

Online Privacy

Dave Raggett, W3C and UWE

Reading, 27 July 2011

What is Privacy?

- We can all recognize effects of its absence
 - Injury or harm
 - Due to prejudice or malicious intentions
 - Financial harm, Physical harm, inability to work or travel
 - Loss of face
 - Parents, friends or colleagues learning something that embarrasses you
 - Loss of control
 - Inability to keep things private
 - Profiling by Search engines and advertisers

Evolution of Privacy on the Web

- HTTP logs
 - IP address, time, URL
- Cookies
 - Originally proposed to reduce burden on server
 - Now have many uses, e.g.
 - Session management
 - User preferences
 - Saving where you left off viewing a video (app state)
 - Setting a unique id for a visitor over repeat visits
 - Recording profile data for targeted advertising

And worse . . .

- Proliferation of means to store data in browser
 - HTML5 local storage, flash
- Fingerprinting
 - “80% of browsers have unique fingerprint”, *Electronic Frontier Foundation, 2010*
 - Try it yourself at <http://panopticklick.eff.org/>
 - Based upon installed fonts and plugins, etc.
- History stealing
 - Via CSS and JavaScript
- Web Bugs
 - Unique tracking ids {hidden images, scripts, iframes, etc.}

Third Parties

- Websites commonly make use of third parties
 - Content distribution
 - e.g. akamai
 - Understanding website traffic
 - e.g. Google analytics and Quantcast
 - Who are your visitors, where do they come from, and what pages do they like?
 - Advertising
 - e.g. DoubleClick (Google) and RightMedia (Yahoo!)

Advertising

- Enabling websites to offer free services
- Initial focus on context-based advertising
 - Ad networks matching ads to web page content
- Now based on profiling users through detailed tracking right across the web
 - Via 3rd party cookies on many websites
- **YOU** are the product websites sell to profilers!

Facebook *like* button



- Added by website developers to allow facebook users to recommend a link to their friends
- Implemented as iframe or script element
- Enables facebook to track its users across all sites with the *Like* button
- You don't even need to click the button to be tracked!
 - Browser sends identification cookie to facebook when loading the script or iframe
 - Facebook tells you whether your friends like this page or invites you to be the first of your friends
 - 1,602 people like this page, be the first of your friends.

Surveillance on the cheap

- “Governments have changed to using data brokers for much of their surveillance, and buying profile data from advertisers” *
- Chris Hoofnagle, Berkeley Center for Law & Technology

Privacy doesn't matter

- “nobody cares about online privacy because they're worried about terrorism and the economy” *
- Russell Glass, Bizo
 - “Bizo is how marketers reach and engage business professionals, wherever they travel online. Bizo's unique ability to precisely target more than 80% of the US business population gives marketers cost-effective access and insight into business professionals – the most valuable online audience segment.”

Privacy doesn't matter

What do you think?

What do your customers think?

Ethnographic Perspective

- Privacy is not dead, but it is deeply misunderstood
 - Danah Boyd, Microsoft researcher
 - Privacy, publicity and visibility
 - “Public by default, private through effort”
 - <http://www.danah.org/papers/talks/2010/TechFest2010.html>
 - Living life in public: why american teens choose publicity over privacy
 - “Using in-jokes and encoding information to limit visibility”
 - <http://www.danah.org/papers/talks/2010/AOIR2010.html>

Emergence of browser extensions for privacy

- Firefox addons relating to privacy that help with blocking ads and inhibiting cookies
 - Adblock Plus
 - BetterPrivacy
 - NoScript
 - Silent Block
 - Privacy Dashboard

Privacy Dashboard

- Developed by EU PrimeLife project
 - <http://primelife.eu/>
 - <http://code.w3.org/privacy-dashboard/>
- See how websites are tracking you, and set per site preferences, e.g. to block 3rd party content or cookies
- Share your findings with others
- Privacy Dashbot
 - Automated survey of top 1000 websites

Information on current site

Privacy Dashboard

Data Track Location **Current Website** Share Findings About

Information about the current website

Review and adjust privacy options for the current website, including, cookies, P3P and more. You can find out more about this website using the queries on the 'Data Track' tab above.

