Society | DOI:10.1145/2686740 Gary Anthes

Data Brokers Are Watching You

You would be surprised by how much they know about you, and what they are doing with your information.

AIDED BY ADVANCES in data science and the increased digitization of analog information, an industry little known to the public is quietly compiling comprehensive dossiers on millions of Americans. The companies, called data brokers, say they operate within the law, but the unprecedented breadth and depth of the data files, the difficulty in correcting erroneous data, and the potential for abuse of personal information are raising alarms from privacy advocates, consumer groups, and government officials.

Data brokers compile information about individuals from a wide variety of online and offline sources, including email, personal websites, social media posts, U.S. Census records, retailers' systems, Department of Motor Vehicles records, and real property records. The data is often collected without the consent or knowledge of the individuals involved, integrated and synthesized using advanced analytic tools, then sold to other data brokers and businesses for a variety of purposes.

"You may not know them, but data brokers know you," said Edith Ramirez, chairwoman of the U.S. Federal Trade Commission (FTC) in a statement last May. "They know where you live, what you buy, your income, your ethnicity, how old your kids are, your health conditions, and your interests and hobbies."

Ramirez' statement accompanied the release of a report, "Data Brokers: A Call for Transparency and Accountability" which, in turn, was based on information supplied by nine such brokers: Acxiom, CoreLogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, RapLeaf, and Recorded Future. One of them reported having a database of 700 billion data elements culled from 1.4 billion consumer transactions; a second had information from $1 trillion worth of consumer transactions, and a third said it was adding three billion new records to its database each month.

Digitization of analog data, along with advances in algorithms behind data analytics, has enabled a dramatic leap in the ability of data brokers to track individuals.

The FTC has called for legislation that would give consumers greater access to data brokerage practices and more control over their own information. The agency suggested Congress should consider requiring brokers to create a centralized portal where consumers could look at all their data and opt out of having it used.

While personal data is sold for a variety of purposes, many of the brokers' customers use the information for targeted marketing. The FTC said one broker segments consumers into handy buckets with labels such as "Urban Scramble" (heavily populated with low-income Latinos and African Americans), "Rural Everlasting" (single men and women over the age of 66 with little education and small net worths), and "Married Sophisticates" (upper-middle-class young adults with no children). More narrowly defined groups included "Expectant Parent," "Diabetes Interest," and "Cholesterol Focus."

It is easy to imagine a financial company aiming sub-prime loans at "Urban Scramblers" while pitching platinum credit cards to "Married Sophisticates," privacy advocates warn. Even worse, how can anyone be sure that membership in one of the medical cohorts doesn't affect one's insurance rates? "The companies see what website you visit, then they add offline information to that," said Ed Mierzwinski, program director at U.S. PIRG, the federation of state Public Interest Research Groups. "Then the website you go to next might look different than it otherwise would."

The process described by Mierzwinski, known as "onboarding," has drawn special scrutiny and concern from the FTC, Congress, and the White House. In onboarding, a data broker will add offline information—data from manual sources or from other systems such as loyalty cards, warranty registrations, and stores' point-of-sale terminals—into the cookies of computers used by individuals to access websites monitored by the broker. Once in place, the cookies can track the Web activity of the person from place to place, serving up targeted ads at each site. "Data brokers are helping to blur the line between online and offline behavior," said the FTC's Ramirez. "[They] use your offline purchases and information to find and target you online."

With so much information coming from so many sources, it is inevitable that errors arise in digital dossiers; moreover, the errors can be difficult or impossible to correct. In a study published in 2013, the FTC reported one in five consumers had an error in one or more of their credit reports. These errors can lead to an unfairly poor credit

rating that can prevent someone from getting a much-needed loan. The FTC also found more than 10% of consumers saw a change in credit score after getting errors corrected.

