The largest U.S. websites are installing new and intrusive consumer-tracking technologies on the computers of people visiting their sites—in some cases, more than 100 tracking tools at a time—a Wall Street Journal investigation has found.

The tracking files represent the leading edge of a lightly regulated, emerging industry of data-gatherers who are in effect establishing a new business model for the Internet: one based on intensive surveillance of people to sell data about, and predictions of, their interests and activities, in real time.

The Journal's study shows the extent to which Web users are in effect exchanging personal data for the broad access to information and services that is a defining feature of the Internet.

In an effort to quantify the reach and sophistication of the tracking industry, the Journal examined the 50 most popular websites in the U.S. to measure the quantity and capabilities of the "cookies," "beacons" and other trackers installed on a visitor's computer by each site. Together, the 50 sites account for roughly 40% of U.S. page-views.

The 50 sites installed a total of 3,180 tracking files on a test computer used to conduct the study. Only one site, the encyclopedia Wikipedia.org, installed none. Twelve sites, including IAC/InterActive Corp.'s Dictionary.com, Comcast Corp.'s Comcast.net and Microsoft Corp.'s MSN.com, installed more than 100 tracking tools apiece in the course of the Journal's test.
It's rarely a coincidence when you see Web ads for products that match your interests. WSJ's Christina Tsuei explains how advertisers use cookies to track your online habits.

The Journal also surveyed its own site, WSJ.com, which doesn't rank among the top 50 by visitors. WSJ.com installed 60 tracking files, slightly below the 64 average for the top 50 sites.

Some two-thirds of the tracking tools installed—2,224—came from 131 companies that, for the most part, are in the business of following Internet users to create rich databases of consumer profiles that can be sold. The companies that placed the most such tools were Google Inc., Microsoft, and Quantcast Corp., all of which are in the business of targeting ads at people online.

Google,

Dig Deeper

The Web's New Gold Mine: Your Secrets
The Journal's Methodology
What They Know About You
Digits: Your Questions on Digital Privacy
Digits: Analyzing What You Have Typed
Digits: Lawsuit Tackles Files That 'Re-Spawn' Cookies
Full Coverage: wsj.com/WTK
Follow @whattheyknow on Twitter

Glossary

Key tracking terminology

How to Protect Yourself

Almost every major website you visit is tracking your online activity. Here's a step-by-step guide to fending off trackers.
Microsoft and Quantcast all said they don’t track individuals by name and offer Internet users a way to remove themselves from their tracking networks. Comcast, MSN and Dictionary.com said they disclose tracking practices in their privacy policies, and said their visitors aren’t identified by name.

The state of the art is growing increasingly intrusive, the Journal found. Some tracking files can record a person’s keystrokes online and then transmit the text to a data-gathering company that analyzes it for content, tone and clues to a person’s social connections. Other tracking files can re-spawn trackers that a person may have deleted.

To measure the sensitivity of the data gathered by tracking companies, the Journal created an "exposure index" for the top 50 sites. Dictionary.com ranked highest in exposing users to potentially aggressive surveillance: It installed 168 tracking tools that didn’t let users decline to be tracked, and 121 tools that, according to their privacy statements, don’t rule out collecting financial or health data. Dictionary.com attributed the number of tools to its use of many different ad networks, each of which puts tools on its site.

Some of the tracking files identified by the Journal were so detailed that they verged on being anonymous in name only. They enabled data-gathering companies to build personal profiles that could include age, gender, race, zip code, income, marital status and health concerns, along with recent purchases and favorite TV shows and movies.

The ad industry says tracking doesn’t violate anyone’s privacy because the data sold doesn’t identify people by name, and the tracking activity is disclosed in privacy policies. And while many companies are involved in collecting, analyzing and selling the data, they provide a useful service by raising the chance Internet users see ads and information relevant to them personally.

"We are delivering free content to consumers," says Mike Zaneis, vice president of public policy for the Interactive Advertising Bureau, a trade group of advertisers and publishers. "Sometimes it means that we get involved in a very complex ecosystem with lots of third parties."

The growing use and power of tracking technology have begun to raise regulatory concerns. Congress is considering laws to limit tracking. The Federal Trade Commission is developing privacy guidelines for the industry.
If "you were in the Gap, and the sales associate said to you, 'OK, from now on, since you shopped here today, we are going to follow you around the mall and view your consumer transactions,' no person would ever agree to that," Sen. George LeMieux, R-Florida, said this week in a Senate hearing on Internet privacy.

Write to Julia Angwin at julia.angwin@wsj.com

- How to Avoid Prying Eyes
- What They Know About You
- Analyzing What You Have Typed
- Decoding the Trackers: A Glossary
- The Journal's Methodology