www.lovefilm.com

<p>This website has:</p> <ul style="list-style-type: none">• 14 session cookies• 6 lasting cookies• a flash cookie• 7 internal third party sites• 11 external third party sites• an external third party session cookie• 24 external third party lasting cookies	<p>Your preferences for this website:</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> Never block content from this site<input checked="" type="checkbox"/> Block external 3rd parties<input checked="" type="checkbox"/> Block external 3rd party cookies<input checked="" type="checkbox"/> Block all lasting cookies<input checked="" type="checkbox"/> Clear flash cookies<input checked="" type="checkbox"/> Disable web page scripting<input checked="" type="checkbox"/> Disable access to your geolocation<input checked="" type="checkbox"/> Disable HTML5 pings<input checked="" type="checkbox"/> Don't send HTTP referrer header<input checked="" type="checkbox"/> Disable web page access to DOM storage <p><input type="button" value="Simple View"/></p> <p><input checked="" type="checkbox"/> Use these by default for all websites</p>
--	--

You can use the following buttons to check the current website in various ways

Query data on each site

Privacy Dashboard

Data Track | Location | Current Website | Share Findings | About

What are websites collecting and how?

See what data is being collected by the websites you visit, what cookies they set, and more.

Query results appear below -- click on

...	se...	se...
*.ki...	/	user_...	false	false
*.ki...	/	_ut...	false	false
*.ki...	/	_utm...	true	false
*.ki...	/	_utma	false	false
*.ki...	/	last_...	false	false
*.ki...	/	PHPS...	true	false
*.ki...	/	_utmz	false	false

1. Select query:

- What http cookies are used by a given website?
- Which websites use long lasting cookies?
- Which websites use session cookies?
- Which websites use flash cookies?
- Which websites use DOM storage?
- Which websites use invisible images?
- Which websites use HTML5 ping attributes?
- Which websites provide P3P privacy policies?
- Which websites are 3rd parties?
- Which websites use a given 3rd party?
- Which internal 3rd parties are used by a given website?
- Which external 3rd parties are used by a given website?
- What http cookies are used by a given website?
- Which websites have permission to access my location?
- What data has been sent to a given website?
- Which websites a given datum value has been sent to?
- Which websites a given datum name has been sent to?
- Which datum names are used for a given value?

Query data on each site

The screenshot shows a web application window titled "Privacy Dashboard". It has a navigation bar with tabs: "Data Track", "Location", "Current Website", "Share Findings", and "About".

What are websites collecting and how?
See what data is being collected by the websites you visit, what cookies they set, and more.

