

Working Party for Information Security and Privacy (WPISP)

Working Party on the Information Economy (WPIE)



Joint WPISP-WPIE Roundtable

**The Economics of Personal Data and Privacy:
30 Years after the OECD Privacy Guidelines**

OECD Conference Centre

1 December 2010 9:30 – 18:00

Background Paper #1

“The Economics Value of Online Customer Data”

by Catherine Tucker, MIT Sloan School of Management

cetucker@mit.edu

This document has been prepared as background for the Roundtable. It summarizes the methods that Internet service firms (including ad-networks, social networks and search engines) use to collect individual data online. It then discusses the economic value of such online data for marketing, product development and other operational purposes.

The views expressed in this paper are those of the author and do not necessarily represent those of the OECD or its Member governments.

Contents

1	Introduction	3
2	Types of Data Collected Online	3
2.1	Description of different online data collection techniques	3
2.1.1	IP addresses	3
2.1.2	Web-bugs	4
2.1.3	Cookies	4
2.1.4	Click-stream data	4
2.1.5	Deep packet inspection	5
2.2	Data collected by individual firms	5
2.3	Data collected by specialized marketing research companies	6
2.3.1	Data harvested from user generated content	6
2.4	Privacy and societal implications	6
3	Data Collection and Use in Online Advertising	7
3.1	Types of online advertising	8
3.1.1	Display advertising	8
3.1.2	Display advertising networks	9
3.1.3	Search advertising	10
3.1.4	Social media advertising	11
3.2	Economic value of advertising	12
3.2.1	Measuring economic value	12
3.2.2	Theoretical economic value	14
3.3	Privacy and societal implications	15
3.3.1	Potential negative sources of economic value for customers	16
3.3.2	Potential positive sources of economic value for customers	16
4	Economic Value of Online Customer Data Outside of Advertising	18
4.1	Website and communications design	18
4.2	Product design	18
4.3	Improving recommendations	19
4.4	Organizational and operational efficiency	19
4.5	Privacy and societal implications	19
5	Conclusion	20

1 Introduction

This report examines how the data generated when consumers use the internet can create value and opportunities for firms to improve their operations and market their products. It starts by describing how firms collect data and then goes on to discuss in turn how this data is used for improving advertising and firm operations. For each application of using this data, the report discusses the privacy and societal implications.

The split between advertising and non-advertising in the report reflects the extent to which much of the pioneering use of technology in this area has been focused on improving advertising. The emphasis on advertising makes sense because most content on the internet is free, so it is a first-order problem for internet firms to monetize their content through advertising. These efforts have been very successful. A recent report put the economic value for consumers of advertising-supported internet services in the US and Europe at 100 billion euros. Several web service firms, however, gone beyond advertising, and use such data to improve their product offering and increase the efficiency of their operations. Such techniques also have the potential to generate considerable economic value.

Correspondingly, much recent discussion over online privacy has debated the need for regulation surrounding the use of online customer data to target advertising. Another point of controversy has been the practices of search engines who have pioneered the use of huge banks online data in order to optimize their products.

This report focuses on the use of data that would not exist, were it not for customers' use of the internet - data explicitly generated by customers browsing the web. This means that it does not study the wide variety of scenarios where the use of online platforms has facilitated the exchange of data concerning ideas, information and products. Various innovations in data-processing that have been made possible by the internet, such as online applications for financial products and online job-search, also represent a huge source of economic value that is not assessed during this study.

2 Types of Data Collected Online

2.1 Description of different online data collection techniques

The major ways that data are collected online are simple. Generally, most information is collected using some combination of web-bugs and cookies.

2.1.1 IP addresses

An Internet Protocol address (IP address) is a numerical label assigned to any device (such as a personal computer) that uses the internet. Any time a user views a website, their IP address makes a request to that website's server for information.

Websites find IP addresses useful because they allow the website to track the movement of one computer across a website over time, by recording the series of requests for websites that particular computer makes. IP addresses are either dynamic or static. Dynamic IP addresses change every time an internet connection is reset. Static IP addresses are persistently configured and stay the same over time. Static IP addresses allow even longer-term tracking, because they allow websites more easily to recognize a returning customer.

2.1.2 Web-bugs

Web-bugs are 1x1-pixel pieces of code that allow advertisers to track customers remotely. These are also sometimes referred to as 'beacons', 'action tags', 'clear GIFs', 'Web tags', or 'pixel tags' (Gilbert, 2008). Web-bugs are different from cookies, because they are designed to be invisible to the user and also are not stored on a user's computer. With web-bugs, a customer cannot know whether they are being tracked without inspecting a webpage's underlying html code. Web-bugs allow advertisers to track customers as they move from one webpage to another. They also allow advertisers to document how far a website visitor scrolls down a page. Combined, this means they are very helpful in determining website visitor interests. Murray and Cowart (2001) found that 96 percent of websites that mentioned a top 50 brand (as determined by the 2000 FT rankings) had a web-bug.

