A Web Pioneer Profiles Users by Name

By EMILY STEEL

Linda Twombly at her Nashua, N.H., home.

In the weeks before the New Hampshire primary last month, Linda Twombly of Nashua says she was peppered with online ads for Republican Senate hopeful Jim Bender. It was no accident. An online tracking company called RapLeaf Inc. had correctly identified her as a conservative who is interested in Republican politics, has an interest in the Bible and contributes to political and environmental causes. Mrs. Twombly’s profile is part of RapLeaf’s rich trove of data, garnered from a variety of sources and which both political parties have tapped.

A company called RapLeaf is building databases on people by tapping voter-registration files, shopping histories, social-networking activities and real estate records. WSJ's Emily Steel and Julia Angwin join the Digits show to discuss which sites are using RapLeaf, and what web users can do to try to protect their privacy.

Cracking the Code

- RapLeaf ties people's email address to a profile about them. It uses that profile to target ads. See data RapLeaf had on one user and how the system works.

The market for data about Web users is hot—and one of the methods used is "scraping," harvesting online conversations. In May, Nielsen scraped private forums where patients discuss illnesses. How can web users prevent their data from being scraped? Julia Angwin joins Digits to discuss.

RapLeaf knows even more about Mrs. Twombly and millions of other Americans: their real names and email addresses. This makes RapLeaf a rare breed. Rival tracking companies also gather minute detail on individual Americans: They know a tremendous amount about what you
do. But most trackers either can't or won't keep the ultimate piece of personal information—your name—in their databases. The industry often cites this layer of anonymity as a reason online tracking shouldn't be considered intrusive. RapLeaf says it never discloses people's names to clients for online advertising. But possessing real names means RapLeaf can build extraordinarily intimate databases on people by tapping voter-registration files, shopping histories, social-networking activities and real estate records, among other things.

"Holy smokes," says Mrs. Twombly, 67 years old, after The Wall Street Journal decoded the information in RapLeaf's file on her. "It is like a watchdog is watching me, and it is not good."

Some early adopters of the service are political campaigns. Democratic political consultant Chris Lehane used RapLeaf in a successful campaign against Proposition 17 in California, which would have changed the way auto-insurance rates are set in the state. RapLeaf ranks among the most sophisticated players in the fast-growing business of profiling people online and trading in personal details of their lives, an industry that is the focus of a Journal investigation. The San Francisco startup says it has 1 billion e-mail addresses in its database.

RapLeaf acknowledges collecting names. It says it doesn't include Web-browsing behavior in its database, and it strips out names, email addresses and other personally identifiable data from profiles before selling them for online advertising. Nevertheless, the Journal found that, in certain circumstances, RapLeaf had transmitted identifying details about Mrs. Twombly—such as a unique Facebook ID number, which can be linked back to a person's real name—to at least 12 companies. The Journal also found RapLeaf had transmitted a unique MySpace ID number (which is sometimes linked to a person's real name), to six companies. MySpace is owned by News Corp., which publishes the Journal.

RapLeaf says its transmission of Facebook and MySpace IDs was inadvertent and the practice was ended after the Journal brought it to the company's attention. The company says people can permanently opt out of its services at RapLeaf.com. RapLeaf executives say their business offers valuable consumer benefits by allowing people to see relevant advertising and content. "The key goal of RapLeaf is to build a more personalizable world for people," says RapLeaf CEO Auren Hoffman. "We think a more personalizable world is a better world." When a person logs in to certain sites, the sites send identifying information to RapLeaf, which looks up that person in its database of email addresses. Then, RapLeaf installs a "cookie," a small text file, on the person's computer containing details about the individual (minus name and other identifiable facts). Sites where this happened include e-card provider Pingg.com, advice portal About.com and picture service TwitPic.com.