1. Select query: Which external 3rd parties are used by a given website?

2. Domain name or URL: www.kinopoisk.ru

3. Execute query

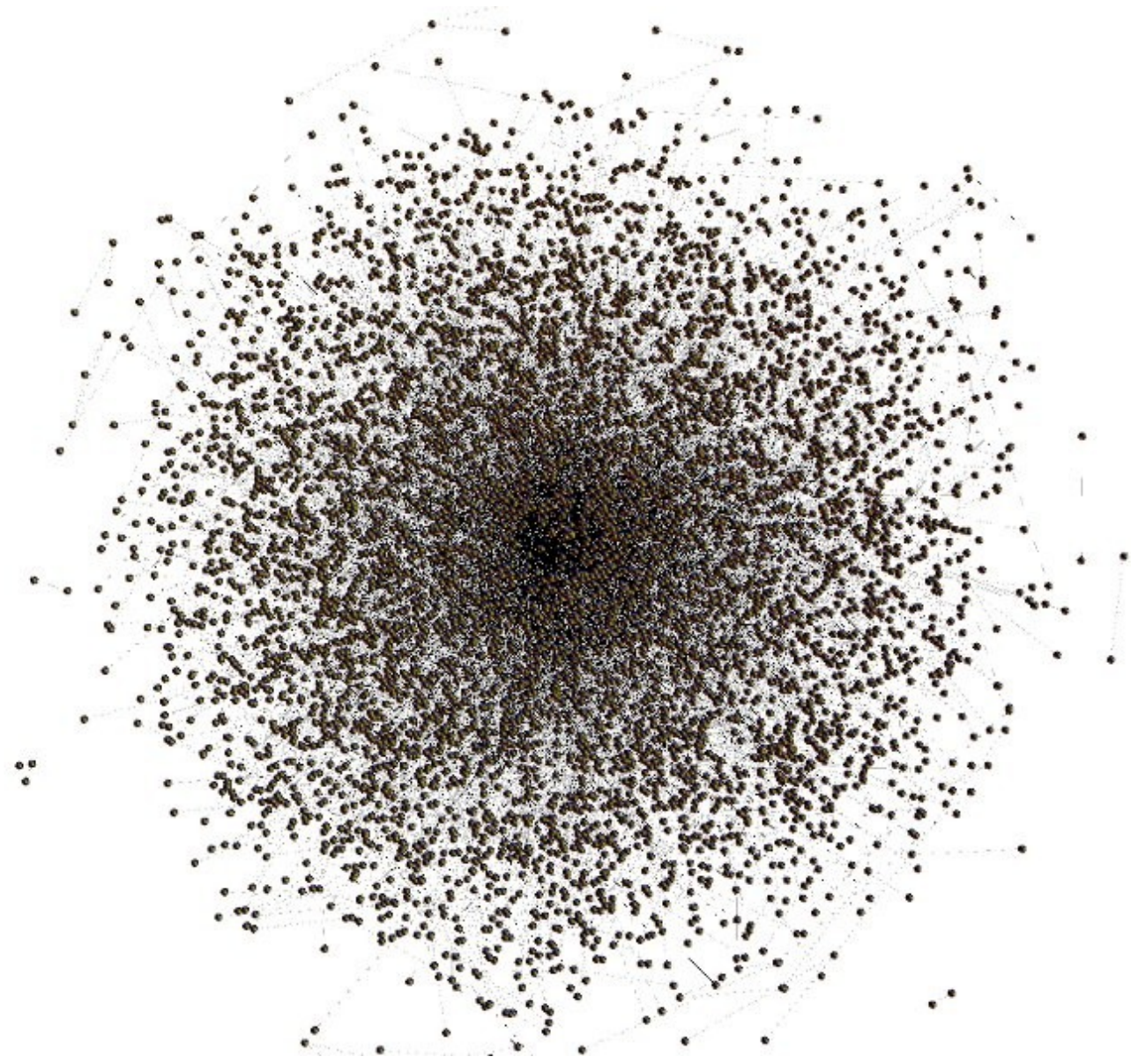
Query results appear below – click on values to set the query parameter above

third_party
ad.adriver.ru
api.conduit.com
apps.conduit.com
counter.rambler.ru
counter.yadro.ru
kinopoiskgaua.hit.gemius.pl
marriottinternational.122.2o7.net
ping.chartbeat.net
static.chartbeat.com
top.list.ru
top3.mail.ru
usage.apps.conduit-services.com
www.google-analytics.com
www.google.com
www.tns-counter.ru
yandex.st