2.1.3 Cookies

A cookie is simply a string of text stored by a user's web browser. Cookies allow firms to track customers' progress across browsing sessions. This can also be done using a user IP address, but cookies are generally more precise, especially when IP addresses are dynamic as is the case for many residential internet services. Advertisers may also use a 'flash cookie' as an alternative to a regular cookie. A flash cookie differs from a regular cookie in that it is saved as a 'Local Shared Object' on an individual's computer, making it harder for users to delete using regular tools on their browser.

Advertiser tend to use cookies and web-bugs in conjunction because of the challenge of customer deletion of cookies. 38.4 percent of survey respondents say that they delete cookies each month (Burst, 2003). Therefore, web-bugs (which a user cannot avoid) have been increasingly used in conjunction with, or even in place of, cookies in targeting advertising. Web-bugs also have greater reach in terms of tracking ability than cookies, because they can be used to track consumers' scrolling within a webpage.

2.1.4 Click-stream data

A click-stream is a series of webpage requests. 'Click-stream data' refers to the collection of data that describes the browsing habits and actions of a particular customer. Typically a customer is identified by a cookie (if available) or an IP address if not. Web-bugs are used in conjunction with these webpage-level data to determine precisely where on a

webpage a customer browsed.

2.1.5 Deep packet inspection

There are also other even more comprehensive ways of obtaining user browsing behavior. One such technique is 'deep-packet' inspection. This occurs when an internet service provider inspects for content the data packets that are sent between one of its clients and websites. This technique was used by Phorm, an advertising agency in the UK, in partnership with internet service providers, to target ads. Since deep-packet inspection is typically done at the IP level, it represents a universal history of browser behavior. This is different from collection of click-stream data where users are tracked across a subset of websites only. Researchers such as Clayton (2008) argued that this is akin to 'warrantless wiretapping,' because theoretically the firm can observe the content of private communications.

2.2 Data collected by individual firms

In the past, when a customer interacted with a firm offline, the trail of information was both scattered and limited. There may have been point-of-sales records, there may have been telephone calls, and in some cases scanner data from the checkout if the firm offered a customer loyalty card. However, in general it was hard for any firm to link behavior to an individual at much more than a zipcode level, and they generally lacked information about what broader range of products that customer was interested in.

However, online the picture is very different. From the moment a customer first visits a website, the firm can collect multiple types of information:

They can obtain information about the website that directed the user to that website, and if the user used a search engine, what search terms they used to reach the website.

They can work out exactly where on an individual webpage a user is looking.

They also have records not only of the decisions that a user made (such as making an actual purchase) but also decisions that the user did not make (for example the decision to abandon a purchase).

This kind of information is at an individual level. However, if the website has agreements with other websites to share users' clickstreams, the reach of this information is potentially much broader. Two particular areas of note are:

If the firm has an agreement with a social networking site such as Facebook, they can use any information that the user chooses to make public in their settings (often their name, friends, and affiliations) to personalize that person's web experience.

More broadly, the firm can try and match its click-stream information with other websites to try and track what other websites that person visited. This is often facilitated by advertising networks.

2.3 Data collected by specialized marketing research companies

In a similar manner, the internet has also permitted marketing research companies to collect details about user behavior that allows more voluminous and accurate marketing data than ever before.

There are also various companies that specialize in more general collection of data on online behavior that they then resell. For example, comScore collects data on the websites that over 2 million worldwide panelists visit, what search terms they use at search engines and their online purchase and shopping history. comScore then repackages this information to sell reports and data services that illuminate e-commerce sales trends, web site traffic and online advertising campaigns, to Fortune 500 companies and media companies. The nature of the data collected has led to some controversy. Harvard Business School researcher Ben Edelman suggests that even if comScore requests permission to track users' names and addresses, browsing, shopping, and 'online accounts ... includ[ing] personal financial [and] health information,' some of their partners who distribute the tracking software do not always get consent (Edelman, 2007).

As one of the pioneers in this field, comScore has received a lot of press and attention. However, there are many other companies with similar business models worldwide. For example, Experian Hitwise collects data on users in Australia, New Zealand, France, Brazil, the United Kingdom, Hong Kong and Singapore. Nielsen/Netratings similarly also has wide-ranging geographical coverage. This kind of firm, contributes increasingly to the broader economy. As of September 2010 Businessweek Investor reports that, comScore has a market capitalization of \$622.5 million and 593 employees.

2.3.1 Data harvested from user generated content

There are also private firms which explicitly harvest content from user-generated content such as reviews and blogs. Firms like 'Sentiment Metrics' monitor news, social media, blogging and micro-blogging websites (such as Twitter). Firms are also able to collect this information themselves using tools such as 'boardtracker.com,' which alerts them when someone mentions a company name or product or forum. Similarly firms are able to use tools such as Yahoo! Pipes to monitor mentions of their product or firm on social networking sites, Twitter or blogs. This allows firms to quickly respond to customers' complaints or suggestions.

