ADVANCED INTERNET POLICY REPORT

ONLINE ADVERTISING:
Confronting the Challenges

MAY 2009

BY

Policy Engagement Network

The Information Systems and Innovation Group

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

PEN paper 2

About this report and the research methodology

This report is a culmination of two years of research on advanced internet policy issues. It was conducted through the Policy Engagement Network within the Information Systems and Innovation Group at the London School of Economics and Political Science.

This series of *Advanced Internet Policy Reports* presents our research in a package that is geared towards policy-makers, regulators, industry groups, and the interested general public. They are research reports that bring together the state of the art of the practices across industry to identify the 'dynamics' of the field. These research reports aim to enhance understanding in order to inform policy-making, technology processes, and enhance market intelligence. As technical reports, they aim to describe the basic concepts in the field, and to identify some of the contentious issues, and to highlight the regulatory issues and emerging solutions.

In devising these reports, the research team at the LSE interacted at length with many of the companies whose technologies and techniques are represented in our work. We met with some of the companies to interview them regarding their technologies, and then circulated draft versions of this report to ensure that our understandings matched theirs. The value of this methodology is that we are able to bring all these companies' approaches together to see how they are similar and how they differ, and to identify some of the trends in this complex and dynamic environment. The purpose, therefore, is not to scrutinise the companies' practices but rather to seek an understanding of their practices to piece together a picture of the larger marketplace of activities. We welcomed comments from these companies to inform any misunderstandings we may have had regarding their practices; and while they could review our analysis, they could not, nor did they try to, influence our objectives nor our analyses.

To ensure independence from the companies that were involved in this research process, we also circulated these reports for review by experts across academia, industry, government, and civil society. These experts assisted us throughout the research and revision processes.

There are three *Advanced Internet Policy Reports* in this current series. The first was the Consent Report, which is the result of research done within a partnership of companies and data protection regulators to highlight the greater challenges in online consent. This second report is on online advertising, conducted through interviews and document analysis, and extensive consultative and review processes. The third report is on Cloud Computing, and was conducted through extensive discussions with leading thinkers in the field, a review of the scholarship to date, and a more limited set of interviews.

Table of Contents

Advertising as An Emerging Market, Practice, and Solution

If done properly, internet advertising has the potential to enable near-universal free or cheap access to services and content. It may also pose a significant threat to online consumer protection, and in particular it may hamper an individual's right to privacy.

These are not by any means mutually exclusive options. The path of least resistance would see advertisers seeking to know as much about individuals as possible so as to better advertise to them, and to increase market share. In reality, the environment is far more complicated and the institutions are far more knowledgeable: key players are moving cautiously as they seek to maximise the amount of information they process while seeking to enhance user confidence by assuring some modicum of privacy protections. The primary battle in the online advertising world is certainly about which company can have the largest market share. But in a close second, the battle for online advertising is also about ensuring consumer confidence through the protection of privacy. This report is about the latter case.

The quest for the perfect form of advertising continues, but the range of methods have grown dramatically more innovative. In recent years we have seen the rise and fall, or the trials and failures, of a number of methods to target advertisements to individuals as they use the Internet. In some cases, the more recent methods have been incredibly successful. In this report, we seek to review the leading approaches and techniques, to highlight the most promising and hazardous developments, and to propose a framework of challenges for the industry, regulators, and users.

Online advertising may achieve what marketers have long been dreaming of: generating advertisements relevant to the individual's interests and behaviour, while embedding advertisements into the fabric of an individual’s daily life, and regularly changing the advertisements to target other issues and interests. In the old world of advertising, an ad in a magazine was the same for all readers of that magazine; and when the user reopened the magazine months later that same advertisement would be there trying to re-entice the reader. In the online environment, advertisements can be targeted to the specific user of a website, or the interests of that specific user, and the ad can be changed every time the user visits that site.

Unsurprisingly, online advertising has been rapidly gaining market share as advertisers opt for more measurable approaches to reach ever-larger audiences with higher susceptibilities to consume. Although estimates vary widely, the most widely repeated figures estimate that the global online advertising market is currently valued at $40 billion and is expected to grow to $80 billion by 2010.

This projected growth figures rely on a healthy growing economy, and a significant shift from traditional marketing techniques to new opportunities online. Online advertising offers increased accountability in terms of return on investment and cost per acquisition. With online marketing the targeting is more precise as more can be known about the specific users, and the effect is more measurable than with print, television, or radio advertising as individuals can click on the ads themselves, and these clicks can be traced to an eventual financial transaction. A recent survey by Epsilon shows that 50% of chief marketing officers prefer using data driven

marketing techniques.[1] According to a recent report from the IAB/PWC,[2] in the United Kingdom alone online advertising share grew from 11.4% to 18.7% in less than two years, reaching £1.68bn in the first half of 2008. By comparison, the market shares of traditional advertising platforms, such as TV (21.7% - 22.1%), Press Display (20.8% - 19.3%), Press classified (16.2% - 14.6%) have stagnated or declined between 2006 and 2008.

The focus on targeting specific users and interests means that online advertisers know far more about internet users than newspapers know about their individual readers. Previously readers would be classified based on the item that is being read, i.e. the weekend FT would advertise different apparel than the Sunday Times. Now the online versions of these papers can target advertisements to the individual reader based on his stated profile of interests, his browsing history on the site, his search history on the site, and additional information about him from third parties such as other websites visited, search terms on search engines, etc. With sufficiently targeted advertisements, based on information that these organisations collect on their users, online services will in turn be able to provide free content because money is being generated through this form of advertising, while traditional costs such as print and distribution can be reduced dramatically. How they do this profiling, however, is a significant part of the problem.

Recognizing that online advertising is a potential driving engine of the Internet economy, but also attentive to the implications for consumer privacy, this report aims to provide a brief overview of the advertising market, a review of a variety of targeting techniques currently used, as well as examples of industry practices. Based on the identified advertising trends, we highlight some of the challenges posed by the collection of personal and non-personal data and suggest issues in need of further consideration as we begin to inform users, enroll advertisers, innovate on the technology, and consider regulation.

The Evolution and Landscape of Online Advertising

Internet advertising could provide the foundation for a new economy for online activities. In the dot-com boom there was optimism for free services but skepticism eventually arose due to the lack of income models. While some businesses tried the advertising models back then, the solutions were often criticized on a number of grounds including being cumbersome, ineffectual, and the most basic concern of being 'creepy'. With the immense interest in online advertising with increased sources of data, online advertising through some form of targeting can be the vehicle for true income generation for free services and applications. For many services and sites, it could also permit the transitioning from a subscriber-based e-commerce model towards an advertising-based one.[3]

Targeting is a general term used to describe a more complex phenomenon and has gone through two discernible stages of development. The first stage arose in the mid 1990s in the form of banner ads, or images appearing at the top of websites,

---

[1] http://www.epsilon.com/epsilonstatic/media/press/2008/09/08_cmo.html

[2] http://www.iabuk.net/en/1/introductiontosearchmarketing.html

[3] See Michael Rappa, *Business Models on the Web*, Available at: http://digitalenterprise.org/models/models.html#Advertising

reflecting primarily the content of visited pages. Ad serving was primarily site-centric and did not guarantee a match with the interests of website visitors.[4] This resulted in *contextual targeting*, where ads were targeted based on the content of the website.

The second development stage, which began in the late 1990s, reflects the evolution of online advertising and the use of more sophisticated tracking techniques, such as *behavioural targeting*. In contrast to contextual targeting, behavioural targeting allowed for a much more precise evaluation of consumer attitudes by advertising firms. Techniques would monitor online users' behaviour, which were in turn provided to companies to be analysed so that they could then sell ad-space. That is, users' *clickstream data*, or browsing patterns, could be analyzed for the purpose of increased ad relevance, and which would ideally lead to individuals clicking on the ads, resulting in an increased return on investment for advertisers. Through the tracking of users, *contextual* analysis could be extended across a number of sites rather than being merely restricted to the site offering a single ad.

In the late 1990s we also saw the advent of search engine advertising. This form of advertising links users' search terms with advertisements, and represents the biggest share of the online advertising market, now constituting more than 40% of the global online ad spending four years in a row, since 2004.[5] According to estimates, in the US alone more than 50% of the funds spent on online advertising in 2007 was spent on search advertising. Currently Google is the clear market leader, followed by Yahoo, Microsoft, and AOL.[6] The same spending trend is observed in the UK where paid search advertising currently represents more than 50% of advertising revenues.

Search-based targeting serves text ads on search results pages in response to user search queries. Google, for instance, offers text ads both on Google.com through AdWords and on partners' search engines, including AOL and Ask.com through AdSense for Search.

Despite the power of search advertising, search results pages constitute only 5% of pages on the Internet[7] and purchase decisions are made all over the web. Also, search-based advertising does not deliver display ads, i.e. ads that incorporate both text and graphics, and offers instead text-based advertising.

While there is some overlap between *search* and *contextual* advertising**,** the latter is different in that it serves ads based on the content of pages a user visits rather than search queries. An example is Google's AdSense for Content, which provides ads to visitors of the Google content network, or third-party partner sites. Contextual advertising now allows advertisers to show ads in the context of other sites. Ads are served as they are matched with keyword themes or categories, set by advertisers.[8] The advantage of contextual advertising is that it displays all formats, such as text, video, image, and flash. Furthermore, contextual targeting is generally more privacy

---

[4] This form of advertising already used 'cookies' (see below) but primarily for ascertaining users' particular geographic area.

[5] http://www.emarketer.com/Reports/Viewer.aspx?code=search_feb05&page=5&src=page5_sample_report&xsrc=page5_reportx

[6] http://www.comscore.com/press/data.asp

[7] Google AdWords Learning Center: http://www.google.com/adwords/learningcenter/

[8] There are some limitations on the use of categories that may be associated with sensitive personal information.

friendly, as it does not use as much personal data to serve ads but rather analyses the content of visited websites.

