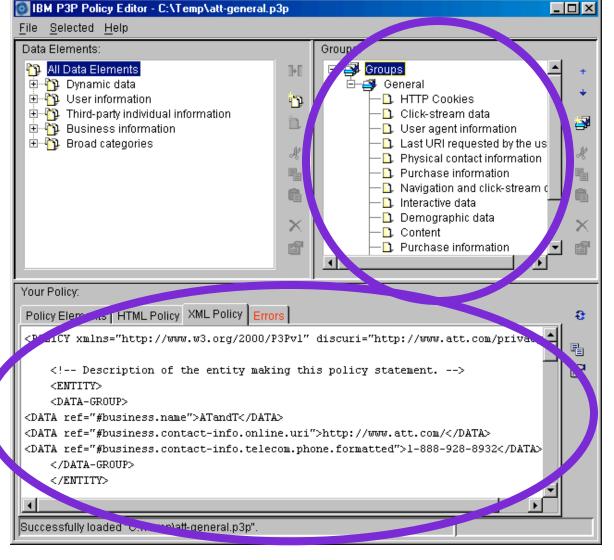
Third-party content

- Third-party content should be P3Penabled by the third-party
- If third-party content sets cookies, IE6 will block them by default unless they have P3P compact policy
- Your first-party cookies may become third-party cookies if your site is framed by another site, a page is sent via email, etc.

Generating a P3P policy

- Edit by hand
 - ★Cut and paste from an example
- Use a P3P policy generator
 - ★ Recommended: IBM P3P policy editor http://www.alphaworks.ibm.com/tech/p3peditor
- Generate compact policy and policy reference file the same way (by hand or with policy editor)
- Get a book
 - ★ Web Privacy with P3P by Lorrie Faith Cranor http://p3pbook.com/

IBM P3P Policy Editor



Sites can
list the types
of data they
collect

And view the corresponding P3P policy

Locating the policy reference file

- Place policy reference file in "well known location" /w3c/p3p.xml
 - ★ Most sites will do this
- Use special P3P HTTP header
 - ★ Recommended only for sites with unusual circumstances, such as those with many P3P policies
- Embed link tags in HTML files
 - ★ Recommended only for sites that exist as a directory on somebody else's server (for example, some congressional representative websites http://www.house.gov/NAME/)

Server configuration

- Only needed for compact policies and/or sites that use P3P HTTP header
- Need to configure server to insert extra headers
- Procedure depends on server see P3P Deployment Guide appendix

http://www.w3.org/TR/p3pdeployment or Appendix B of Web Privacy with P3P

Policy updates

- Changing your P3P policy is difficult, but possible
- New policy applies only to new data (old policy applies to old data unless you have informed consent to apply new policy)
- Technically you can indicate exact moment when old policy will cease to apply and new policy will apply
- But, generally it's easiest to have a policy phase-in period where your practices are consistent with both policies

P3P/XML encoding

```
P3P version
      <POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1"> <
       <POLICY discuri="http://p3pbook.com/privacy.html" <
                                                                     Location of
               name="policy">
                                          P3P policy name
         <ENTITY>
                                                                      human-readable
         <DATA-GROUP>
                                                                     privacy policy
           <DATA
Site's
             ref="#business.contact-info.online.email">privacy@p3pbook.com
           </DATA>
name
           <DATA
and
             ref="#business.contact-info.online.uri">http://p3pbook.com/
contact
           </DATA>
info
           <DATA ref="#business.name">Web Privacy With P3P</DATA>
         </DATA-GROUP>
                                             Access disclosure
         </ENTITY>
                                                                  Human-readable
         <ACCESS><nonident/></ACCESS>
                                                                  explanation
         <STATEMENT>
           <CONSEQUENCE>We keep standard web server logs.
           <PURPOSE><admin/><current/><develop/></PURPOSE>
                                                                   — How data may
  Statement
           <RECIPIENT><ours/></RECIPIENT>
                                                                      be used
           <RETENTION><indefinitely/></RETENTION>
           <DATA-GROUP>
                                                               Data recipients
              <DATA ref="#dynamic.clickstream"/>
              <DATA ref="#dynamic.http"/>
                                                              Data retention policy
           </DATA-GROUP>
         </statement>
                                                        Types of data collected
       </POLICY>
       </POLICIES>
```

