**Data**

Getting to know you

**Everything people do online is avidly followed by advertisers and third-party trackers**

Sep 13th 2014 | [From the print edition](http://www.economist.com/printedition/2014-09-13)

* 

Close

Save this article

Click this to add articles to your Timekeeper reading list. Learn more »





IN “DIVERGENT”, A book series and Hollywood film, humans in a post-apocalypse Chicago are split into five different groups according to their aptitudes and values. All 16-year-olds take a test to be categorised for life. The world of online advertising is not quite as rigid as that, but gathering information about users and grouping them into sellable “segments” has become big business. Data are crucial to the $120 billion online advertising economy.

“This is an information war,” says Omar Tawakol, the boss of BlueKai, a data broker, which tracks users online and sells that intelligence to companies. “This is 100% about having more information about the customer and being able to generate more commerce as a result of it.” The internet has made it much easier to gather data about users because they leave traces wherever they go. Facebook and Twitter accumulate heaps of information, including ages, friends and interests, about people who sign up for accounts and spend time on their sites. Some of it is collected without users being aware of it. For example, Facebook’s “Like” and Twitter’s “Tweet” buttons on other websites carry a code that enables the social-networking companies to track users’ movements even if they do not click those buttons, says Peter Stabler, an internet analyst at Wells Fargo Securities.

The advertising industry obtains its data in two main ways. “First-party” data are collected by firms with which the user has a direct relationship. Advertisers and publishers can compile them by requiring users to register online. This enables the companies to recognise consumers across multiple devices and see what they read and buy on their site.

“Third-party” data are gathered by thousands of specialist firms across the web. “We have this tremendous growth of companies that people do not talk about as household names,” says Mahi de Silva, the boss of Opera Mediaworks, a mobile-advertising company that is one of them. To gather information about users and help serve appropriate ads, sites often host a slew of third parties that observe who comes to the site and build up digital dossiers about them. BlueKai, for example, compiles around 1 billion profiles of potential customers around the world, each with an average of 50 attributes.

To identify users as they move from site to site, third parties use technologies such as cookies, web beacons, e-tags and a variety of other tools. Cookies, widely used on desktop computers, are small pieces of code that are dropped on a user’s browser. According to TRUSTe, the 100 most widely used websites are monitored by more than 1,300 firms. Some of these firms share data with other outsiders, an arrangement known as “piggybacking”.

All this allows firms to glean what sites users have visited, what they have shopped for, what postcode they live in and so on. From this the firms can infer other personal details, such as their income, the size of their home and whether it is rented or owned. Typically web users are tagged when they visit a particular website, but companies are getting cleverer about expanding their reach. RadiumOne, an advertising-technology company, puts cookies on users, normally unbeknown to them, when they click on a weblink sent by a friend.

Data-gathering on mobile devices can be even more precise. PubMatic, a firm that helps publishers sell advertising space in real time, provides some 50-70 data points about users on desktops and around 100 on mobile, including the mobile device’s precise position. Mobile users spend close to 90% of their time online in mobile applications, or “apps”, which do not support cookies, so advertisers, app developers and intermediaries use other tools, such as their device’s ID, to recognise them.

Companies stress that they do not know users’ names. But they identify them by numbers, and as they build up detailed profiles about those numbered users, there is concern that the information might be traced to individuals. This puts the companies in an awkward spot. They like to boast about their robust tracking and data offerings but do not want to spook users by appearing to know too much. Firms such as Facebook, which have people’s names and other personal information, insist that they respect users’ privacy when selling information to advertisers.

**Tag, you’re it**

Collecting and dealing with all that information requires a large cast of characters. Data brokers earn their living by helping advertisers and publishers manage their own first-party data, as well as selling them more data about users. They divide them into segments defined by location, device, marital status, income, job, shopping habits, travel plans and a host of other factors, and auction those segments off to buyers of ad space in real time. This segmentation can become highly specialised. For example, eXelate, a data broker, sells “men in trouble”, presumed to have relationship problems because they are shopping online for chocolate and flowers. Another data firm, IXI, sells a segment called “burdened by debt: small-town singles”.

Most consumers have never heard of the companies that make a full-time business of gathering data about them, but they do know some of the firms that do it as a sideline. Forbes, a publishing company, sells data about readers who visit its site. Political campaign groups rent out their lists to firms as a way to generate cash. OkCupid, a dating website, used to sell information about users’ alcohol consumption and drug use, but says it no longer does.

Credit-card companies, including Visa, MasterCard and American Express, all sell anonymised data about their cardholders to advertising companies. Bidders for advertising space can go to MasterCard to buy aggregated segments of consumers who are likely to subscribe to particular telecommunications services, for example, or stay at particular hotel chains. American Express has an edge, says someone in the data business who has worked with the company, because it actually issues the card (whereas MasterCard and Visa are in partnership with banks), enabling it to put cookies on users when they log in to check their statements and see where else they go online.

Auctions can also be data mines. Some companies plug into the exchanges where firms buy and sell advertising just to glean information about users and publishers. Brokers that buy and sell advertising, known as ad networks, collect reams of data across the web. For example, Mindshare, a media buyer, wanted to find the best place to advertise for its client Kleenex, a tissue manufacturer. It took part in search auctions to see where people were Googling for cold and flu remedies, but deliberately kept its bids low enough to lose. Then it concentrated its marketing on regions where lots of people seemed to have the sniffles.