Behavioural advertising represents a small but important part of the advertising market, as it is also probably the most controversial targeting technique. It is used for the categorization of likely consumer interest segments, inferred through the collection of personal information on the users, including user browsing patterns and publicly available information.[9]

The Techniques

This entire market hinges on some essential tracking technologies. Standard behavioural targeting on websites is executed by placing a string of text, called "cookie", on the user's computer each time they visit a website. The cookie is assigned a unique cookie ID, held by the database of the website, allowing it to recognize users when they revisit the website. The danger of this practice is that it may lead to the accumulation of significant amount of data about a single user.

Behavioural targeting also enables the building of user profiles based on geo-location, demographics, lifestyle and affinities – all enabled by cookies. Some companies even go one step further and target users based on data from customer relationship management systems (CMS) within websites. This enables extremely precise targeting as the system recognizes when a customer is searching for a product or bidding on one (e.g. eBay) and serves relevant ads. Once customers have selected products and have them in their basket, behavioural trends can be used to present optimized cross-sell (similar) and up-sell (higher price) products.[10]

Naturally, privacy concerns arise. Consumers have expressed their concerns about invasive practices through a variety of mechanisms, including moving their shopping elsewhere, complaining to companies, running campaigns, using technologies that circumvent these practices, and even complaining to regulators. This can be seen through the failures of business models that relied upon intrusive techniques in exchange for free services including free computers or free phone calls,[11] and advertising methods such as pop-up advertising[12] and spam.

Industry leaders are increasingly sensitive to privacy concerns. In order to prevent correlation between different data sets and to increase privacy companies use different methods of de-identification and anonymisation. These consist of structural separation of personally identifiable information (address, name, social security number) from anonymous data, such as unique ID numbers assigned to cookies placed on users' computers and used to serve targeted ads. This results in behaviourally targeted ads, served to users that the provider cannot easily identify, as they are hidden behind unique numbers.

---

[9] It is important to note that display advertising can also be delivered based on information unrelated to user's behavior on different websites, such as optimization of ad placement based on it popularity on certain types of websites.

[10] http://www.iabuk.net/en/1/behaviouraltargetinginsearch.html

[11] e.g. 'Internet war hots up', BBC News, July 23, 1999, discussing the rise of advertising funded free-telephone calls with 30 second advertisements every 5 minutes, or free computers if users are subjected to advertising.

[12] e.g. 'Pop-ups chase eyeballs', BBC News, July 24, 2002, discussing the annoyances and concerns in pop-up advertising, particularly geared towards children.

One potential hindrance to achieving full transparency is a technique, developed by DoubleClick more than 10 years ago, called 'retargeting'. Retargeting allows you to identify your past visitors when they are elsewhere on the web and drive them back to your website. This is usually done by advertising networks who serve ads to thousands of users and can easily track their online movements. For example, if you visit a website that sells climbing gear, you might continue seeing ads from that site while browsing other web properties for a week. Unfortunately, few have been willing to openly talk about retargeting and as a result no efforts have been made by advertisers to promote this practice and explain to users what it entails.

Technological safeguards are increasingly important, but despite all of these protections and as seen above, there are still risks from the 'creepiness' factor, where the level of advertising to which users are subjected alarms them. Advertising that is too targeted to a user's profile could alienate that individual. Moreover, if consumers are not sufficiently empowered to decide how their personal information is processed, we run the risk that they will reject this powerful business opportunity.

The Landscape of Advertising: Publishers vs advertising networks

In addition to different types of targeting techniques, advertising models also differ depending on whether ads are served by a publisher (first party) or through an advertising network (third party). Publishers are big web portals, such as AOL and Yahoo!, that attract large Internet audiences and can to a certain extent be self-sufficient in their ad serving capabilities.

Advertising networks, on the other hand, reach thousands of partner websites across the web and thus exponentially increase not only the reach but also the level of precision in ad targeting, as they capture a much broader range of user activity. Network advertising offers the benefit of serving ads on less frequented pages or less desirable spaces. While newspapers like the Washington Post and the New York Times are still able to sell expensive ad space on their home pages, they also use ad networks, which allow them to utilize space on less attractive or newly created pages, which would otherwise remain ad-free. According to a study by Bain & Company and the Interactive Advertising Bureau,[13] in 2007 30% of the ad spaces sold on premier publishers' sites came from networks (see Figure 1), up from 5% in 2006. Despite the significant increase of impressions sold through intermediaries, network monetization has been low, with revenue lower than 2% of total display ad revenue. One reason for this discrepancy is the significant price compression experienced by large publishers relying on third party ad networks for the monetization of their inventory. The creation of branded vertical ad networks can thus be of high strategic value for large publishers.

[13] Bain & Company and the Interactive Advertising Bureau, *Digital Pricing Study*, 2008. Available at: http://blog.adify.com/2008/08/.

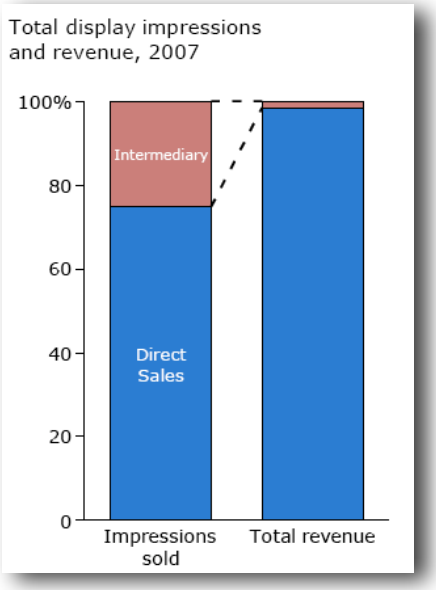

Figure 1. Total display impressions and revenue, 2007[14]

The only advertising segment that does not currently make use of intermediaries is video advertising. Demand for video advertising is still higher than the supply.

Complexity of advertising actor partnerships

Despite attempts to put advertising models in different functional categories, today's online advertising constitutes a complex web of partnerships, mergers, and acquisitions. As innovation continues, companies strive to employ newer and more sophisticated targeting techniques. In order to achieve this, they utilize the combined expertise of ad serving platforms, mobile media networks, video serving networks, ISPs, and other.

To begin with, there has been a great deal of consolidation in the marketplace. In the past two years alone, AOL, Google, Microsoft and Yahoo! have all added behavioural targeting firms to their portfolio: Tacoda, DoubleClick, aQuantive, and BlueLithium, respectively.

AOL's Platform A, launched recently to simplify trading experience for both publishers and advertisers, is another one example of the industry's attempt to deal with the complexity of the environment. AOL consolidated TACODA's behavioural advertising techniques with the Advertising.com network, and combined this with AOL's web presence. This significantly expanded not only the reach of AOL's advertising but also the amount of personal information collected.

Collaborations can also take place across infrastructures. The collaboration between advertising technology firms and internet service providers has proven to be the most controversial. For instance, Phorm's *Webwise* technology monitors the transactions of users of an ISP to identify an individual's interests. By positioning their technology at the ISP, Phorm is able to monitor the grand majority of transactions made by an individual on the web, rather than merely limit that profiling activity at specific sites.

The multitude of advertising technologies and platforms become even more relevant

[14] Id.

when one considers user notification, consent, and *opt-in* and *opt-out*. After the acquisition of DoubleClick, for example, Google implemented a DoubleClick ad serving cookie called DART, allowing for a single user opt-out point that applies to both DoubleClick's technology and Google's content network.[15] Yet another example of the industry's simplification of services is Yahoo!, who launched in late 2008 a new ad-serving platform called APT. APT will allow advertisers to sell targeted ads on their and their partners' sites with one simple transaction. Yahoo! aimed to have 784 newspaper partners in the US using the system by the end of 2008, and is promising to take the system worldwide in 2009.

However, some experts have warned against a soon-to-be overcrowded market.[16] While in the past companies wishing to advertise could choose between the different offerings of the giant display advertising companies, now there is less and less differentiation between their products and often there is even cross-over, i.e. they sell across the same sites.

Industry practices

The public policy debate around this issue used to be framed as *those-companies-who-want-to-make-money* vs. *privacy-and-consumer-advocates-who-want-everyone-to-pay-for-services*. This situation has changed dramatically as companies, for the most part, have taken privacy concerns under some consideration in their design and policy processes.

As a result, a key differentiator in the operation of these advertising platforms comes down to how personal information is processed. Some make use of extensive personal information, even while claiming that they are not doing so, so the very definition of *personal information* comes into question. Below we review the practices of some of the leading and innovative firms to identify the key dynamics that must be addressed to enhance consumer confidence and end-user privacy while providing a minimal amount of information to third parties to target advertisements.

An important element in this discussion is data retention, and particularly that of search queries, IP addresses, and cookies often used by search engines for targeted advertising. How long personal information is kept by advertisers is a highly contested issue, and has given rise to significant regulatory interest. While the variance in periods of retention is interesting, there are by far many more variances in how companies actually collect personal information, and how they protect that information through techniques such as de-identification, one-way cookie identifier encryption with the use of ephemeral keys, or partial anonymisation through the removal of octets of the IP address (for comparison of company practices, as done by Microsoft see Fig. 2).

The method and strength of anonymisation are equally if not more important than the time period after which data is anonymised. For example, if cross-session identifiers remain, data can possibly be correlated and maybe even linked to an individual at a later stage. That is, a common misconception is that if IP addresses are removed

[15] The single opt out button can be found here: http://www.google.com/privacy_ads.html

[16] Farber, A. (2008, October 2). Ad giants back with new act. (cover story). New Media Age, Retrieved November 3, 2008, from Business Source Premier database.

from search logs, for instance, then the searches have been 'anonymised'. Often, however, search firms may remove/modify the IP address but introduce a unique identifier to differentiate between the searches of UserA and UserB. As was uncovered by the AOL search log scandal in 2006 when AOL released its 'anonymised' search logs, by removing the IP addresses, it was still possible to re-identify individuals by aggregating all of an individual's searches and indexing them by the cross-session identifier.[17] Complete anonymisation thus seems to offer better protection for user search data used for advertising. It is also arguable that only complete deletion of search logs can offer adequate protections. Deletion of the entirety of the IP address, as well as all other cross-session identifiers, such as persistent cookie IDs, prevents any correlation between anonymised and personally identifiable information. Currently, this seems to be the most reliable method, especially when coupled with the separation of search queries from personally identifying information (de-identification) during the initial stage of data processing.