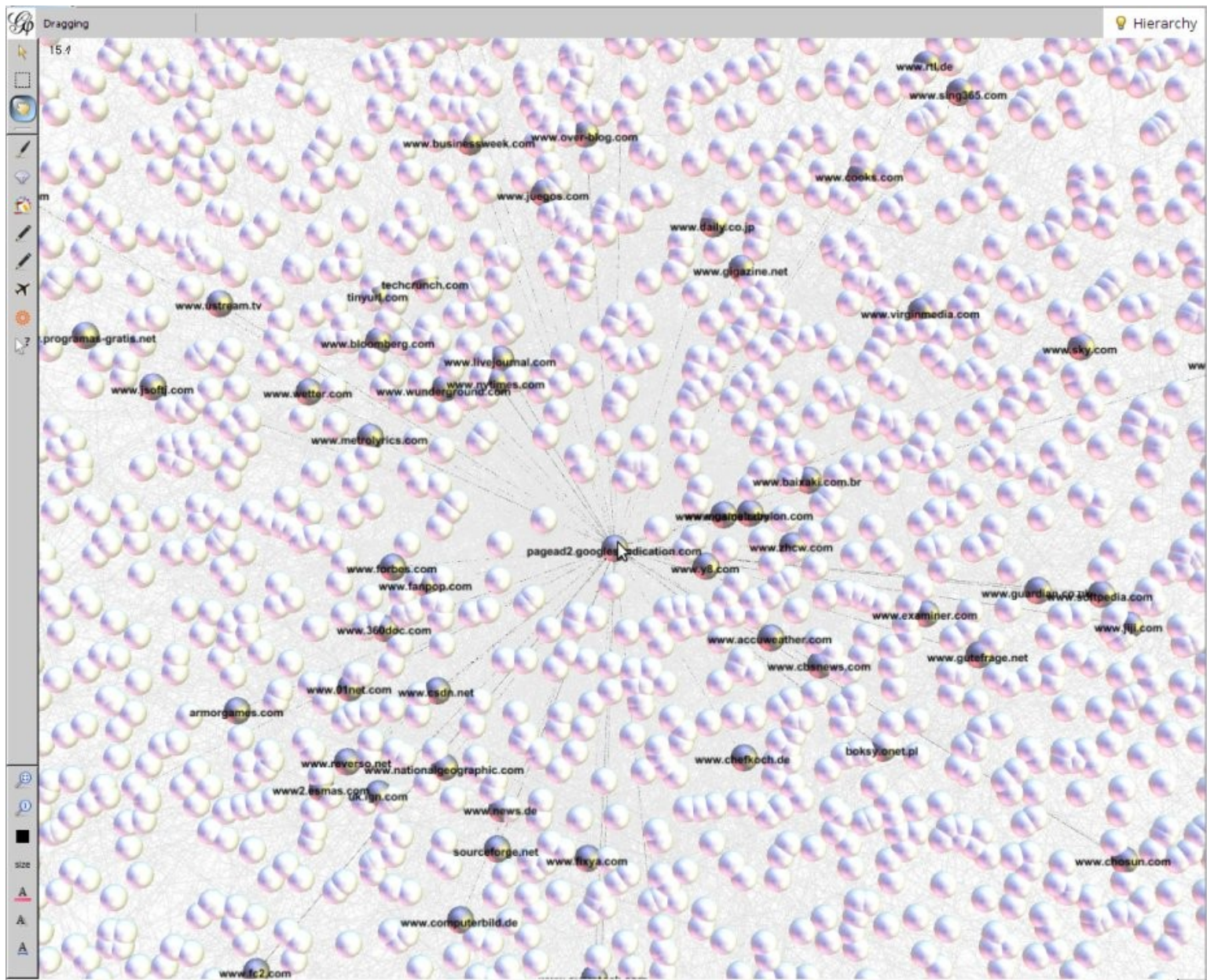
Privacy Dashbot

- Adaptation of Dashboard to scan top 1000 sites
 - Based upon list provided by Google
- Determine associated 3rd party sites and form into clusters
- Rank hosts by the number of hosts citing them as direct 3rd parties