2.4 Privacy and societal implications

Companies have collected data on their customers since the beginning of commercial transactions. The fact that data are being collected is not new. Firms have also been able for decades to buy data from external parties such as magazine subscription and car ownership data and integrate it into their mailing lists, so the fact that firms have data reflecting customer actions beyond what occurs within the confines of the customer's relationship with the firm is not new. What

is new about the collection of online data is the scope of the data collected, the precision with which the company can associated an action with a customer and the sheer quantity of information. In some sense, the data collection can be thought of as like being able to video someone's consumption of media and engagement with products and then accurately encode various commercially relevant actions in a way which is easily scalable.

3 Data Collection and Use in Online Advertising

Online advertising is the canonical example of how firms can use online data to improve operations. Online advertising is characterized by a need for detailed data collection online to support it. Key to this data collection is the notion of targeting, that is, that data are collected in order to determine which kind of customers would be most likely to be influenced by a particular ad.

Ad targeting occurs when an advertiser selects a particular subset of potential viewers of the ad to show the ad to, and displays the ad online to that subset rather than to everyone using the media platform. An example would be choosing to advertise, not to the hundreds of millions of Facebook users in general, but only to those Facebook users who are female and aged between 26 and 54 and list on their profile that they like the poet Maya Angelou. No newspaper can offer this level of targeting. The targetability of online advertising can be thought of as reducing the search costs for advertisers of identifying consumers. Targeting ads has always been known to be desirable, but internet advertising has two primary advantages over offline advertising. First, the internet has made it virtually costless for advertisers to collect huge amounts of customer data. In contrast, the costs of collecting detailed enough individual-level data to target offline, for example through 'direct response mail', has generally been prohibitive. Second, internet technology makes it relatively easy to serve different customers different ads because packets are sent to individual computers. In contrast, with current technology, splitting cable TV ads across consumer types or sending hundreds of different newspaper ads to different households is prohibitively expensive.

Between the data collection and the serving of ads, targeting requires sophisticated algorithms and data processing capabilities to serve the right ads to the right people. There is nothing inherent in internet technology that makes online advertising better at this computational step, but the data collection and individual-level sending of ads make this computational step particularly fruitful.

These different targeting methods generally, of course, require media platforms to collect comprehensive data on the webpages that customers have previously browsed. Typically, advertisers and website owners track and identify users using a combination of cookies, flash cookies and web-bugs. By examining past surfing and click behavior, firms can learn about current needs as well as general preferences.

3.1 Types of online advertising

There are many different types of online advertising. Table 1 summarizes how spending was distributed across the different types in the United States in 2009. Such advertising can be divided into three broad categories: Search advertising, display advertising and social media advertising.

Table 1: How \$22.7 billion Spent in the US on Online Advertising in 2009 was Distributed

Type of Ad	Percentage Spent
Search	47
Display Banner Ads	22
Classifieds	10
Rich Media	7
Lead Generation	6
Digital Video	4
Sponsorship	2
E-mail	1

Source: IAB (2010)

3.1.1 Display advertising

“Display advertising” includes display banner ads, media-rich ads and digital video ads. Display advertising is the major mechanism by which web-pages that provide non-search content finance their websites.

In the early years of the World Wide Web, banner ads were the predominant advertising medium. However, click-through rates for banner ads have fallen sharply since the early 2000s, perhaps because they are less novel to the consumer, and because there are many more of them. In response to falling effectiveness (as measured by, for example, click-through rates), display advertising has evolved substantially beyond the ‘electronic billboard’. Display advertising (which encompasses plain banner ads as well as new rich media and video ads) is now a multi-billion-dollar market where ads include many sophisticated visual and auditory features that make ads more obtrusive and harder to ignore. Another separate but noteworthy development has been Google’s development of a highly profitable non-search display advertising division (called “AdSense”) that generates an estimated \$6 billion in revenue by displaying plain content-targeted text ads (Goldfarb and Tucker, 2010b).

Many different types of online targeting are used today for display advertising. Table 2 summarizes the broad categories of targeting used by advertisers in 2010. There are of course many hybrids of these categories. The targeting techniques used vary across different advertising formats. The targeting of display advertising is particularly important on websites where the content is not easily monetizable or matched to consumers (such as news, web services and entertainment). Websites in these categories generally use various forms of targeting based on user behavior rather than content to try to match ads with consumers. Other websites, for example theknot.com, whose services are focused

on a very specific topic (in this case weddings), attract advertisers who rely primarily on the fact that the user base of theknot.com will be interested in buying wedding-related products and services.

Behavioral targeting in its most basic form uses data from a user's clickstream to try to work out whether the user is interested in a particular service. For example, if Yahoo! observes a person looking at Infiniti cars on Yahoo! Autos, they can use this information to then serve them ads for Infiniti cars when they look at news stories on Yahoo! News. This kind of targeting can either be conducted by a single content provider, or by an ad network that provides ads to many different websites. An ad network can use the information that someone is looking at information on a certain Caribbean vacation destination on one website, to serve ads for resorts in that destination when a user starts reading celebrity gossip on a totally separate website.

Another increasingly important form of targeting is 'retargeting'. Retargeting is when a website tracks whether or not a customer has expressed an interest in a particular product or service. If the customer fails to purchase the product, then the advertiser shows new ads to those specific customers in an effort to get them to return to the website. This kind of targeting is also possible in search advertising, where search ads can be replayed after the initial search and the user is searching using different search terms.