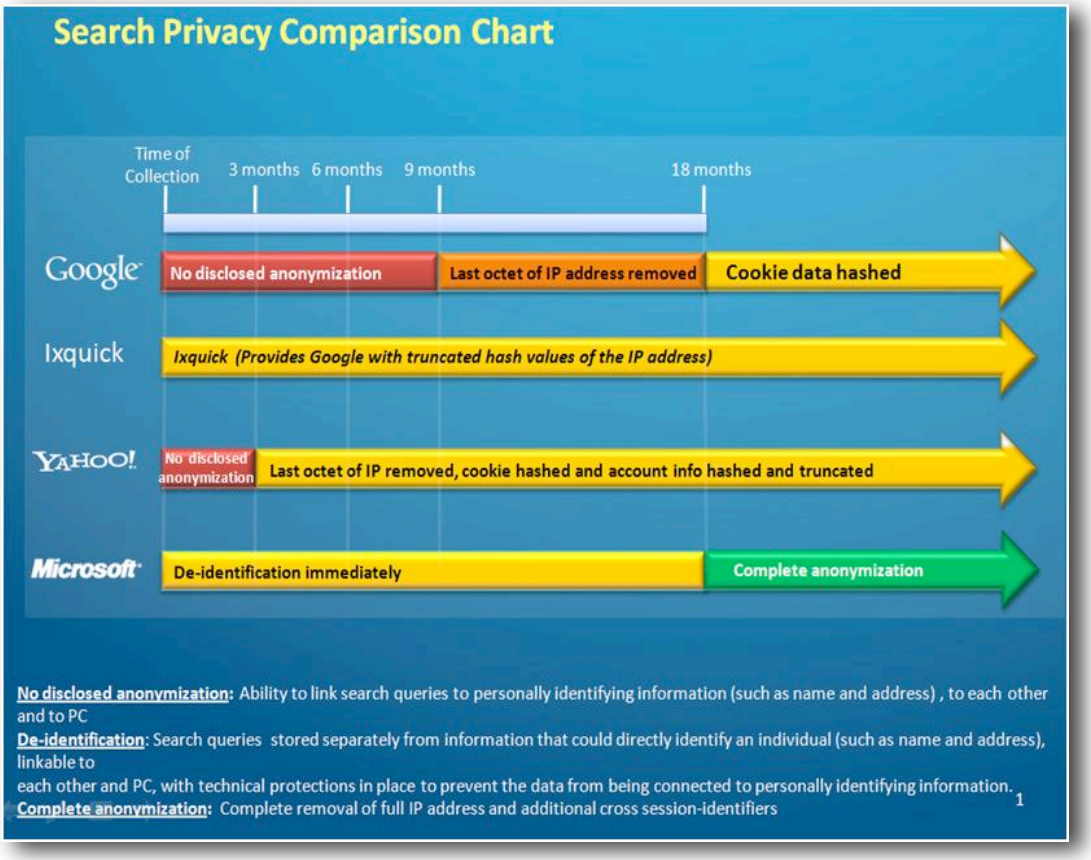

Figure 2. Search privacy comparison chart[18].

As Yahoo! showed in their retention announcement in November 2008, the retention of search terms is merely one part of the retention problem. Yahoo! was the first company to announce that they were going to also delete advertising profile information after a certain period of time. Therefore many of the issues that apply to

[17] 'AOL apology for search data error', BBC News, August 8, 2006, available at http://news.bbc.co.uk/1/hi/technology/5255732.stm

[18] Available at: http://microsoftontheissues.com/cs/blogs/mscorp/archive/2009/02/10/comparing-search-data-retention-policies-of-major-search-engines-before-the-eu.aspx For the sake of disclosure, we used our research to advise Microsoft on some of the earlier drafts of this graphic.

one form of advertising may also apply to the others.

In this section we provide an overview of some of the practices in online advertising. This is not by any means a comprehensive review of the products and services, but we cover some of the practices that we were able to identify. We are not presenting these examples to highlight some companies over others, or to shame them. Rather we selected the most advanced and innovative practices that we could research, through document analyses and interviews with the companies. We distinguish the range of services and products by grouping them together as follows:

- Advertising in search
- Advertising in spaces 'owned' by the companies serving advertisements
- Advertising infrastructures through methods such as 'deep packet inspection',

and then covering a variety of other practices in using and protecting personal information.

Search companies

Microsoft

Microsoft anonymises all search query data after 18 months, unless they receive user consent for a longer period of time. At the time of writing, this was the longest retention 'period'. However, this policy includes removing the entirety of the IP address and all other cross-session identifiers, such as cookie ID or other machine identifiers, from the search terms. Furthermore, Microsoft stores live search query and click stream data separately from user account information, such as name, telephone number, e-mail address, etc.

As mentioned above, the process of *de-identification* used by Microsoft to structurally separate personal from non-personal data in its ad serving system is an example of privacy by design. For that purpose Microsoft uses three different cookies – the machine unique ID (MUID), the Windows Live user ID (LiveID) and the “Anonymous” ID (ANID). The latter two are part of the system that separates personally identifiable information from data used for ad personalization.

MUID is a standard cookie with randomly generated unique identifier (Fig. 3). In turn, LiveID is a user-based ID, assigned on a per-login basis to users who have already established a relationship with the website. Once placed on a user’s computer, this cookie is used to remind an MSN or Windows Live service to continue to grant access to that user. When the user logs out, the LiveID cookie expires (unless the user has opted to “save the password”).

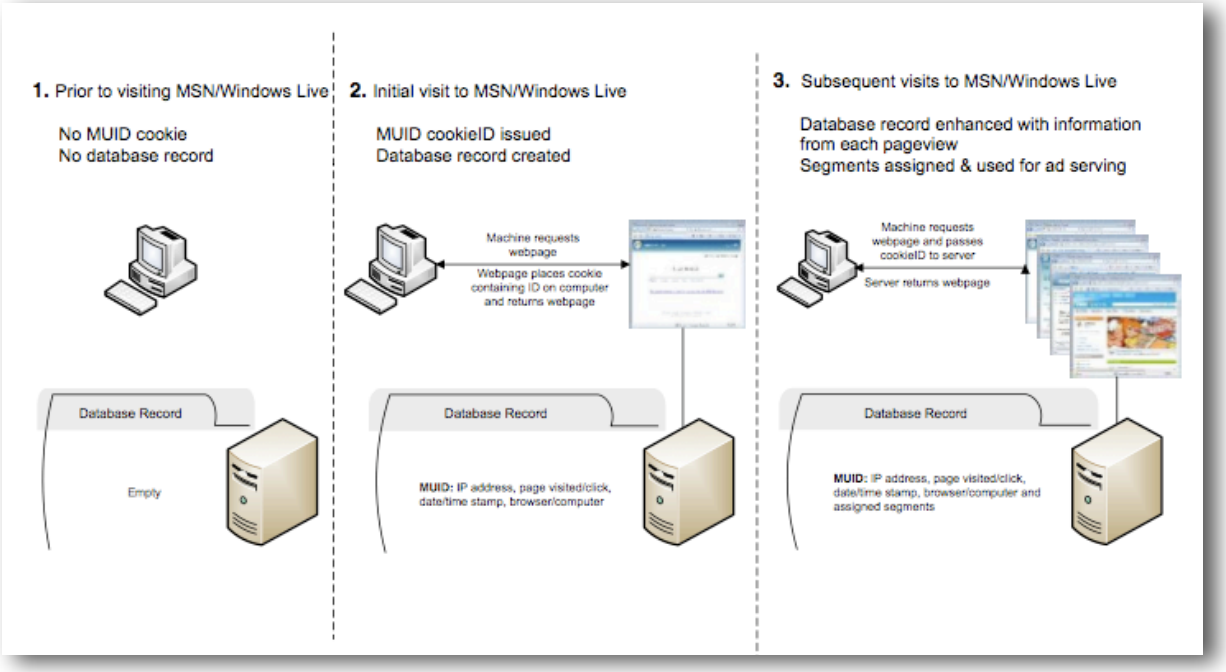

Figure 3. The Machine Unique ID (MUID)[19]

Therefore, the ad targeting and ad serving capabilities are based on the ANID cookie. When a user registers with a Live or MSN service a LiveID and an ANID are created simultaneously. The ANID is derived by applying a one-way cryptographic hash function to the LiveID ensuring that there is no practical way of deriving the original value from the resulting hash value. When created, each ID is put in a separate cookie on the computer. Only some demographic information that is highly unlikely to be useful in identifying a user is indexed on the ANID. Thus Microsoft's ad serving infrastructure only uses data associated with the ANID and not the LiveID. When a user logs out, the LiveID is deleted from her computer. However, the ANID cookie remains until a different Windows Live account is accessed from that computer, until the user deletes the cookie or until it expires.

Microsoft also offers two opt-out options (Fig. 4). Its users may chose to opt-out of receiving ads based on their search logs and page views in a particular browser or computer they use and do this without the need for authentication. Alternatively, they may choose a more 'roam-able' opt-out that is associated with their Windows LiveID and thus requires authentication. Once they have opted out using the second option, users' preferences will be added to their account settings and will persist until another LiveID user logs in on the same computer. The disadvantage of both options, however is that if a user inadvertently deletes the opt-out cookie, she may not realise that her preferences have reverted back to the default settings and targeted advertising is once again taking place.