XML syntax basics

```
Element opening tag
                                 Attribute
          <BIG-ELEMENT>
doesn't contain --> <element name="value" />
other elements </BIG-ELEMENT>
                                           Comment
(ending slash)
Element
             -- This is a comment
closing tag
(beginning
slash)
          <ELEMENT>Sometimes data goes
          between opening and closing
Element
          tags</ELEMENT>
that contains
character
data
```

Assertions in a P3P policy

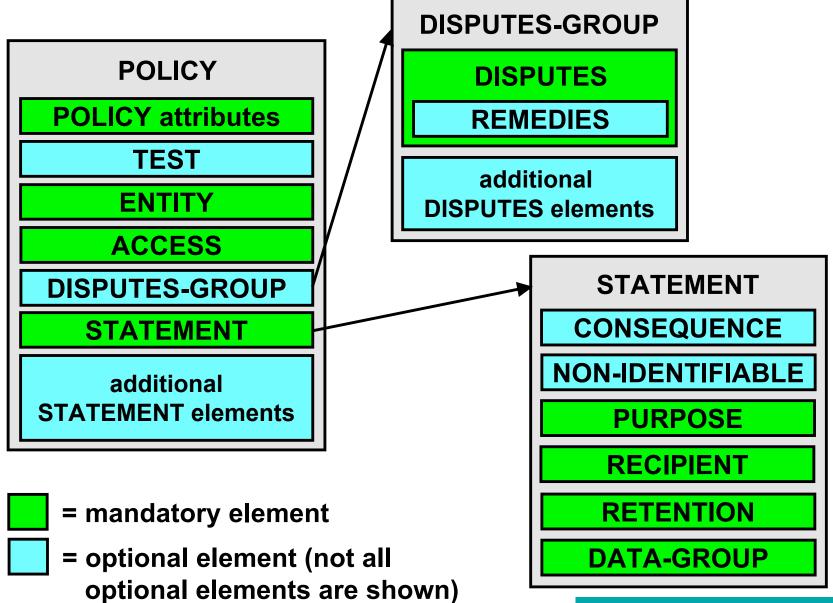
General assertions

- Location of human-readable policies and opt-out mechanisms discuri, opturi attributes of <POLICY>
- ★ Indication that policy is for testing only <TEST> (optional)
- ★ Web site contact information <ENTITY>
- ★ Access information <ACCESS>
- ★ Information about dispute resolution <DISPUTES> (optional)

Data-Specific Assertions

- **★** Consequence of providing data **<CONSEQUENCE>** (optional)
- Indication that no identifiable data is collected -<NON-IDENTIFIABLE> (optional)
- ★ How data will be used <PURPOSE>
- ★ With whom data may be shared <RECIPIENT>
- ★ Data retention policy <RETENTION>
- ★ What kind of data is collected <DATA>

Structure of a P3P policy



The ACCESS Element

Indicates the ability of individuals to access their data

```
*<nonident/>
*<all/>
*<contact-and-other/>
*<ident-contact/>
*<other-ident/>
*<none/>
```

Example

```
<ACCESS><nonident/></ACCESS>
```

The DISPUTES Element

Describes a dispute resolution procedure

may be followed for disputes about a service's privacy practices

Part of a

<DISPUTES-GROUP>

* allows multiple dispute resolution procedures to be listed

Attributes:

- ★ resolution-type
 - customer service
 - independent organization
 - court
 - applicable law
- * service
- * short-description
 (optional)
- ★ Verification (optional)

Sub-Elements

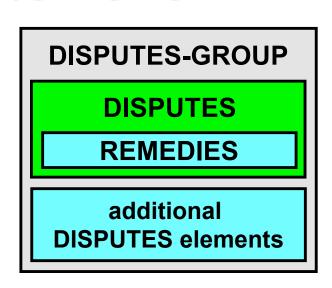
- **★ <IMAGE>** (optional)
- *** <LONG-DESCRIPTION>**(optional)
- **★ <REMEDIES>** (optional)