Companies have always tried to find out as much as they could about their consumers. Direct marketers used to hunt through public records, such as birth and marriage certificates and property deeds, and catalogue companies would sell lists of their customers to competitors. But the internet has vastly expanded the scope of data collection. Sometimes users explicitly allow services to track information about them, but often they are not asked, and the information is gathered by third parties that can use it without consumers or regulators knowing how.

Firms keep trying to get a rounder picture of users’ lives. One way of doing that is trying to work out which devices belong to the same owner. Companies that require users to log in, such as Facebook, Google and Twitter, have an advantage, because they are able to recognise the same user across devices. Being logged into the same Wi-Fi also provides a clue.

Companies are also keen to connect the offline and online worlds. Facebook, for example, has joined with Datalogix, a data provider, to link purchases in both spheres. Acxiom, one of the largest data brokers with expertise in the offline world, recently paid more than $300m to buy LiveRamp, a firm that helps match offline data about customers with online information.

This is not as new as it sounds. Fifteen years ago DoubleClick, an online-advertising firm that was later snapped up by Google, bought Abacus, a firm with troves of data about people’s offline purchases, but privacy advocates kicked up such a fuss that DoubleClick abandoned the project and in 2006 sold Abacus. The fuss has died down. “The technology has improved, so it’s easier to anonymise stuff,” says Scott Knoll, the boss of Integral Ad Science, an analytics firm, who formerly worked at DoubleClick. “Companies are doing it under the radar screen.”

Data firms say they take pains to protect users’ personal information, and sometimes have trouble keeping track of them. Privacy-conscious consumers regularly delete their cookies. And advertisers point out that they do not want sensitive information. “I don’t care if you’re cheating on your taxes or on your spouse. We are not trolling for personal information,” says one digital-advertising executive. “We are trying to figure out if you are a high-value customer and are in the market for a car.”



“We can do more technologically than we’re permitted to culturally”

Sometimes advertisers do not use information they have because they do not want to look as though they are spying on customers. “We can do more technologically than we’re permitted to culturally,” says Tony Weisman of DigitasLBi, a digital-advertising firm. Some advertisers wait for a few days before targeting users who had been shopping for a particular item because they do not want to let on how much they know. “We are actively trying to figure out where the boundaries are,” says Simon Fleming-Wood, the chief marketing officer of Pandora, a digital-music company. “And in the meantime we’re being conservative.”

**Breathing down your browser**

The system of data-gathering that underpins online advertising raises several questions. One is consumer privacy. Ad companies say they will not use sensitive personal and health information for advertising purposes. But Kate Kaye, who covers the data industry for *Advertising Age* and did some research on sexually transmitted diseases for a story, found herself targeted with ads offering support to HIV sufferers days later.

Another concern is how to prevent data leakage. Many companies are wary of giving third parties access to their data in case they are laxer about security or share it with competitors. In June Reuters, a news agency, had its website attacked by the Syrian Electronic Army through a third-party advertising network called Taboola which sat on its site. Others worry about a data breach from perhaps a rogue programmer who could de-anonymise the vast amount of information firms have collected. “The question becomes, who is policing that? And are those checks and balances really there?” asks Mr Knoll of Integral Ad Science.

As more information is attached to cookies and devices, it becomes easier to identify users, says Ed Felten, a professor of computer science at Princeton University. Mr Felten and others have shown that, given enough information, anonymous data sets can be de-anonymised. One study found that it took only two data points to identify more than half the users. “The idea of personally identifiable information not being identifiable is completely laughable in computer-science circles,” says Jonathan Mayer, a Stanford University computer-science researcher.

Besides, different countries have different standards of what data count as personal information. Germany forbids any marketing to people of specific ethnic groups or political affiliations without their consent, but America does not. More broadly, in Europe an e-mail or IP address is considered personal, whereas in America it might not be. Data-gathering and digital marketing there have largely escaped the regulator’s grip, except in the finance industry.

Regulators around the world increasingly find that technology has outrun them and are trying to catch up. In Europe a new privacy directive, now being drafted and likely to come into effect in 2016, will introduce extremely strict (some say stifling) rules on data collection that will apply across the European Union. Websites already have to make it clear to users that third-party cookies are tracking them. Even in China, where individuals’ rights have not loomed large, President Xi Jinping has asked his prime minister to look into data security and privacy issues.

In America a government proposal to make it harder to track people online has fallen flat. Instead, under the digital-advertising industry’s system of self- regulation users can go online to opt out of being targeted with ads (but not of being tracked). Ads delivered by firms that have signed up to the self-regulation programme feature a small “Ad Choices” icon on which people can click to opt out, though according to Chris Babel of TRUSTe, a mere 0.00015% of those who see the icon take advantage of that option. And users who delete their cookies are automatically opted back in and keep having to repeat the process.

Some American advertising executives see more regulation as inevitable, especially in relation to third parties and data brokers. According to Jim Halpert of DLA Piper, a law firm, who co-chairs its global data-security practice, “the issue is not advertising. It is rather that some entities can sell lots of information about individuals without those individuals knowing about it.” There is very little oversight of how this information is used or where it is sold. Annual audits of third-party data collectors could help ensure that the information is used fairly, says Mr Halpert. So far concerns about unfair practices have been raised mainly by academics, tech geeks and some vigilante consumers, not the public in general, but that may be because most people do not even know that they are being followed.

[From the print edition: Special report](http://www.economist.com/printedition/2014-09-13)

fa