[19] *Privacy Protections in Microsoft's Ad Serving System and the Process of "De-identification*. Microsoft Corporation, October 2007.

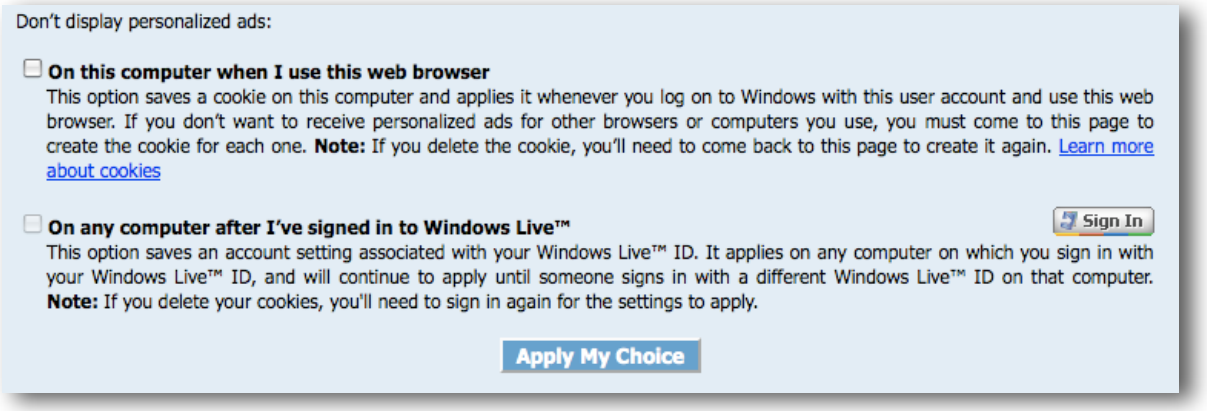

Figure 4. Choosing not to receive targeted advertising from Microsoft[20]

AOL

As pointed out earlier, there is some confusion about the exact roles and responsibilities of different players in the online advertising industry. The reason for this is the complexity of the technology and the resulting complexity of partnerships. As a global Internet and media company, AOL is naturally involved in online advertising. What most users are not aware of, however, is that AOL serves ads primarily through its partners (e.g. Google Ad Sense for Search) or smaller companies it has acquired but which are in essence separate entities. AOL has two behavioural advertising networks - Advertising.com and Tacoda. AOL itself is a publisher, but recently created Platform A in an attempt to consolidate all of its advertising businesses. Platform A includes a range of smaller ad networks or ad servers amongst which Adtech (ad serving platform similar to DoubleClick), Quigo, and others. Adtech is a company that doesn't sell ads but rather offers their technology to publishers who sell their own ads. As a result, it does not have any rights to user data. Quigo Technologies, on the other hand offers self-service advertising services to publishers who can themselves sell links.

The differentiation between unauthenticated and authenticated users is key for the understanding of privacy controls and safeguards in online advertising. In the case of AOL search targeting, when an unauthenticated user conducts search, his or her query is sent from the AOL gateway to Google, which powers the search. No redirection of the user's browser occurs at any point of the search query transmission to Google. Google's role is to simply create an index of the web and deliver it to AOL upon request. When a user generates a search transaction, AOL's log file records details regarding the user's cookie, the IP address, and search terms. The raw log file is then deleted after 7-14 days, the IP address in search logs is deleted after 30 days and granular search data is kept anonymously for 13 months. The entire set of data is discarded after 13 months.

When users have authenticated themselves by logging into the AOL portal, an encrypted version of their screen name - similar to Microsoft's ANID - is logged with the data. AOL's ad serving infrastructure however is a separate entity and actual user search terms are not shared directly with the advertising portion of the business. Unlike Microsoft who have their own ad serving mechanisms, until recently AOL used the services of DoubleClick, the display ad serving technology company.

[20] https://choice.live.com/advertisementchoice/Default.aspx

AOL now uses their own ad server AdTech or/and DoubleClick. As a result of the structural separation between the different technology companies, no sharing of authenticated user information between AOL and the ad server takes place. AOL passes to the ad server only the criteria of the ad the server should deliver. When a user does a search query, the latter can be used to tailor an ad on AOL or on the ad network, creating a user profile.

AOL members have access to a preference to elect not to have search terms used for personalization, including advertising. Other privacy efforts include an educational campaign called "Mr. Penguin" (Fig. 5), aimed at raising awareness about behavioural advertising.[21]

Figure 5. AOL's Mr. Penguin video.

Service Providers: Companies that Own their 'Space'

Google

Until recently, Google conducted online advertising based solely on aggregated non-personal information, such as the aggregate number of users who searched for a particular term or who clicked on a particular ad. Their advertising products were contextual in nature and served relevant ads based on users' search queries or/and the content of websites users visited while browsing the web.

In March 2009, Google announced that it would start offering behavioural advertising, currently embraced by most of its competitors. It might be argued that Google's involvement in behavioural targeting was just a matter of time, especially as previously they had openly expressed their support for "the efforts to establish strong self-regulatory principles for online advertising that involves the collection of user data for the purpose of creating behavioural and demographic profiles".[22] Another sign of their interest was the acquisition of DoubleClick, a display ad-serving company that specialises in behavioural targeting.

But before we look into Google's newly evolved advertising model, let us elaborate on some of the particularities around their search and contextual techniques. Search targeting, offered through AdWords, is an advertiser-facing service, which delivers ads based on key words and shows adverts next to search results. Google serves

[21] http://corp.aol.com/o/mr-penguin/

[22] Davidson, Alan, *Google Responses to Questions from the House Energy & Commerce Committee*, August 8, 2008.

search ads on Google Search or on the sites of its AdSense for Search partners. The results are based, amongst other factors, on recent search queries, standard log information, including cookie information, IP address, browser type, operating system and the date and time of the user request.

Google also offers language and general location (region, city) targeting, making ads geographically relevant. Both language targeting and geo-targeting are IP-based and no user information is retained. In terms of IP addresses being considered personal data, Google believes that a more nuanced analysis is needed, one that considers the context in which IP addresses are used, in order to apply the correct legal characterizations.[23] According to the company, an IP address should not automatically be considered personal data if a website has no ability to identify the user (as the IP address only has true meaning to the ISP that issues it to the subscriber/user). Furthermore, Google's search engine is not responsible for the creation of content on the web, nor are its search results intended to form a profile of any individual. Rather, Google responds to user search queries with links to what appear to be relevant pages thus effectively serving ads to relatively 'anonymous' users.

In an attempt to reassure users, policy makers, consumer groups, and privacy advocates, and following the recommendations of the European Commission's Article 29 Working Party, on September 9, 2008 Google announced that it would reduce the search data retention time from 18 to 9 months (see figure 2). The question is whether only logs of "unauthenticated users" are anonymised. What is also not clear in the case of "web history" accounts is whether data is removed from search logs as well as stored databases that are used for auditing and service improvements. Similarly, Google has not clarified whether they anonymise all logs regardless of whether a user is logged in or not. These criticisms could also apply to other search engines.

In addition, Google's content network,[24] a large number of sites and other web-based products who partner with Google to display AdWords ads, allows advertisers to serve text, animated, and video ads, targeted to the content of websites and based on the associated keywords. Google offers this type of behavioural advertising primarily through AdSense for Content. The latter is a publisher-facing service, which allows websites to monetize clicks or impressions by displaying targeted ads on their property.

For the purpose of contextual ad serving Google also partners with social networking sites. This practice has proven quite controversial despite Google's reassurance that the advertising model used in these cases is not a classic AdSense service, as content from profile pages is not used. Instead, according to Google, social networking websites send requests to Google with the relevant information, containing no PII, and Google serves the most appropriate ads. All information Google keeps from each such transaction is a unique ID, provided by the partner website.

To serve ads that are relevant and tailored to users' interests, Google uses

---

[23] Google Response to the Article 29 Working Party Opinion On Data Protection Issues Related to Search Engines, 8 September 2008, p. 6.

[24] The Google content network reaches over 75% of unique internet users in more than 20 languages and over 100 countries.

information about their activity on AdSense partner sites or Google services that use the DoubleClick cookie. In the above cases, however, users are presented with the option to opt-out of data collection. They may choose to opt-out of the DoubleClick cookie, called DART, in three ways: 1) by visiting the DoubleClick opt-out page;[25] 2) by visiting Google's Privacy Center[26]and clicking on the opt-out button; and 3) by opting out of the DoubleClick cookie at the NAI opt out page[27]. The DART cookie is not client-specific. This means that, if a user chooses to opt-out of DoubleClick's DART cookie, their opt-out cookie will be effective for all websites and advertisers that use DoubleClick's ad-serving or search technologies, including the Google content network.

In this context we may better understand Google's most recent innovation in behavioural targeting. In March 2008 the company announced that it will now offer 'interest-based' ads, based on the types of pages visited or the type of content viewed. Equipped with this information it will then associate a user's browser with all relevant interest categories (Fig. 6).

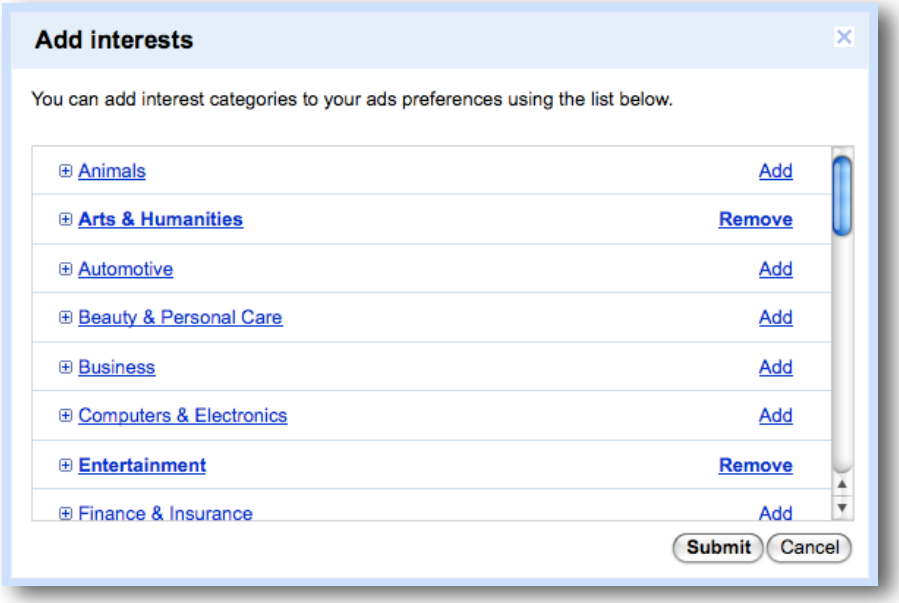

Figure 6. Google's ad interests pop up window. Can be accessed from the Ad Preferences Manager.