An important development in ad targeting has been the advent of 'real-time targeting'. The distinction between this and the two methods discussed above is that real-time targeting puts the anonymous data about users into advertisers' hands directly, rather than the media platform deciding whom to serve an ad to at any one time. The label 'real-time' emphasizes the fact that this type of targeting is implemented by real-time auctions based on recent movements by a certain user.

3.1.2 Display advertising networks

Collecting targeting data is reasonably burdensome for an individual website or advertiser, so online advertising networks have sprung up to collect such information. These platforms or intermediaries connect advertisers to websites that want to host online ads.

Advertising networks are intermediaries, connecting website publishers with advertisers seeking to reach an online audience. They are often referred to as 'third-party ad networks' because they purchase advertising space from websites and resell it to the ultimate advertisers. Advertising networks are significant within the online economy. The top nine firms averaged 159 million unique visitors in December 2009 with advertising revenues of over \$3 billion that year, accounting for approximately 40 percent of total internet display advertising revenue (Beales, 2010). As well as collecting targeting data, such ad networks also perform the useful optimizing task of aggregating advertising space and allocating it among advertisers and websites, potentially reducing transaction costs within their economies. Such ad-network services are particularly important for small websites that cannot afford a large advertising sales force or

Table 2: Different Types of Online Targeting

Name	Description
Contextual Targeting	Ad is matched to content it is displayed alongside
Behavioral Targeting	Use prior click-stream data of customer to determine whether they are a good match for the ad. Scope generally depends on whether ad network or website publisher controls which ads get displayed
Retargeting	Online ad is shown to user who previously visited a website but did not 'convert'
Real-Time Targeting	Advertiser has power to decide in 'real time' whether to serve an ad to a user based on data the website shares with them about that user
'Look-alike' Targeting	Targeting based on users having similar characteristics to current customers
'Act-alike' Targeting	Targeting based on users having click-through paths which resemble successful conversions
Demographic Targeting	Publisher uses data that customer has volunteered, such as age, gender, location and interests, to choose whom to display ads to

the search costs associated with finding potential advertisers. Beales (2010) found that, in 2009, 55 percent of revenues in ad-networks were shared with websites, suggesting that websites benefit too.

Advertising networks have an advantage over individual websites in collecting and using clickstream data for targeting purposes. First, because of their scope, they can invest in large scale data collection, processing and optimization techniques. Second, because they contract with a wide variety of websites, they are able to usefully construct an anonymous user information profile and identify users who have been visiting, for example, travel websites in order to identify potential target customers for hotel and airline ads on non-travel websites.

3.1.3 Search advertising

Search engine advertising is enormously important because of the limited extent to which consumers search on the internet from website to website. Search engines are a key gateway to all websites on the internet, and as such provide a crucial advertising venue. 'Search ads' or 'paid search' are the ads that appear alongside search results after a consumer types a search term or 'keyword' into a search engine. Typically, these ads are placed above or to the right hand side of the 'organic' or main search results.

Generally, search engines rely on context-based advertising where ads are displayed based on the search term typed into the search engine. This means that if someone searches for information on Aviation Accident Attorneys, they see only ads that were bid on by advertisers who are advertising some form of aviation accident legal service. Search

engine ads are therefore placed in the context of the stated intent of the user. Search is a statement of intent, so ads that are targeted and priced to these statements of intent are highly effective. A search engine ad is a direct response to a statement of intent and therefore targets potential customers at exactly the time they are looking for something. Search engines also sometimes engage in retargeting. This is where an online ad is shown to a user who previously searched using a particular search term. To identify such users, search engines typically collect logs of search queries.

Usually search engines collect data for an individual user-profile using either a cookie or an IP address. Associated with this profile are the search queries and subsequent clicks made by each user. The length of time that data is retained is controversial. The EU parliament's privacy working party has requested that search engines retain data for only six months. Currently Google anonymizes IP addresses on its server logs after nine months, but keeps queries associated with a cookie for 18 months. Microsoft has stated that it deletes them after six months at the EU's request. This may change, however. In June 2010, the proposed 'European Data Retention Directive' would request search engines to keep data for 2 years in order to identify pedophiles and other illegal activity better.

Search engines also typically distinguish between the data that they collect that are personally identifiable, for example that are associated with a Google account, and search logs. For example, Google emphasizes that 'Web History databases', associated with a user account, are maintained separately from the search log data. They also allow users to control what goes into their web history through their Google Account settings.

3.1.4 Social media advertising

Marketers have always recognized that word-of-mouth and customers' social relationships were important for transmitting positive information about their product. However, before online media evolved, it was very expensive and difficult to measure word-of-mouth or for firms to pro-actively encourage it.

Online social media advertising allows firms explicitly to direct word-of-mouth. New advertising agencies, such as Bzz Agent and Tremor, now exist to encourage positive online word-of-mouth. These practices are somewhat controversial, as it is not clear whether consumers realize that the information being shared is not necessarily authentic. To address such possible behavior, October 2009 the Federal Trade Commission introduced Guidelines Concerning the Use of Endorsements and Testimonials in Advertising. These were directed specifically at clarifying what constitutes an endorsement when the message is conveyed by bloggers or other 'word-of-mouth' marketers.