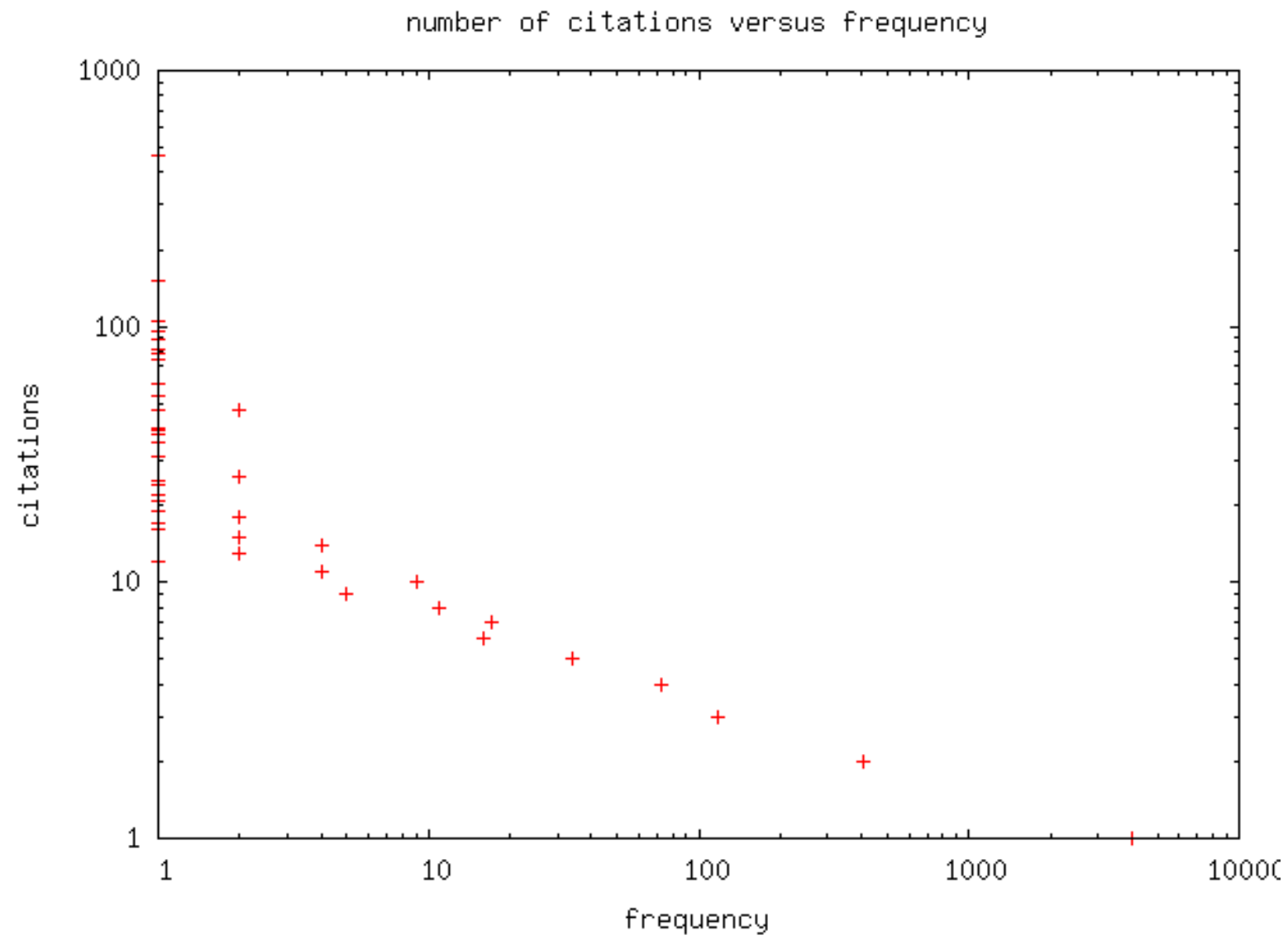
The Web as a galaxy of sites



The Web as a galaxy of sites



Frequency of citations counts as 3rd parties



Is self regulation enough?

- Network Advertising Initiative opt out cookies
 - NAI is a cooperative of online marketing and analytics companies
 - Visit NAI opt-out page to set cookies as signal to 3rd parties to avoid targeted ads
 - You are still tracked and the data may be sold on
 - Switch browsers or devices and start all over again!
- Need for better regulation and better tools

Do Not Track (DNT)

- HTTP header or DOM property
 - Set by browser to signal to 3rd parties to avoid tracking the user
 - Persistent user preference setting in browser
 - Relies on websites honouring user's request
 - What does “do not track” mean exactly?
 - Vested interests seeking to define it to their advantage
 - Supported by US FTC
- Not set by default – i.e. users have to opt out of behavioural tracking

Tracking Protection Lists

- Microsoft feature for IE9 and later
- Lists of rules for which 3rd party sites should be blocked or enabled
- Users decide which lists to apply
- Lists supplied by Internet privacy organizations
- Not enabled by default – i.e. users have to opt out of behavioural tracking
- Relationship to P3P (machine readable privacy policies)

European Perspective

- Neelie Kroes, *European Commission VP for Digital Agenda*
 - *Principles for privacy in the digital age*
 - **Transparency, fairness and user control**
 - <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/461>
 - **E-Privacy directive 2002/58 on Privacy and Electronic Communications**
 - Only UK, Estonia, Finland, Sweden and the Netherlands have implemented the directive in their domestic laws as of June 2011

Transparency, fairness & user control

- Users must be given clear and comprehensive information on the purposes and retention policy
- This must be fair – sites must not present a one sided agreement to the detriment of the user
- The user must be given a means to opt in or out of behavioural tracking
- The EU legislator's desire to leave the implementation of the directive open to future technological innovations
 - No guidance on what constitutes an opt-out
 - “Self-regulatory efforts are clearly part of the equation on the implementation side. But we need to go further and look beyond cookies and specific sectors. DNT can help us do this.”, Neelie Kroes