In contrast to many other behavioural targeting advertisers, Google has made a remarkable effort to offer users increased control over their interest categories, i.e. the categories based on which ads are served. In a way, the company has adopted a privacy-by-design approach and will allow users to view their current interest categories through an 'Ad Preferences Manager' tool and modify them if they wish to do so instead of relying solely on opt-out cookies. Showing users what data has been collected about them is not a novel idea, as advertising data companies such as BlueKai already offer similar services, but the scale of Google's reach makes this a very important development.

---

[25] www.doubleclick.com/privacy

[26] www.google.com/privacy_ads.html

[27] www.networkadvertising.org/managing/opt_out.asp

Google's behavioural targeting system also offers two advances in the opt-out mechanisms. Unlike rival marketers, in addition to the traditional cookie-based opt-out, the company will offer persistent opt-out through a browser plug in (Fig. 7).

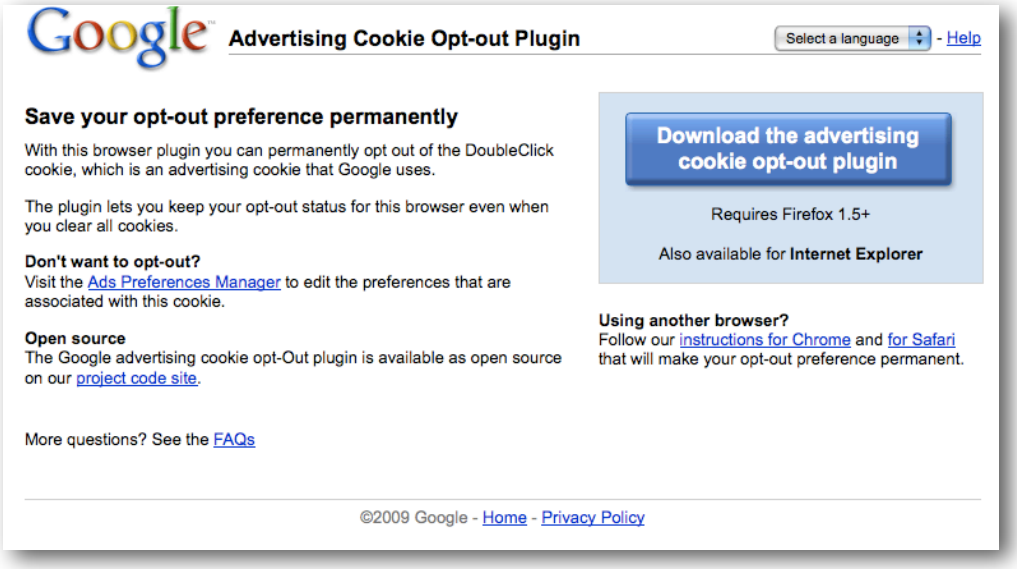

Figure 7. Google's browser-based persistent opt-out.

This persistent opt-out ensures that if the individual accidentally deletes his or her opt-out cookie, the cookie will be reset by the browser plug-in, thus ensuring that the user is not tracked after he or she has clearly stated such an intention.

Second, Google will offer prominent notices to targeted users through an 'Ads by Google' link appearing on the ads themselves. eBay is another advertiser which has been using similar prominent notices as part of its AdChoice program for over a year now. Google is however going one step further by placing on-ad links not only to their privacy policy but also to the advertisers behind each ad (Fig. 8). One possible disadvantage of this initiative could be that users may find the additional text and links too distracting. Furthermore, advertisers might not be too eager to give up part of their valuable ad space as the notification text overlaps the bottom of ads.

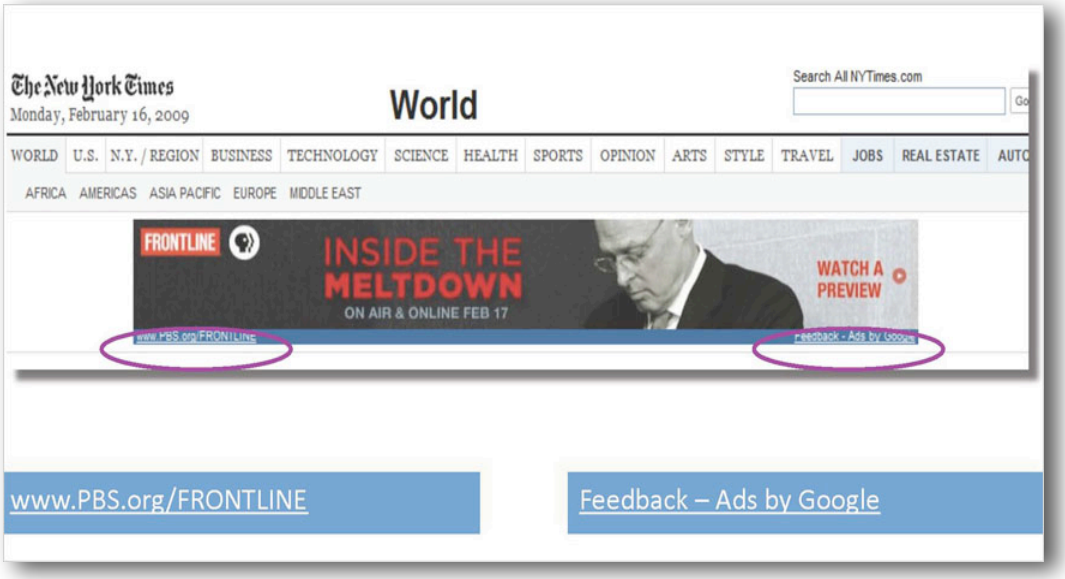

Figure 8. Google's new behavioural targeting on-ad notification.

The challenge for Google is to now enforce this practice across all their services. This is no simple task, but since Google has such a strong position in the marketplace,[28] it is possibly uniquely able to do so.

eBay

eBay is a global provider of e-commerce (through eBay.com, World of Good, Kijiji), payments and financial services (through PayPal), and communications (through Skype). While eBay is not a search engine, it offers product search on its site. eBay is not an advertising network either but it manages promotional content while holding a huge personal information database, which makes its ad targeting practices relevant to the present discussion.

eBay began placing retargeting[29] eBay ads in July 2007 and placing targeted third party ads on the eBay site in November of the same year. eBay does not do one-to-one direct marketing. It records users' activities and classifies them into segments, in effect achieving mass customization. Due to their business model, however eBay tends to collect more personal data – including financial information - than regular advertisers. In this sense, they differ from advertisers that tailor ads based solely on cookies.

Last year eBay initiated the labeling of its ads through AdChoice as it recognizes that its unique ability to identify users when they are elsewhere on the web since they are personally registered on it's website implies additional accountability.

The AdChoice program is a privacy preference for targeted advertising based on eBay's e-commerce-related customer data. It provides users with more information and control of their privacy preferences. AdChoice has two distinctive components:

- *Prominent notice* with on-ad links to a pop up with more additional information and opt out (Fig. 9)

Figure 9. eBay's on-ad notice.

- *Persistent opt-out in "My eBay"* (Fig. 10)

[28] See, for instance, the recent report by Joshua Gomez, Travis Pinnick, and Ashkan Soltani, "KnowPrivacy" (June 1, 2009), available at http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf

[29] "Retargeting" is personalization of own eBay ad placement across the web.

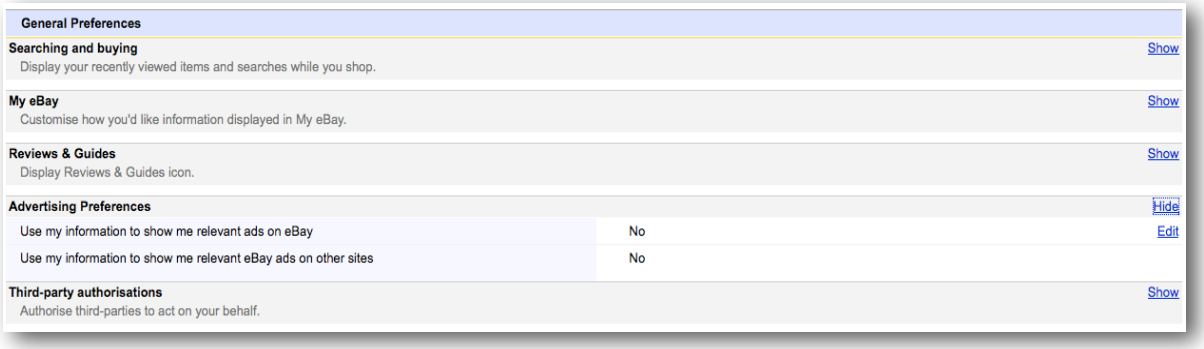

Figure 10. eBay users can opt-out of targeted advertising from the *My eBay* 'Site Preferences' page.