The REMEDIES element

- Sub element of DISPUTES element
- Specifies possible remedies in case a policy breach occurs

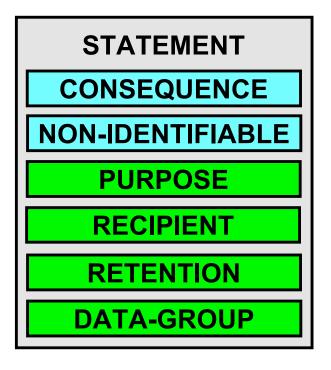
```
★ <correct/>, <money/>, <law/>
```

■ Example of DISPUTES and REMEDIES



The STATEMENT element

- Data practices applied to data elements
 - ★ mostly serves as a grouping mechanism
- Contains the following sub-elements
 - ★ <CONSEQUENCE>
 (optional)
 - ★ <NON-IDENTIFIABLE> (optional)
 - ★ <PURPOSE>
 - ★ <RECIPIENT>
 - ★ <RETENTION>
 - ★ <DATA-GROUP>



The CONSEQUENCE element

- Consequences that can be shown to a human user to explain why the suggested practice may be valuable in a particular instance, even if the user would not normally allow the practice
- Example

<CONSEQUENCE>We offer a 10% discount
to all individuals who join our Cool
Deals Club and allow us to send them
information about cool deals that
they might be interested
in.</CONSEQUENCE>

The NON-IDENTIFIABLE element

- Can optionally be used to declare that no data or no identifiable data is collected
 - ★ non-identifiable: there is no reasonable way to attach collected data to identity of a natural person, even with assistance from a third-party
 - ★ Stronger requirements than *non-identified*
- Must have a human readable explanation how this is done at the discuri
- Other STATEMENT elements are optional when NON-IDENTIFIABLE is present
 - <NON-IDENTIFIABLE/>

The PURPOSE element

Purposes of data collection, or uses of data

```
* <current/>
* <admin/>
* <develop/>
* <tailoring/>
* <pseudo-analysis/>
* <pseudo-decision/>
* <individual-analysis/>
* <individual-decision/>
* <individual-decision/>
* <contact/>
* <historical/>
* <telemarketing/>
* <other-purpose/>
```

Optional attribute:

```
★ required
```

- always (default)
- opt-in
- opt-out

Example

Only use as last resort

Customization purposes

Purpose	Does this involve creating a profile of the user?	How is the user identified?	Does this result in a decision that directly affects the user?
Research and development	No	user is not identified	No
One-time tailoring	No	user may not be identified at all, or may be identified with a pseudonym or with personally-identifiable information	Yes
Pseudonymous analysis	Yes	pseudonym	No
Pseudonymous decision	Yes	pseudonym	Yes
Individual analysis	Yes	personally- identifiable information	No
Individual decision	Yes	personally- identifiable information	Yes

The RECIPIENT element

Recipients of the collected data

Optional attribute

- * required
 - always (default)
 - opt-in
 - opt-out

Optional sub-element

```
★ <recipient-
description>
```

Example

```
<RECIPIENT>
  <ours/>
  <same required=
   "opt-out"/>
   <delivery>
    <recipient-description>
     FedEx
    </recipient-description>
     </delivery>
  </RECIPIENT>
```

The RETENTION element

Indicates the kind or retention policy that applies to the referenced data

Example

<RETENTION><indefinitely/></RETENTION>

The DATA element

- Describes the data to be transferred or inferred
- Contained in a DATA-GROUP
- Attributes:

```
* ref

* optional (optional, default is no, not optional=required)
```

- Sub-Elements:
 - ★ <CATEGORIES>
- Example

The CATEGORIES element

Provides hints to user agents as to the intended uses of the data

- ★ Physical contact information
- ★ Online contact information
- ★ Unique identifiers
- ★ Purchase information
- ★ Financial information
- ★ Computer information
- Navigation and clickstream data
- ★ Interactive data