In some cases, RapLeaf also transmits data about the person to advertising companies it partners with. Data gathered and sold by RapLeaf can be very specific. According to documents reviewed by the Journal, RapLeaf's segments recently included a person's household income range, age range, political leaning, and gender and age of children in the household, as well as interests in topics including religion, the Bible, gambling, tobacco, adult entertainment and "get rich quick" offers. In all, RapLeaf segmented people into more than 400 categories, the documents indicated.
RapLeaf's privacy policy states it won't "collect or work with sensitive data on children, health or medical conditions, sexual preferences, financial account information or religious beliefs." After the Journal asked RapLeaf whether some of its profile segments contradicted its privacy policy, the company eliminated many of those segments. Segments eliminated include: interest in the Bible, Hispanic and Asian ethnic products, gambling, tobacco, adult entertainment, "get rich quick" offers and age and gender of children in household.

RapLeaf says many of its segments are also "used widely by the direct-marketing industry today." In this year's hotly contested midterm elections, some political organizations are tapping RapLeaf's technology. With traditional postal mailing lists, "We used to bombard their house with mail. Now we can bombard their house with online ads," says Robert Willington, the Republican online campaign strategist who worked on behalf of Mr. Bender's New Hampshire campaign. RapLeaf helped Mr. Bender's campaign target likely Republican voters with ads online. (Mr. Bender, who confirms working with RapLeaf, lost the election.)

In Mr. Lehane's California effort against Proposition 17 this year, RapLeaf found online about 200,000 suburban women over the age of 40 in Southern California, a demographic the campaign considered swing voters. Mr. Lehane says the 4-percentage-point margin of defeat suggested the technology was effective. "With an election that close, every voter you can reach matters," he says. Mr. Lehane says he was considering using RapLeaf as part of a campaign against Meg Whitman, who is running for governor in California. That campaign is being run by a political group, Level the Playing Field 2010, which was funded by several labor unions and which Mr. Lehane led.

RapLeaf says it has participated in about 10 campaigns this season, declining to identify them. "We expect that forward-thinking campaigns will begin to use it this year more widely as an alternative to direct mail, email and phone calls," says Joel Jewitt, RapLeaf's vice president of business development.

Co-founded in 2006 by Mr. Hoffman, a Silicon Valley entrepreneur, RapLeaf began as an online service letting people rate each other based on their business transactions. The company raised an initial $1 million in funding from well-known Silicon Valley investors including PayPal co-founder and Facebook investor Peter Thiel. A person familiar with the situation says the company closed a $15 million fund-raising round this month.

Soon after it was founded, RapLeaf began "scraping"—or collecting information from—social networks to build a people search engine. It matched data from social-networking profiles with email addresses. RapLeaf says data it collects are public. It sold a service giving companies information about the customers on their e-mail lists.

By 2009, RapLeaf had indexed more than 600 million unique email addresses, it said in a press release that year, and was adding more at a rate of 35 million per month. Meanwhile, the business of helping marketers with their email lists (RapLeaf's core) was lagging in the recession. And the online-tracking business was taking off.
The Wall Street Journal analyzed the tracking files installed on people's computers by the 50 most popular websites, plus WSJ.com. Explore the data here and see separate analysis of the files on popular children's sites.

RapLeaf's Mr. Jewitt says the company saw an opportunity: It decided to connect its database of dossiers on people to cookies placed on those same individuals' computers, for ad targeting. "If you are a modern information company, you have to be involved in that," he says.

Combining off-line profiles with online tracking has raised red flags ever since another company first tried it 10 years ago. Privacy advocates argued that connecting people's Web-browsing habits with their names was too intrusive. RapLeaf says it doesn't share or sell emails. However, under some circumstances it will provide names and other personal details if a client already possesses that person's email address.

For example, a company might come to RapLeaf with an email-address mailing list, and RapLeaf will try to provide information about the people on that list. This year, RapLeaf began offering services to target these people with online ads for the client. For that to work, RapLeaf relies on a network of cooperating websites that use email addresses as part of the sign-on process. Those sites agree to transmit their users' email addresses (in encrypted form) to RapLeaf. Then, RapLeaf "drops," or installs, cookies on users' computers.