When ads are served on eBay, there is a link next to the ad that opens the Ad Choice pop up. On ads promoting eBay but served on third party websites the link is embedded within the ad itself (see bottom left corner of above ad, Fig.9). eBay is able to do this as they *own* the ad space on their own ads or the frame around the ads on their sites. Depending on the type of ads users receive, they are informed that they can prohibit eBay from using their eBay-collected information for (1) targeted ad delivery on the eBay website from eBay's partners; and (2) targeted advertisements on third party-sites from eBay. It is important to note that the choices provided by eBay are not about limiting the collection of the data itself, which eBay must process for the purpose of providing their sites and services, but for the use of that data for targeted advertising.

The challenge in e-commerce is that tailored content is often difficult to distinguish from advertising. E-commerce platforms collect personal data for a number of legitimate purposes, such as accounting, tax purposes, reporting, and prevention of fraud. Thus, providing user opt-out for the collection of PII is not a workable solution. For companies like eBay advertising and content are thus interrelated (Fig. 11).

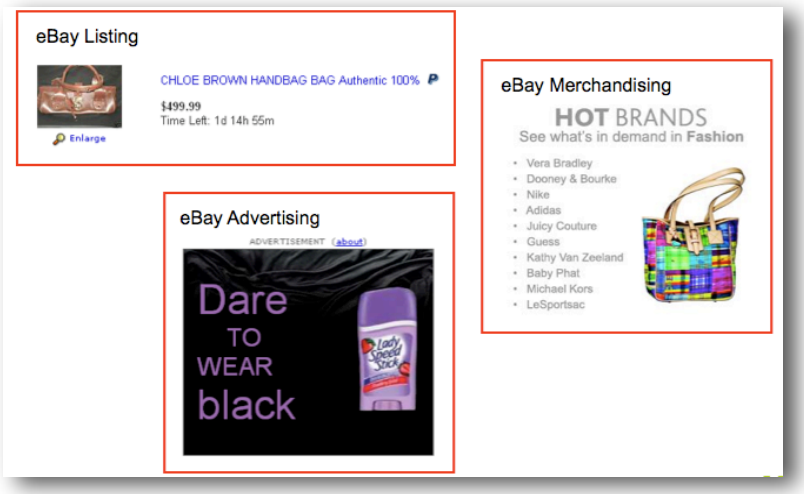

Figure 11. Advertising and content are hard to tell apart on sites like eBay.

Users do a number of things on eBay. They browse listings of products generated by other users/sellers (content) but those listings are also commercial in nature. eBay also offers information to users about other relevant listings, promotions and services related to the eBay platform. And finally, eBay serves ads in banner format from

other eBay services or unrelated third parties.

The ability to use eBay's services on other websites and even off the web on mobile phones is referred to by the company as *Distributed eBay*. When eBay provides *Distributed eBay* services, they try to ensure that users are aware of being subject to eBay's terms and conditions, including the privacy policies. When behavioural data is used by eBay for third party advertising or off-platform advertising, an on-ad link to additional information and preferences is provided.

It is difficult to differentiate between first and third party advertising when ads are served by a company on other websites or even off the web, on mobile devices. Could the behavioural information collected by eBay through widgets be considered as collected by a third party site? Or do the directly conveyed policies and disclosures suffice to classify this as first party data collection? This merits further research and discussion.

Facebook

Facebook's default privacy settings limit the information displayed in users' profiles to their networks. Users may further control the information they share and the people they share it with through the privacy settings on the Privacy page. Facebook has also introduced "friends lists", which allow users to create subsets of their confirmed friends who may see certain content (Fig. 12).

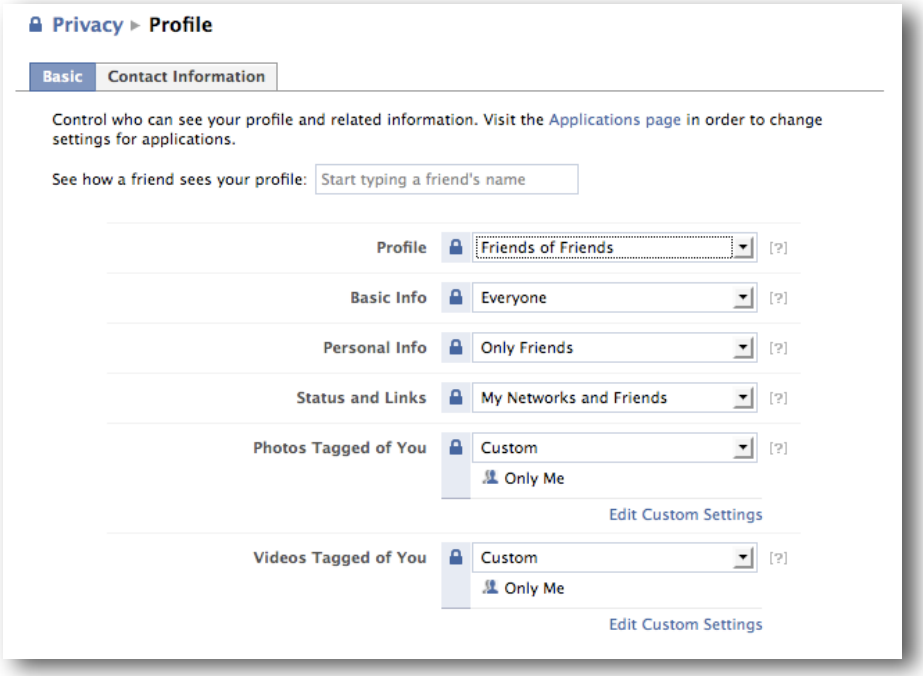

Figure 12. Control who can see your profile and personal information.

In general, Facebook discloses to third parties[30] non-personally identifiable information from users' profiles. This information is used for data aggregation serving behavioural advertising, which Facebook believes benefits their users. Although

[30] A user may decide to willingly share PII with an advertiser by entering a contest for example.

most Facebook data is collected transparently in personally identifiable form[31], only non-personally identifiable information is disclosed to advertisers.

Facebook has two advertising services – Facebook Ads and Social Ads, both introduced in November 2007. Facebook Ads is a self-service targeting advertising structure based on non-personally-identifiable key terms derived from profile data. Facebook Ads is a standard targeting platform that ensures that users with certain characteristics and interests are targeted with the appropriate messages and ads. Social Ads, on the other hand, allow for paid promotion of certain interactions users take online to those users' friends in conjunction with an advertising message. Social ads are triggered by user actions and presented to confirmed friends only, rather than Internet users at large. Third party advertisers have no access to personally identifiable information but rather they receive notification that a certain number of users have taken relevant actions. For example, a political campaign could pay to promote to a user's network the fact that they have become a supporter of a political figure on Facebook. If users do not feel comfortable with this service, they can choose to opt-out of appearing in their friends' social ads from the privacy settings page (Fig. 13).

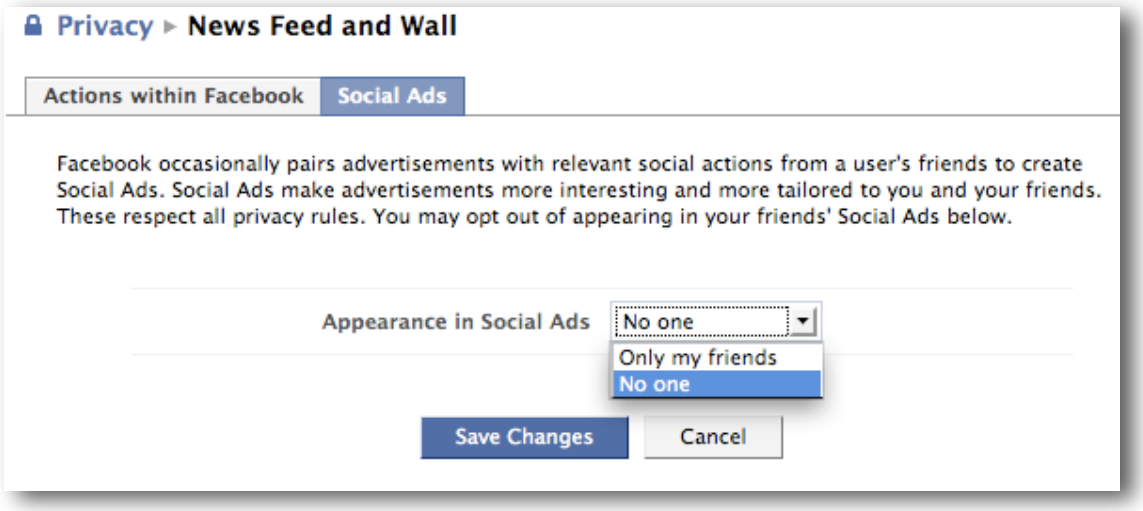

Figure 13. Facebook's Social Ads opt-out page.

The launch of Facebook ads was accompanied by the introduction of another product, called *Beacon*. Beacon allows users to bring actions they take on third-party websites to Facebook and share them with their networks. Participating parties do not pay for the use of Beacon, nor do they need to purchase Facebook Ads in order to use it. Although Facebook did not sell or share any personally identifiable information from user profiles, the product became quite controversial due to the lack of adequate user control over the information about their activities and whom it was shared with, at least within Facebook's space. Just a few weeks following the launch, Facebook modified the system and made it fully opt-in.

Connect is the newest Facebook product allowing for interactions with third party websites. It allows users to connect their Facebook account with any partner Web site using trusted authentication. Once a website publisher knows that there are

[31] Currently only four pieces of personal information are required to establish a Facebook account – e-mail address, birth date, age, and gender.

Facebook users on their site, the publisher will be able to start publishing Feed stories in Facebook for all actions the users take there. For this to happen a user needs to have authorized Facebook Connect, and to be logged into Facebook. If a user is not logged into Facebook no feeds will be posted.