E-Privacy Directive

- Security obligations
 - Duty to inform subscribers of risks from viruses or malware
 - Prohibition on listening, tapping, storage or other kinds of interception or surveillance of communication and “related traffic”, unless the users have given their consent
 - some exceptions under Article 15(1)
- Data retention
 - Obligation to erase or anonymize traffic data
 - Right to non-itemised billing

E-Privacy Directive

- Spam
 - Prohibition of use of email and SMS for marketing except with prior agreement of the recipient
- Cookies
 - May only be set if the user is provided with clear and comprehensive information about the purpose, and the user is offered a means to opt-out
 - Does this apply to other means of implementing behavioural tracking or is it limited to cookies?
 - Is a browser based cookie blocker sufficient?

Privacy Policies

- Many websites have a privacy policy
- It is often hard to find, and hard for ordinary people to understand
- Cookies have lots of useful purposes
- Policy should state what broad classes of information are being collected, for what purposes, who it may be shared with and for how long it will be retained
- “Operational purposes” is no longer acceptable!

Fairness

- If the user has opted out of behavioural tracking the user shouldn't be penalized in an unfair manner
- If a site is ad-supported, is it reasonable to ask users who opt out of effectively targeted ads to pay in some manner?
- Can we look forward to online equivalent of the nectar card and other loyalty schemes that act across a group of non-competing businesses?

Privacy friendly strong authentication

- Increasing use of email addresses as user ids
 - Facilitates linkability of personal information across sites
- Sites need strong assurances as to attributes of identified party, but not a globally linkable id
- User defined personae can help
 - Browser based account management
- New cryptographic techniques for proving user has trusted credential with given attributes BUT without disclosing user's identity
 - Age, address, nationality, membership of named group, etc.
- We need better credentials with robust processes

W3C's Role

- Initial work on P3P
 - Platform for Privacy Preferences
- Involvement in EU PrimeLife project
- Many workshops related to privacy
- Policy Languages Interest Group
- Privacy by design for Web APIs, e.g. geolocation
- New Privacy Interest Group
- New WG's expected on DNT and Protection Lists
- Implementation work on new technologies
 - Funded through W3C involvement in EU webinos project

Platform for Privacy Preferences

- P3P 1.0 Recommendation issued April 2001
- Machine readable privacy policies
- But initial specification too flexible!
 - Makes it hard to generate report of mismatch between user preferences and site policy
 - Complicates UI for setting user preferences
- Compact policies as a (bad) compromise
 - Implemented in Internet Explorer and other browsers
 - 3rd party cookies blocked if there is no (compact) P3P policy
 - Only covers cookies

Fresh take on P3P

- Developed as part of EU PrimeLife project
- Improves on compact policies to cover a much bigger subset of P3P
 - Expressed in JSON
- Chosen to make it easy to create UI for user preferences and to generate reports of mismatch with site policy
- Includes link to full human readable policy
- For details see <http://www.w3.org/2011/D1.2.3/>

Machine generated view of P3P privacy policy

PrimeLife Policy Preferences Editor

Generic Privacy Policy for ACME widgets online inc.

We may collect the following types of information about you:

- Information about the computer you are using, such as its hardware, software, or Internet address
- Which pages you visited on this web site and how long you stayed at each page
- Activities you engaged in at this web site, such as your searches and transactions

The ways your information may be used:

- To provide the service you requested
- To perform web site and system administration
- To customize the site for your current visit only
- To do research and analysis that uses information about you

With whom we may share your information:

- Companies that help us fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose
- Delivery companies that help us fulfill your requests and who may also use your information in other ways
- Companies that have privacy policies similar to ours

How long we may keep your information:

- Our full privacy policy explains how long we keep your information


Further information

- [Link to full privacy policy not provided](#)

[Return to view of preferences warning](#)