A specific area of online social media marketing that is growing in importance is advertising on social networks such as Facebook. Table 3 summarizes the average time spent each month by users on such sites. What is striking is just how long users are spending on these sites relative to other kinds of websites. For example, in the US people spend on average over six hours a month on these sites, which is more than double the time spent on other popular types of websites such as portals or search engines.

Table 3: Global Reach of Social Networking Sites

Country	Unique Audience (000)	Time each Month per Person (hh:mm)
United States	142,052	6:09
Japan	46,558	2:50
Brazil	31,345	4:33
United Kingdom	29,129	6:07
Germany	28,057	4:11
France	26,786	4:04
Spain	19,456	5:30
Italy	18,256	6:00
Australia	9,895	6:52
Switzerland	2,451	3:54

Source: The Nielsen Company, January 2010

In addition to the typical advertising data used for display advertising, social networks also collect and provide additional information about their users to advertisers. For example, Facebook allows targeting based on whether or not someone's friends have expressed a liking for a product. It also allows targeting on its users' profile demographics, such as birthdates, educational and occupational background, and interests. Though this information has generally been provided anonymously, there has been recent controversy because Facebook appeared to be directly giving advertisers the profile ID of each of its users, rendering the data non-anonymous (Krishnamurthy and Wills, 2009). This was corrected by Facebook in May 2010, but potentially other means of de-anonymizing such data may be found.

3.2 Economic value of advertising

3.2.1 Measuring economic value

Since the first banner ad was shown in October 1994, online advertising has grown quickly. By 2009, online ads accounted for \$22 billion in spending (IAB, 2010). Online advertising is also important for what it enables. In the United States alone, websites supported by advertising represent 2.1% of the total U.S. gross domestic product (GDP) and directly employ more than 1.2 million people (Deighton and Quelch, 2009).

Figure 1 compares the growth of internet advertising for its first fifteen years (1995-2009) relative to cable television (1980-94) and broadcast television (1949-1963), in current inflation-adjusted dollars. It is clear that online advertising has grown much faster, especially in the last 5 years, than traditional advertising channels did in their first fifteen years.

Figure 1: Ad Revenue for the first 15 years of online media, cable TV, and broadcast TV



Source: IAB (2010)

Figure 2: Many European countries spend proportionally more on online advertising than the US (13 percent)

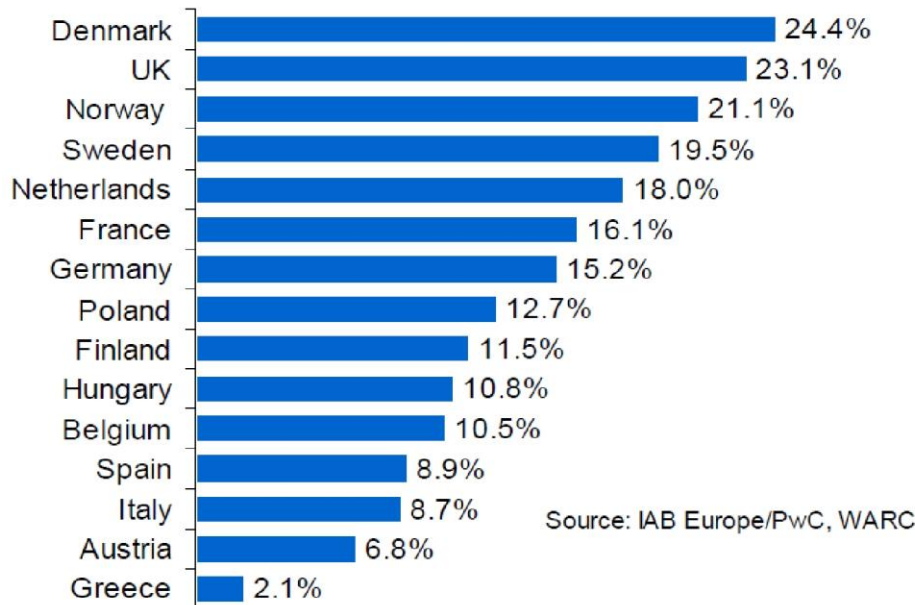


Figure 2 compares the proportion of advertising money spent in different European countries on online advertising relative to traditional offline formats. In most large European countries, the proportion of advertising spending online is actually higher than in the US.

The economic value generated by such targeted advertising activity does only reflect advertising revenues for web-platforms. There are also benefits for consumers. (McKinsey, 2010) use conjoint techniques to estimate that in the US and Europe consumers received 100 billion euros in value in 2010 from advertising-supported web services. This is three times greater than current revenue from advertising, suggesting that the consumer value created is larger than advertising revenues would indicate.

3.2.2 Theoretical economic value

The obvious question is why advertisers have seen so much economic value in marketing themselves online.