It's tough to build up a network of such sites, because many don't want to let outsiders track their visitors. This summer, RapLeaf sent a marketing email offering to pay one website an unspecified sum for this kind of access, according to documents reviewed by the Journal. The website chose not to take the offer.

RapLeaf declined to name the sites it works with, citing nondisclosure agreements. The Journal found that sites installing RapLeaf cookies included About.com, owned by the New York Times Co.; online invitation site Pingg.com; photo-sharing sites TwitPic.com and Plixi.com; movie site Flixster.com; discount site Tester-Rewards.com; and some Facebook.com and MySpace.com applications.

The Journal last week reported on the Facebook and MySpace apps sending data to RapLeaf. Both sites say they prohibit applications from sharing user data with outside data companies, and that they took steps to stop the apps that were transmitting user data to RapLeaf.

A Facebook spokesman says the company is acting to "dramatically limit" the exposure of users' personal information. Facebook says the user ID allows access only to information that Facebook requires people to make public in their profile.

MySpace says it uses RapLeaf data for its "friend recommendation" system, but doesn't share user data or let RapLeaf track MySpace users. After receiving user IDs from some MySpace and Facebook apps, RapLeaf was then transmitting data about users to its advertising partners. After being contacted by the Journal, RapLeaf says it "acted immediately" to strip out identifying information from the data it shared with partners.
An About.com spokeswoman says the company doesn't have a relationship with RapLeaf. She says users' information was sent to RapLeaf via a partner that operates on its site, and that About.com wasn't aware its users' email addresses were being sent to RapLeaf.

Plixi.com says the company is "in experiment mode right now with behavioral-targeting companies like RapLeaf." Flixster.com says it "does not sell any of our users' personal information to anyone" and declined to comment further.

Pingg.com declined to comment. TwitPic and Tester-Rewards didn't respond to requests for comment.

The Journal decoded RapLeaf's information on Gordon McCormack Jr., a 52-year-old who lives in Ashland, N.H. RapLeaf correctly identified Mr. McCormack's income range, number of cars (one), his interests in gardening and the Beatles, and his interest in playing the online game Mafia Wars, among other topics. Mr. McCormack says he plays Mafia Wars almost every day before going to bed. RapLeaf also identified Mr. McCormack as someone with an interest in online personals. He says he isn't currently active in online dating, but might have a couple of profiles "lurking on the Internet." When Mrs. Twombly, the New Hampshire Republican, registered at Pingg.com using her email address, RapLeaf matched her to dozens of "segments," according to a Journal analysis of the computer code transmitted while she was on the site.

The Journal was able to decode 26 of the segments, including her income range and age range and the fact that she is interested in the Bible and in cooking, crafts, rural farming and wildlife. Mrs. Twombly says all the decoded segments describe her accurately. RapLeaf says some of the segments in Mrs. Twombly's and Mr. McCormack's profiles "do not exist," possibly due to changes in RapLeaf's overall segment list in the time since their web traffic was decoded for this article last month. In Mrs. Twombly's case, RapLeaf transmitted data about her to at least 23 data and advertising companies after she logged into Pingg, according to the analysis of the computer code.

Twenty-two companies, including Google's Invite Media, confirmed receiving data from RapLeaf. RapLeaf declined to comment on its relationships with the companies. Since talking with the Journal, Mrs. Twombly tweaked her Web browser to limit cookie installation. As a result, she says, some websites don't always work properly for her, a common side effect of restricting cookies. Mrs. Twombly also removed applications from her Facebook profile that were transmitting data to RapLeaf, the Best Friends Gifts and Colorful Butterflies apps. The maker of those apps, Lolapps Media Inc., says it stopped working with RapLeaf.

Still, Mrs. Twombly is no longer using those apps to send virtual gifts and butterflies to her online friends. "My neighbor did send me a hug or a rainbow or a heart or something like that, but I didn't respond," Mrs. Twombly says. "Once burned, twice shy."

—Julia Angwin and Peter Wallsten contributed to this article.