Machine generated view of P3P privacy policy

PrimeLife Policy Preferences Editor

 **Generic Privacy Preferences Warning**

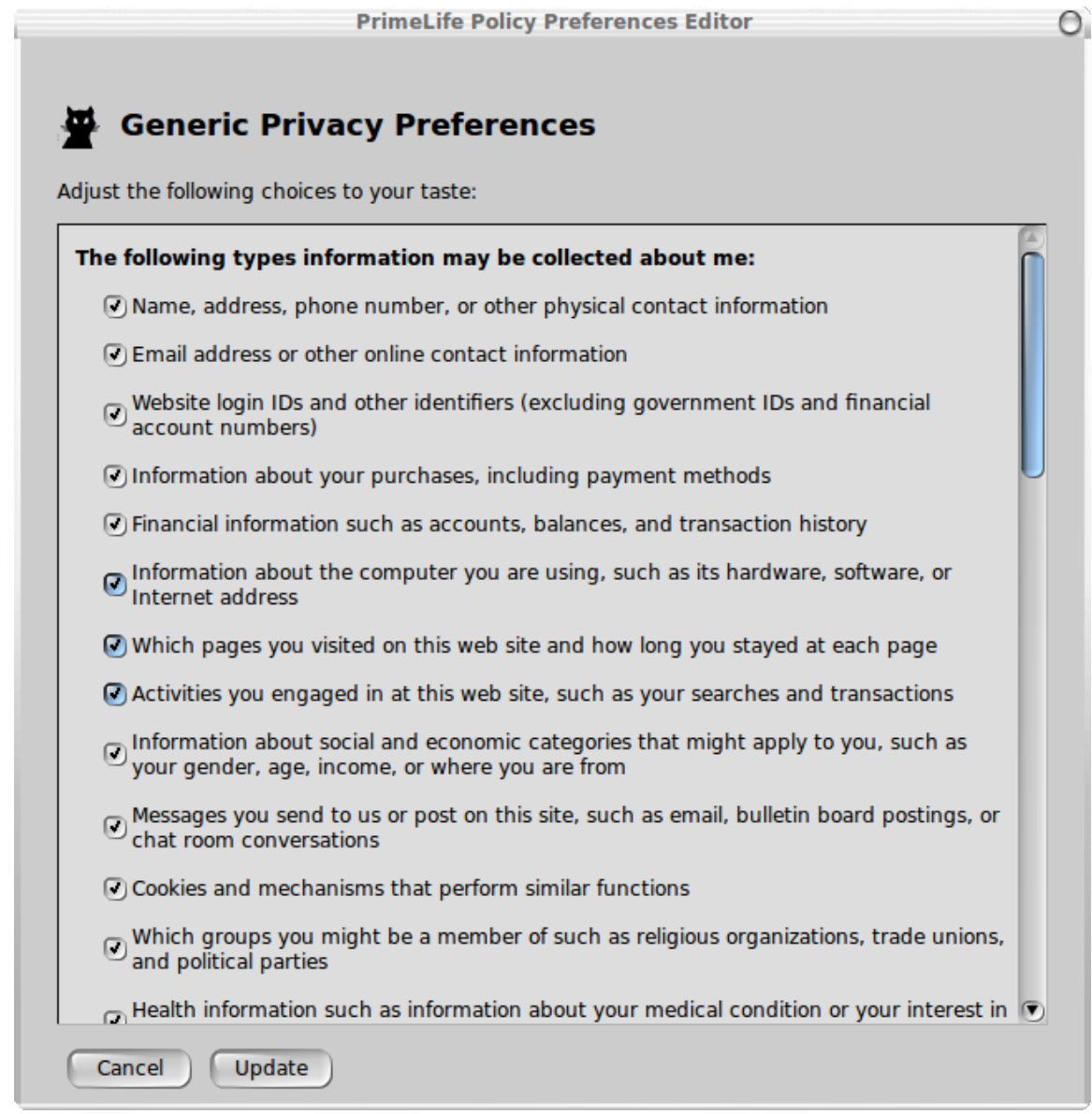
Each time your browser sends a request to a website, some information is disclosed in the HTTP headers. This includes your browser's external IP address, information about your browser and operating system, and your language preferences. The IP address may provide information about your location and your identity.