The answer lies in two features that are unique to the kind of data that can be collected online: Measurability and Targetability. Measurability is higher because the digital nature of online advertising means that who sees what ad and whether they respond can be tracked relatively easily. Targetability is higher because firms know about consumers' browsing habits at the individual level and can then choose whether to serve them an ad based on that profile. These two features mean that online advertising can overcome the legendary critique of offline advertising:

'I know half my advertising is wasted, I just don't know which half' - John Wanamaker, department store

innovator, 1838-1922.

Measurability

The measurability of online advertising creates economic value because offline it is hard to observe the link between a consumer seeing an ad and the same consumer subsequently buying the product. It appears to work (i.e., people who see the ads might be more likely to buy than people who do not), but the firm cannot see how. The firm does not know whether a consumer was motivated to buy because of a particular newspaper ad, or because of a TV ad, a specific billboard, or their new radio jingle. For long-term advertising campaigns that try to build affection over time for a particular brand, this problem is especially acute. Macy's can observe who uses their 20% off coupon in the Sunday paper, but Budweiser cannot observe whether their Bud Light ad shown during the Superbowl is linked to higher sales in the long run. Further, even if firms can observe a clear link between someone seeing an ad and then buying the product, it is not clear that there is a causal link between the two.

By contrast, online advertising is inherently measurable. The digital nature of online advertising means that individual responses to ads can be easily recorded. For example, the effectiveness of many forms of online advertising can be measured by whether or not someone clicks on an ad. Often, through the use of cookies, IP addresses, and other tracking technologies, advertisers can go beyond this simple click metric and observe directly whether users engage in a certain online action (such as an online purchase, or subscribing to receive more information) after being exposed to an ad.

Targetability

Targetability increases the value of advertising to firms because they no longer have to pay for wasted 'eyeballs'. In other words, firms do not have to pay money to serve ads to people who are unlikely to buy a car or a vacation or suffer from that particular health complaint. Instead, firms can be reassured they are allocating their money to serve ads to customers who are potentially likely to buy their product.

Behavioral targeting has empirically been shown to increase the economic value that firms place on advertising. Beales (2010) finds that in 2009 the price of behaviorally targeted advertising was 2.68 times the price of untargeted advertising.

3.3 Privacy and societal implications

This section has described the benefits of detailed user-level data for advertisers and websites. However, there may be costs and benefits to customers and the broader society that are not captured sufficiently in the simple figures that reflect the amount of money spent on online advertising.

3.3.1 Potential negative sources of economic value for customers

Collecting advertising data online is often argued to be harmless, because it typically involves a series of actions linked by an IP address or otherwise anonymous cookie-ID number. However, attempts by advertisers to use such information has met with resistance from consumers due to privacy concerns. In a well-publicized survey, Turow et al. (2009) found that 66 percent of Americans do not want marketers to tailor advertisements to their interests. This customer resistance to tailored advertising is a major problem for advertisers. Fear that users may react unfavorably because of privacy concerns has led advertisers to limit their tailoring of ads. A recent survey suggested that concerns about consumer response have led advertisers to reduce the targeting of advertising based on online behavior by 75 percent (Lohr, 2010).

There also may be costs to consumers in the form of behavioral price discrimination. Behavioral price discrimination means that firms use past consumer actions to distinguish between customers who have low and high willingness to pay for their product and offer them low and high prices as a consequences. One example may be that firms may offer ads that offer discounted coupons to consumers whom they observed browsing their products but not purchasing the product, in order to provide a final incentive for that customer to buy the product. Therefore consumers could be offered very different effective prices based on their click-stream data without their knowledge. This may be harmful, especially if it distorts consumer decisions - that is, consumers might strategically waste time exhibiting behavior (such as browsing a website and not purchasing a product) in order to attract a discounted ad.

Third, at the moment it is not clear that property rights over such click-stream data used for targeting are assigned in a way which ensure equitable sharing of the economic value created by such targeting between firms and customers. Instead, firms collect information on customers' click-streams, often without informing consumers in an upfront way that they are doing so, and use that to conduct profitable advertising campaigns. Consumers are not given a chance to negotiate with advertising networks to share the profits being made from their personal data.

3.3.2 Potential positive sources of economic value for customers

There is obvious economic value created by online advertising for firms in terms of advertising revenues. However, there are also sources of economic value for customers. First, targeted internet advertising may serve a useful informational role. Instead of being forced to view untargeted mass-media ads on TV, consumers see ads that are related to their potentially unique interests and desires. For example, people who are not in the market for booking a vacation are less likely to have to see ads for travel companies, but instead may see ads related to their actual hobbies. Supporting this contention that there is utility of targeted ads for consumers, is the observation that conversion rates for behaviorally targeted ads are more than twice the rate of non-targeted ads (Beales, 2010).

Second, there is evidence that growing evidence that targeted ads are optimally less obtrusive in design than non-targeted ads (Goldfarb and Tucker, 2010a). Obtrusive ads are ads that are deliberately designed to intrude on users' web browsing experience such as 'take-over ads' or 'floating-ads'. Since consumers have expressed dislike of these obtrusive ads, there may be benefits to targeted ads if users do not have to experience ads that are also deliberately obtrusive.