- Demographic and socioeconomic data
- * Content
- State management mechanisms
- ★ Political information
- ★ Health information
- ★ Preference data
- Government-issued identifiers
- ★ other

Base Data Schema

- User data user
 - name, bdate, cert, gender, employer,
 department, jobtitle, home-info,
 business-info
- Third party data thirdparty
 - ★ Same as user
- Business data business
 - *name, department, cert, contact-info
- Dynamically generated Dynamic
 - *clickstream, http, clientevents,
 cookies, miscdata, searchtext,
 interactionrecord

dynamic.miscdata

- Used to represent data described only by category (without any other specific data element name)
- Must list applicable categories
- Example

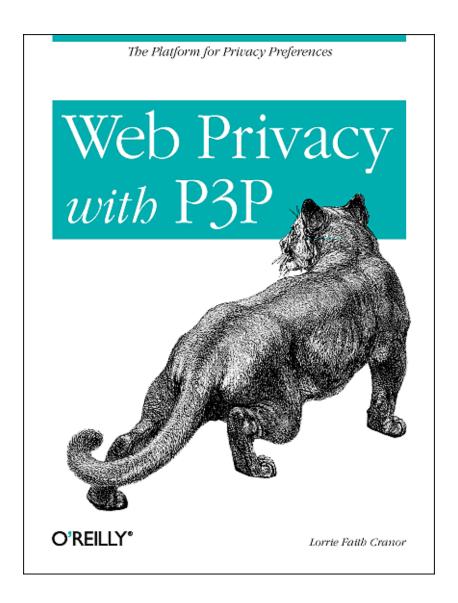
Resources

For further information on P3P see:

http://www.w3.org/P3P/

http://p3ptoolbox.org/

http://p3pbook.com/



Validating your site and looking for common errors

Lorrie Faith Cranor

P3P Specification Working Group Chair Carnegie Mellon University

January 2004

http://lorrie.cranor.org/

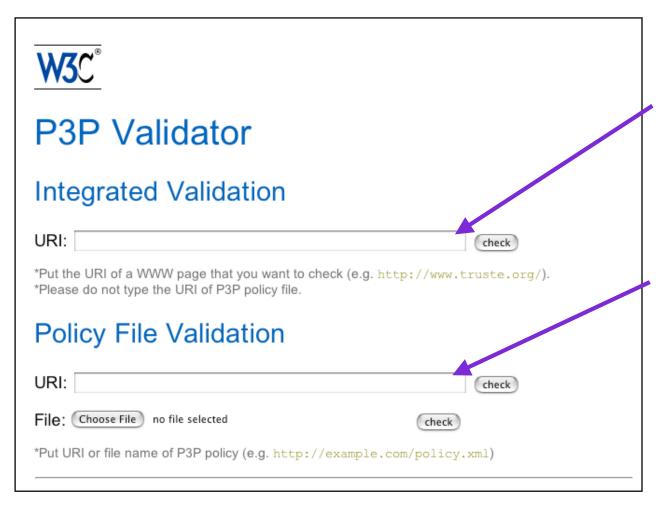
Don't forget to test!

Make sure you use the P3P validator to check for syntax errors and make sure files are in the right place

http://www.w3.org/P3P/validator/

- ★But validator can't tell whether your policy is accurate
- Use P3P user agents to view your policy and read their policy summaries carefully
- Test multiple pages on your site

P3P Validator



Use to validate fully P3P-enabled site

Use to validate
P3P policy before
posting to web site

Common errors to avoid

- Failing to mention web logs in policy
- Failing to mention data linked to cookies
- Unnecessarily disclosing contact purpose
- Failing to give opturi when opt-in or opt-out are used
- Hand coding a PRF and using / or \ instead of /* to apply policy to entire site
- Omitting policy name or putting a space in the policy name
- Omitting COOKIE-INCLUDE in PRF of site with cookies
- Putting P3P files on password protected sections of site
- Failing to P3P-enable secure server
- Failing to test