Embedded into the infrastructure: Deep packet inspection based advertising

Phorm

Phorm offers technology, which analyses Internet traffic and provides targeted advertising based on information provided by partner ISPs. There are currently three ISPs working with Phorm in the UK – Virgin Media, BT, and Talk Talk.[32]

Phorm technology within UK ISP networks is offered to customers as a package of security and customization features called Webwise (Fig. 14). Webwise assigns a unique, randomly generated number to a user's browser, and then matches the categories of browsing activity with relevant advertising. When the user's interests match an advertiser's category, the user can see a relevant ad. What allows Phorm to do this is the creation of a copy of the Internet traffic that passes between an end user and a website[33] with the help of partner ISPs.

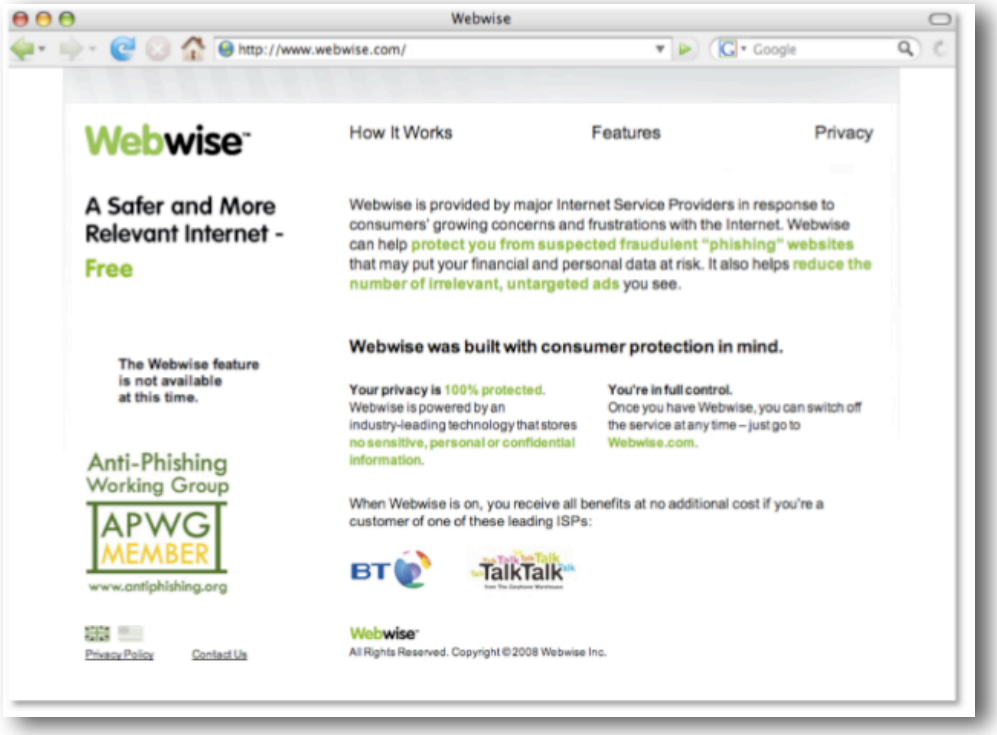

Figure 14. Webwise front page features a prominent link to the privacy policy.

The specific technique used by Phorm to conduct traffic analysis, called deep packet inspection has generated serious public concern.[34] The 'DPI' is operated by ISPs,

[32] http://webwise.phorm.com/current_pilots.php

[33] The Phorm "Webwise" System, Richard Clayton, May 18th, 2008.

[34] See http://www.badphorm.co.uk and http://www.nodpi.org as examples.

while Phorm itself processes a subset of Internet traffic information and does not process IP addresses. Furthermore, ISPs analyze traffic from Port 80, which is the conventional port used for web browsing using HTTP protocol. If the traffic is not HTTP traffic or is encrypted, it is ignored by the Phorm system. Each user is assigned a unique 16 byte ID, which is a random number and does not contain any additional information. This ID is only used by the browser and Phorm and is not shared with third parties. Recently Phorm has announced that it will encrypt these IDs to ensure that they can not be matched across websites. The URLs of sites that a user visits are kept for the minimum time period necessary for channel matching to take place. Channel matching is essentially a correlation between pre-determined advertising categories and UIDs, accompanied by a time stamp. The channels themselves contain no personal information and advertising categories that might relate to sensitive data are excluded.

Phorm has an opt-out system that claims to offer users control over their participation in the system. If a user chooses the generic opt-out, their webwise.com cookie will be discarded and no UID will be associated with their account. No subsequent redirection to webwise.com will occur, although a user will continue to be tracked under his or her old identity for up to three days. They will however acquire a new UID for all new websites they visit. Phorm has also committed to giving users notice and banner status reminders of their participation in the program – regardless of whether they have opted-out or opted-in.

The main challenge remains the relationship between Phorm and their partner ISPs, who are in possession of extensive and highly personal data of their users, including IP addresses, financial data, names, addresses, and e-mail addresses. Therefore, consent at the network level is critical. It is our understanding that at least one of the participating ISPs is experimenting with consent mechanisms that go beyond mere cookie-related procedures. Notwithstanding all technical safeguards, further and stronger leadership is required in dealing with ISPs in regard to their deep packet inspection practices.

NebuAd

NebuAd provides online advertising in partnership with ISPs, in a similar arrangement to Phorm. They argue that they do not collect personally identifiable information to provide services and allow users to exercise their choice of whether to participate in the program or not through opt-out. Users may opt-out both prior to using the service or at any time thereafter, and opt out persistently. NebuAd targets users through anonymous profiling based on a subset of HTTP traffic and matches these profiles with predefined marketing categories. The firm argues that none of the data collected contains personally identifiable information as secured-web traffic is not analyzed, and profiles are anonymous. User profiles do not contain real IP addresses or URLs navigated.

NebuAd uses a proprietary enhanced opt-out system, which makes opt-out more persistent and assists users in understanding the nature of their opt-out choices. In addition, NebuAd's ISP partners are also required to provide robust direct notice in advance of the launch of the service. According to NebuAd, no user-specific data is exchanged between their database and the ISPs they work with. Inadvertent disclosure is prevented through one-way encryption of IP address and other anonymous user identifiers used by NebuAd within its system.

NebuAd has been very controversial in the United States, and has led to Congressional investigations and legal threats. In May 2009 the firm announced that it is winding down its operations.[35]

Other protections and preference management tools

Having conceived the limited application of cookies as privacy enhancing tools and the conflict between using them for both behavioural targeting and opting-out of targeting, some companies are exploring placing more user control within browsers themselves.

Internet Explorer 8, for example, offers browser-based *InPrivate* filtering (Fig. 15), which provides users with an added level of control and choice about the information that third party websites can potentially use to track browsing activity. By analyzing third parties who are in position to aggregate user data and create user profiles this new browsing mode promises to offer enhanced user control over data sharing. It allows you to browse the web without leaving a trail in the browser's history or cookie cash, and allows users to filter out third-party content on a given website. This means when a user visits a site that has ads served by third parties such as Google, AOL, Yahoo! or even Microsoft, the ads and the cookies will be blocked.

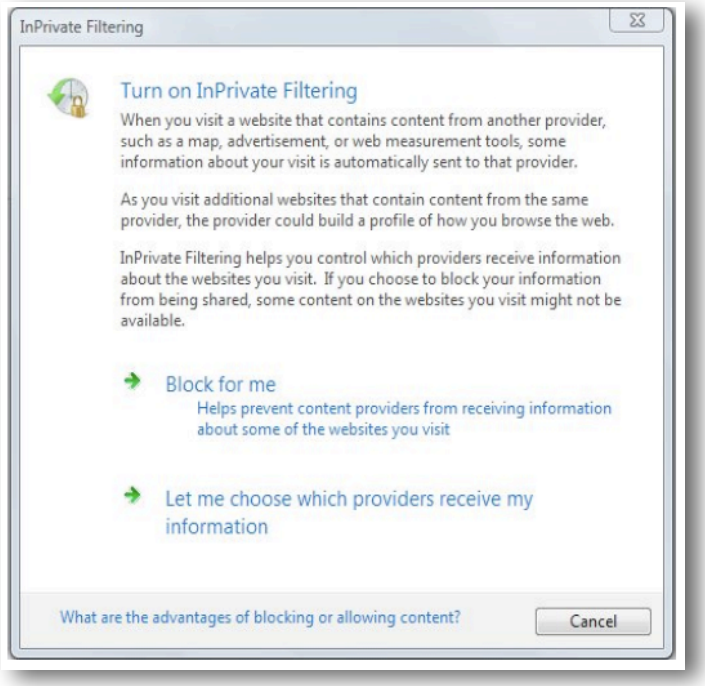

Figure 15. Internet Explorer 8™ InPrivate Filtering options.[36]

InPrivate Browsing in Internet Explorer 8 is another feature that helps prevent user browsing history, temporary Internet files, form data, cookies, and usernames and passwords from being retained by the browser.

Apple's Safari version 3.1.2. also offers a far simpler mode of 'Private Browsing' (Fig. 16). The only drawback is that once a user clicks 'OK', Safari does not open a new window and there are no visual cues to indicate that browsing is 'private'.

---

[35] NebuAd closing doors after Internet privacy woes, Deborah Yao, Associated Press, May 21, 2009.

[36] Courtesy of the Future of Privacy Forum.

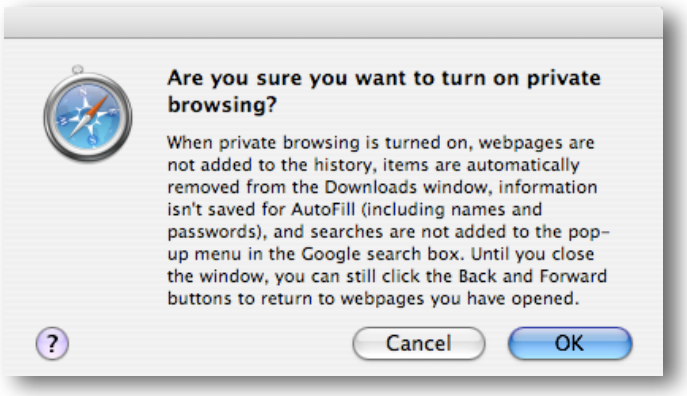

Figure 16. Safari's Private Browsing pop up notification window.