Last, the benefit of profitable targeted advertising is that it funds and enables a wide variety of free-web content and services. It is possible that these services dwarf the negative impact of privacy concerns. McKinsey (2010) use conjoint analysis to suggest that for each euro an Internet user is willing to spend to limit privacy and advertising disturbance, the user gets a value of six euros from using current ad-funded Web application services.

4 Economic Value of Online Customer Data Outside of Advertising

The use of customer online data is both deeper and more widespread in advertising than elsewhere. Most content and online services are advertising-supported, so there has been a first-order need to use such data to optimize advertising and maximize revenues. However, there is a growing group of web services firms that are pushing the boundaries of using such data to meet both broader operational and marketing needs.

4.1 Website and communications design

There are benefits to firms that offer services online from the retention and use of customer click-stream data. The data can be used to tailor products automatically to a consumers' needs and interests. Data can also be used for immediate feedback if some aspect of the online service appears to not be being used by consumers in the manner intended, to immediately start the process of fixing and improving the product. For example, Google retains user clickstream data in order to continuously improve both its search algorithms and online product services such as youtube.com, based on terminated user queries and actions.

Website content can also be optimized on the basis of online data. A company's website can automatically gear the content to the geographical region of the user suggested by their IP address. Alternatively, a website could potentially serve news that may be relevant to someone with a particular browsing history. For example, if an individual has browsed many travel websites, a news sites could highlight travel news that may be of interest.

There have also been experiments at changing the way that web content or the interface of websites is provided to match consumers' cognitive styles. A recent study launched by BT showed that such morphing of website-content could increase sales by 20 percent (Hauser et al., 2009). Here the website interface morphed to match whether the user responded more to graphical or textual information and the level of depth of explanation that they appeared to prefer. Firms are also able to use customer data to reduce flaws in website design. For example, Expedia, an online travel agency, was able to save \$12 million by using clickstream data to identify a flaw in the website design that was leading customers to insert the wrong billing address (Heath, 2010).

It is also possible to use customer online data to improve a firm's direct communications with its customers. For example, Publishers Clearing House managed to increase click-throughs by 40% by incorporating customer online activity (Miller, 2010).

4.2 Product design

One of the unique features of the internet is the extent to which users are able to generate their own content. This content is often useful to firms both for marketing and operational purposes. For example, firms are able to collect data from online reviews to optimize their current product design process. As described by Zhang et al. (2010), this

online information can identify the ‘voice of the customer’ and help firms to understand what consumers like and dislike about their product. The authors show that it is easy for camera manufacturers to use online reviews to spot ways that a camera tends to malfunction.

4.3 Improving recommendations

Another place where the collection of online data has proved exceptionally useful is the development of recommender systems. These recommender systems use similar customers’ purchase decisions to offer recommendations about products of interest to another customer. For example, if a website observes a customer buying a DVD of the TV series *Lost*, they use the purchase histories of other customers who have also bought *Lost* to suggest other DVDs that the customer might also enjoy. Early evidence suggested that such recommendation systems could yield large returns. For example, a home and garden retailer claimed an increase in average order size of 16.5% when they used earlier purchase and browsing behavior to recommend products to clients (Peters, 2001). Recent evidence differs, suggesting that such systems can increase revenues by 0.3 percent (Dias et al., 2008). This is still large given the relatively low cost of implementing such systems and the high costs of increasing revenues through alternative marketing actions.

4.4 Organizational and operational efficiency

This report has focused largely on how the sharing of online information has been used by firms to improve the accuracy of their efforts to increase demand and improve customer satisfaction. However, the sharing of online information can also be used to reduce firms’ operational costs. In particular, information on consumer trends enables firms to manage their supply chains more effectively and to communicate more easily both within the firm and with customers and partners. One example is exploiting data on online wishlists, online grocery lists and registries to project future demand for certain products.

Search data is also useful to firms for predicting demand. For example, Choi and Varian (2009) show that data about who is searching for what on search engines can actually reasonably accurately predict travel and retail demand. They show that data on who searched for tourist information about Hong Kong, predicted the official figures from Hong Kong Tourist Board on visitor numbers. What is crucial, however, from a firm perspective, is that this information is instantaneously available, so a restaurant or hotel can anticipate rises and falls in demand in advance rather than waiting for official visitor figures, which typically take weeks to come out.

4.5 Privacy and societal implications

In general, the uses of customer data described in this section are designed to reduce firm’s costs and the attractiveness of their product offering by reducing information asymmetries. As such, they represent a real gain to society. However, in some sense the storage of this data represents a larger potential privacy risk to individuals than advertising data. Data

like this tends both to be stored for longer than data used for advertising purposes and to be more easily tied back to an individual. We discuss these risks in turn.

First, most data stored for online advertising is attached to an anonymous profile attached to a particular IP address. It is far harder for an external party to tie such data back to a specific individual user than the kind of data used for product personalization discussed in this section, which has the explicit purpose of linking online data to a real person and their actions.