This website's generic policy conflicts with your preferences in the following ways:

1. The website says it would like to use your information
 - *To do research and analysis that uses information about you*
2. The website says it would like to share your information with
 - *Companies that have privacy policies similar to ours*
3. The website says it would like to keep your information for a longer period
 - *Our full privacy policy explains how long we keep your information*
4. The website doesn't provide a link to its full privacy policy

Please select between the following actions:

UI for user preferences



Dealing with the click-thru effect

- Tendency for most users to click through security and privacy dialogues
 - Seen as an annoying interruption
 - Most users aren't in good position to make informed decision
- Solution is to delegate decision to a trusted third party
- White lists for sites which are known to be trustworthy – no warning UI presented
- Black lists for sites which are deemed bad
- Otherwise show user the security/privacy dialogue

Demo of Anonymous Credentials

- Student union issues credentials to new students
- Only current students can access the student union's social website
 - No university staff, no would be employers, no ex students
- Website exposes machine readable policy covering authentication and privacy
- User authenticates to browser, then browser to website
- Website has strong proof that user is current student, but doesn't learn which student!
- Zero knowledge proof with IBM's idemix library
 - No need to contact credential issuer to create proof

Common-Room.com - a hypothetical social website for students

A private meeting place for students and off limits to the College staff!

Come on in: **Connect!**

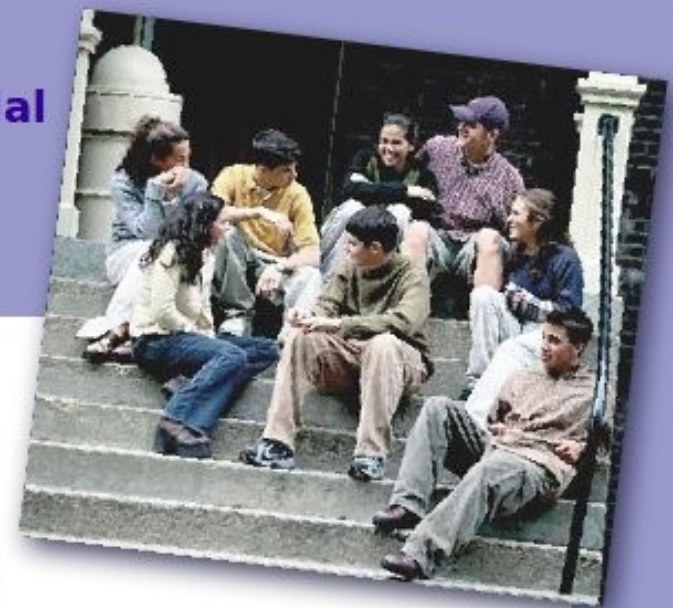
This site allows you to talk about anything you feel like, and to upload photos and videos. At any time of the day or night, you can see who is online and chat to them in real time in anonymity. You can rant or rave about what's happening, and leave notes for others to comment on.

To use this site, you must have a credential attesting that you are a current undergraduate student at Redbrick College. You can get one at the Student Union offices at 47a University Avenue. Don't forget to bring along some visual id for us to check against our records. We will give you a welcome pack with the details of how to get started.

For the technically minded, pressing the connect button invokes a browser extension to search for the electronic version of your student id card and then creates an anonymous proof that you are a current student. Your identity is **not** revealed to us, although, you may find it convenient to give yourself a nick name for conversing with fellow students. After that it is really up to you how much you want to share with other students. Just remember that this site is off limits to College staff, employers and even past students!

Students are issued with a USB stick with their credential as a current student. This is accessed by the browser extension to generate a proof for the website, without disclosing the student's id.

Students have to install the security extension from the memory stick before they can use the web site. Their credential remains on the memory stick enabling them to switch computers with safety. A PIN has to be entered to unlock the memory stick as a precaution against it being mislaid or falling into the hands of others.



Authenticating the User

Common-Room.com - a hypothetical social website for students

A page of limits to the
Co

Student Id Card

Please enter your 4 digit pin and press Enter to prove to localhost that you possess this credential, note that the credential itself is not passed to the website

Co

This is something you feel like. At any time of the day or night, you can see who is online and chat to them in real time in anonymity. You can rant or rave about what's happening, and leave notes for others to comment on.

Welcome Back to Common-Room.com - a hypothetical social website for students

A private meeting place for students and off limits to the College staff!

Your nick name:

If you like, you can use a nickname to identify yourself to other students

❁ What's On

See's what's on and what people are saying about it.

❁ Live Chat

Chat with others online in meeting places you create.

❁ Scrap Book

Upload your photos and videos, and write notes for others to respond to



This page is locked - you can only see this if you are the owner of a valid student credential.

Questions?

Dave Raggett <dsr@w3.org>