Firefox's Private Browsing (to be added to the upcoming 3.2 release) and Google Chrome's Incognito work in a similar fashion to Safari. Despite the fact that sites you visit during your Internet sessions are not able to access your cookies, web history or browser data, they will still be able to track users' visits and run third-party content. Thus the question arises of whether browser-based privacy enhancing solutions do in fact increase security and privacy or simply allow for a modicum of increased secrecy.

Other more robust advertising-specific privacy enhancing solutions could certainly be conceived if the companies behind the different browsers were willing to come together and come up with a common solution that is interoperable and serves as a default setting.

Until then, browser ad-ons, such as the Targeted Advertising Cookie Opt-Out[37], TACO, are available free of charge to those wishing to enhance existing cookie-based opt-out solutions. Unlike Google's opt-out ad-on, which allows users to opt-out of being targeted on Google's own network, TACO spans across 27 advertising networks, including Google's.

What TACO seems to offer is a temporary fix but it also serves as proof that there is demand for more tools that offer persistent opt-out, executed in a single step, and spanning across all networks, behavioural ad serving companies, publishers, and everyone else in the technology world that might be interested in tracking users online. And this is very important, not only because to achieve such coherence across the advertising market would require tremendous effort and unprecedented cooperation between competitors but also because it would require users to speak up as to which tools make sense to them.

Conclusions and Key Questions

This report has shown that the state of privacy and online advertising is far more sophisticated than is often assumed. This is not necessarily a battle between privacy and online advertising. The truth is that some companies have worked quite hard to consider privacy issues in the development of their advertising platforms.

[37] The ad-on can be downloaded here: https://addons.mozilla.org/en-US/firefox/addon/11073.

The problems emerge a bit further along in the process, however. Some firms believe that they have adequately addressed data privacy concerns. Others believe that they have resolved issues about user confidence through the creation of extensive user education campaigns. But so much more is required because the stakes are so high.

To encourage more meaningful and enforceable self-regulation, the U.S. Federal Trade Commission had originally proposed a set of Behavioural Advertising Privacy Principles[38] back in 2007. Concerned about the industry's inability to self-regulate its practices, it has now re-issued the Principles with minor modifications, giving online advertisers one last chance to improve their practices before the federal regulator proceeds with the introduction of privacy legislation.

Most noteworthy are the FTC's requirements for prominent ad-tracking disclosure mechanisms, data collection limitation, and the necessity to obtain informed consent before the collection and processing of sensitive data. Unsurprisingly, some companies have already met a number of these requirements and a noticeable effort is made by others to catch up with the trendsetters. The FTC does however offer first party advertising and contextual advertising an exemption from the requirement to comply with the Principles. This proves yet again that the most significant challenge associated with online advertising stems from the complexity of partnerships and the ever-evolving business models, which call for regular practices review.

Is Privacy-by-design necessary or sufficient?

In order to make progress in the offering of truly empowering data sharing controls, we need to start relying more heavily on privacy-by-design as the default. Because users tend to rely on default settings, by adopting a more conservative approach to privacy-friendly system and service design we would have gone halfway in assisting them to make the right choices.

There have been some remarkable advances in this domain, but we believe that a more thorough and comprehensive approach to privacy principles in the earliest design phases through to the deployment and implementation stages is required. Furthermore, despite some important attempts to offer prominent notices and choices to users, users remain unaware of the multitude of online advertising targeting techniques they are exposed to. Most advertising networks now offer cookie-based opt-out but it is often difficult to find out which networks are serving ads on certain websites, which companies are doing analytics, and which ones are simply aggregating anonymous data. The most significant developments here include the limitations on third-party scripts that are provided in some browsers, and comprehensive opt-out add-ons. Still, these require exceptional steps by users.

Even though the right tools for users to exercise relative control over the processing of their data are now in existence, users need to be educated about where to find them and how to use them. Above all, users need to be adequately informed of all advertising company practices currently taking place in the background. The argument that 'online advertising is complex' is no longer a sufficient explanation as to why companies are withholding valuable information on who is involved in personal information processing.

---

[38] http://www.ftc.gov/opa/2007/12/principles.shtm.

What about notice and consent?

Another important question that needs further consideration is: what qualifies as adequate notice? How can an individual be made to understand the issues involved, and how can he or she be required to take some personal responsibility over these issues when they are still so complex? There is no simple answer here, so more trials and research is needed on the design of user interfaces, the transparency of processing possibly through reviews by independent third parties, and strict adherence to codes of practice.

The lack of transparency and notice is not necessarily down to some form of conspiracy where industry is trying to maximise information collection and minimise consumer education. The online advertising industry has long been struggling with determining the appropriate level of consent for various services. Prominent notice plays a key role in this debate but it should be treated with care as it is often seen as an act of placing the burden upon the individual. As the argument goes: as long as users are aware of company data processing practices, as long as they understand the risks, and are given the tools to accept or reject them, then the users should be the decision-makers. However, given that reading privacy policies and terms of service stretches the limits of acceptability due to their inherent complexity, and that individuals may be subjected to profiling by firms who they may never have actually knowingly encountered, new forms of informing users of data collection practices need to be explored. Prominent notice and simple policy language are a start but they need to be backed up by new models of shared responsibility, ones that evolve with the growing intricacy of modern communications technologies.

What qualifies as 'acceptable' use of personal information?

There still isn't a consensus on what constitutes an acceptable use of personal information, and what doesn't. The debate has traditionally circled around the definition of 'personally identifiable information', thus obscuring the need for adequate protection of anonymous data. Many of the companies mentioned in this report presume that IP addresses are not personally identifiable information. They also presume that unique identifiers are also not personally identifiable information. This is highly dependent on who you ask. Even the companies themselves are conflicted as to whether or not IP addresses are personal information, as we saw Google defend the IP logs of YouTube users as a defense of their users' personal information, even as Google, Microsoft, Yahoo! and all the other companies argue to regulators that IP addresses are not personal information for advertising purposes. Meanwhile, there has been some regulatory guidance, but the conclusions are not universal.

Given the increasing value and power of data analytics and data aggregation services, we must question whether this separation between *personal information* and *non-personally identifiable information* is still appropriate. Further research is thus needed in the area of secure anonymisation and privacy-by-design personal data separation practices. There have been numerous interesting and ground-

breaking research reports on dataset de-anonymisation experiments.[39] The companies in this space must learn quickly from these problems. We therefore need to study the risks of 'anonymised' personal data, such as preferences, and de-anonymisation attacks.

How do we detect and ensure against non-compliance?

Even if privacy legislation is not introduced in the US, and Europe decides to adopts a relaxed approach to regulation by simply offering guidance on a case-by-case (service-by-service) basis, we need to start pondering over the means of detecting potential non-compliance with codes of conduct and the measures to be taken against those who may not abide by best practices. Would a name and shame strategy suffice, or do we need to adjust the current model based primarily on company honour?

The European Internet Advertising Bureau recently published a list of Good Practice Principles for online behavioural targeting, however their recommendations do not offer much guidance on enforcement and compliance. They simply state that "[e]ach Member shall have in place an effective process for handling complaints and enquiries from users about behavioural advertising and the Member's compliance with these Principles" [40]. Stronger measures are likely to be required, but these must be implemented with great care.

Our report has covered some of the leading companies in the advertising industry, but we have not covered the industry in its totality. There are still companies out there that use more invasive and non-consensual techniques like web bugs, flash cookies, even pop-up advertisements. Forcing the entire industry into some framework appears inevitable, and ensuring compliance will likely require a strong regulatory hand that can reach across borders and industrial sectors.

Do we have to authenticate to protect privacy?

In order to be able to evaluate significant deviations from best practices, we would need to ensure the auditability of different services and products. One possible solution might be allowing users to access their profiles and empowering them to review and manage their data sharing preferences, similarly to BlueKai and Google.

Such a solution, however gives rise to yet another challenge, that of reliable authentication. What are the risks of having multiple users access the same account and simultaneously modify advertising category preferences? What are the privacy risks associated with audit trails of such modifications and the potential public exposure of preference lists, whether compiled knowingly or unknowingly? Should users be required to authenticate themselves before they can modify these? The creation of a weak authentication-regime will permit others to gain access to an

---

39 Researchers from the University of Texas, Austin described an intriguing methodology for the de-anonymisation of large datasets of individual preferences, recommendations, and transaction records. They used a large dataset of movie ratings, released as part of a competition for improving the recommendation service of Netflix, the world's largest online DVD rental service. Even though Netflix claimed to have stripped the data of any personally identifiable information, the researchers managed to demonstrate that knowing just a little bit about a subscriber allowed for that their record to be identified in the dataset.

40 Internet Advertising Bureau, Good Practice Principles for Online Behavioral Advertising, 2008.

individual user's profile to discover his or her interests; while a strong-authentication regime would require users to register with all the advertising companies, thus disclosing even greater amounts of personal information.

Is openness overrated?

The promotion of privacy-by-design and notification standards probably requires cooperation across industry, and possibly with regulators as well. But what about the users? As we have seen in the recent public debate around Facebook's updated terms of service, users are increasingly interested in the shaping of privacy policies. While this new participatory approach is encouraging it needs to be further supported through open debate and transparency of services.

Only full openness on behalf of all relevant actors will ensure that the right decisions are made not only now but also in the future. As technology evolves and online advertising comfortably migrates to hand-held devices, such as mobile phones and possibly even smaller devices, we will be faced with new challenges.

We will have to put in extra effort in safeguarding children's privacy and data, in building in notification mechanisms in location services, and last but not least, we need great attention to resolving issues of consent and authentication, seen yet from a different angle. In order to acquire the versatile skills and knowledge to tackle such challenges, we will need to willingly shift our priorities and put users first no matter how high the price for businesses and no matter how mind-boggling the policy making process.