Second, the majority of online advertising data is stored for a short time. Indeed, the IAB suggested in 2010 that such data collection could be limited to a 48-hour window (http://www.theregister.co.uk/2010/10/04/iab_cookie_advice/) Though this met with some controversy, it is indicative of the extent to which data for advertising is short-lived. Purchase decisions occur relatively quickly, so prior browsing behavior quickly becomes irrelevant to predicting whether a customer will buy. However, customer data for operational purposes tends to be stored for longer. For example, currently there is a large debate over whether search engines should be able to store user search profiles for 18 months in the EU as they currently do (Kirk, 2010). One of the risks of longer storage time frames is that it enables a fuller and more widespread profile of a users' habits to emerge, which could more adversely affect a consumer if used for surveillance or malicious purposes.

5 Conclusion

This report describes the different kinds of information and data collected in the online services sector. Every time a consumer browses a website the content they select, the searches they make, the links they click on and the links they do not click on represent valuable data for firms. The use of such information has been particularly transformative for the online advertising sector, and the report describes in detail both the ways that information is used and the privacy concerns this engenders. This focus on advertising reflects both much of the privacy debate about the collection of online customer data and the fact that so many internet business models rely on advertising to provide free content and services. However, there are further ways the automatic collection of such data can be used to improve a firms' operations and product development process. This has been pioneered by the online services sector, which has managed to use the data to improve product offerings and optimize the user experience on their websites.

There are of course broader social and privacy concerns engendered by the collection and use of this data. The first set of issues arise because currently it is not clear that consumers' are well-informed about what data is being collected about them and mechanisms to allow them to control this process are poorly-developed. The second set of issues arise in circumstances where consumers' offline identities are linked explicitly to the data that is collected about them online. With such personally identifiable information, there are potentially negative implications for the

consumer if the data is breached or shared with external parties that the consumer did not authorize.

References

- Beales, H. (2010). The value of behavioral targeting. Mimeo, George Washington University.
- Burst (2003, June). Burst cookie survey: Consumers don't understand, say maybe useful, but some delete anyhow. Technical report, Marketing Vox.
- Choi, H. and H. Varian (2009). Predicting the present with Google trends. Technical report.
- Clayton, D. R. (2008, 15 April). Problems with Phorm. Technical report, University of Cambridge.
- Deighton, J. and J. Quelch (2009, June). Economic value of the advertising-supported internet ecosystem. IAB Report.
- Dias, M. B., D. Locher, M. Li, W. El-Deredy, and P. J. Lisboa (2008). The value of personalised recommender systems to e-business: a case study. In RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems, New York, NY, USA, pp. 291–294. ACM.
- Edelman, B. (2007, June 20). Comscore doesn't always get consent. Benedelman.org.
- Gilbert, F. (2008, May). Beacons, Bugs, and Pixel Tags: Do You Comply with the FTC Behavioral Marketing Principles and Foreign Law Requirements? Journal of Internet Law.
- Goldfarb, A. and C. Tucker (2010a). Online display advertising: Targeting and obtrusiveness. Forthcoming, Marketing Science.
- Goldfarb, A. and C. Tucker (2010b). Search engine advertising: Pricing ads to context. forthcoming, Management Science.
- Hauser, J. R., G. L. Urban, G. Liberali, and M. Braun (2009). Website morphing. Marketing Science 28(2), 202–223.
- Heath, N. (2010, 1 Nov). Expedia: on how one extra data field can cost \$12m. Silicon.com.
- IAB (2010, April). IAB internet advertising revenue report: 2009 full-year results. Technical report, IAB and PriceWaterhouseCoopers.
- Kirk, J. (2010, May 27). Europe warns search companies over data retention. PC World Business Center.
- Krishnamurthy, B. and C. E. Wills (2009, August). On the leakage of personally identifiable information via online social networks. In WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks, New York, NY, USA, pp. 7–12. ACM.

Lohr, S. (2010, April 30). Privacy concerns limit online ads, study says. *New York Times*.

McKinsey (2010, September). Consumers driving the digital uptake. Technical report, McKinsey and Company and IAB Europe.

Miller, S. (2010, February 9). 50% click-through rates? Yes, you can! Technical report, Marketing Profs.

Murray, B. H. and J. J. Cowart (2001). WEBBUGS: A study of the presence and growth rate of web bugs on the internet. Cyveillance, Inc.

Peters, K. (2001, June 27). Shoppers will buy what a retailer recommends, Smith & Hawken learns. *Internet Retailer*.

Turow, J., J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy (2009). Americans Reject Tailored Advertising and Three Activities that Enable It. Mimeo, Berkeley.

Zhang, K., R. Narayanan, and A. Choudhary (2010). Voice of the customers: Mining online customer reviews for product feature-based ranking. In *WOSN'10: Proceedings of the 3rd conference on Online social networks*, Berkeley, CA, USA, pp. 11–11. USENIX Association.

About the Author

Catherine Tucker is the Douglas Drane Career Development Professor in IT and Management and Assistant Professor of Marketing at MIT Sloan School of Management. She is interested in understanding how networks, privacy concerns and regulation affect marketing outcomes. She received an undergraduate degree in Politics, Philosophy and Economics from Oxford University and a Ph.D. in Economics from Stanford University.