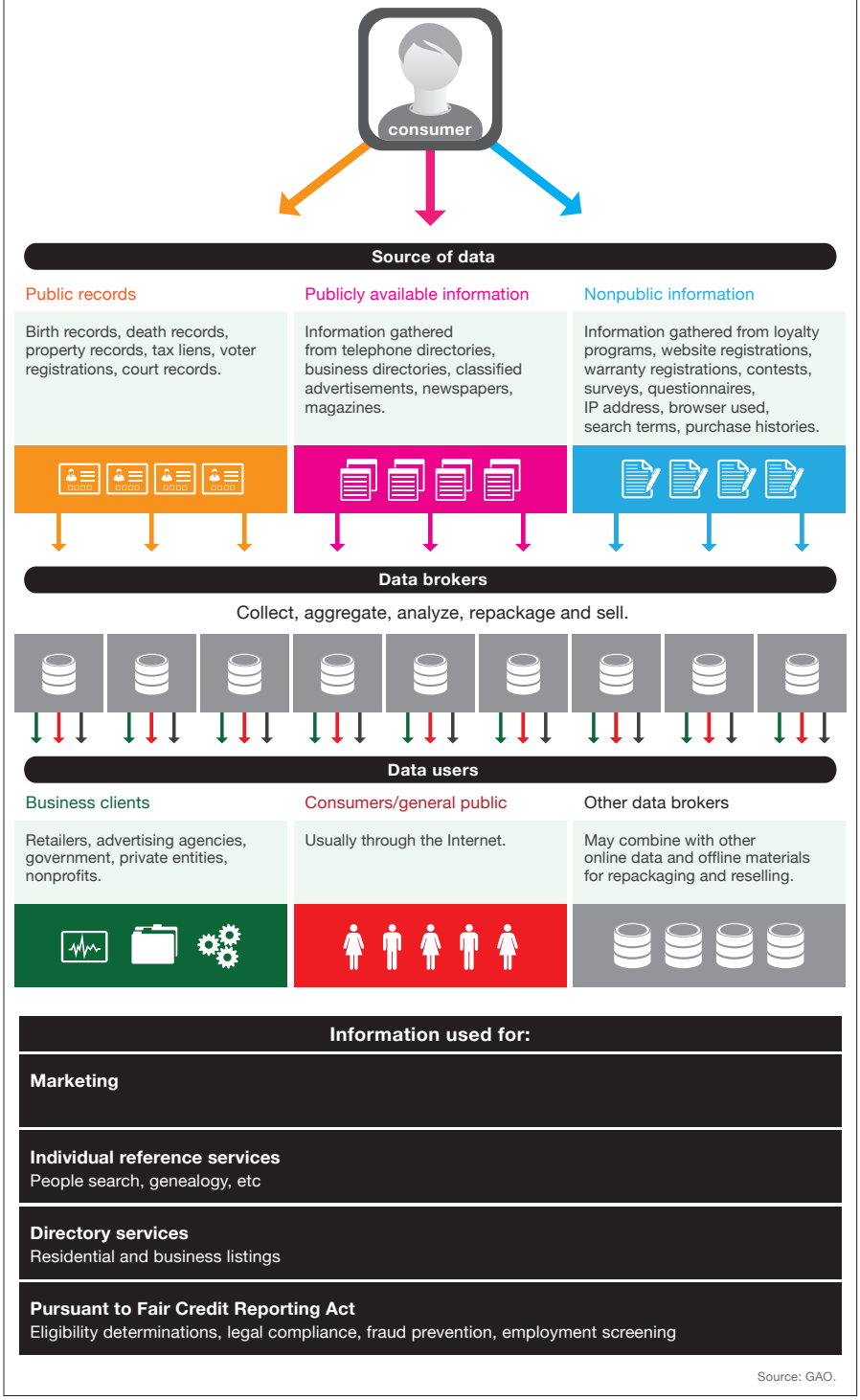
Many of the gatherers of online information offer ways to correct the personal data they hold, but the methods for doing so are often not well publicized, or are difficult to follow. In any case, consumers often do not know who is collecting or using their data, and there is no central clearinghouse for verifying all of one's online data. Even the most diligent, tech-savvy people cannot ensure the integrity of all their data, anymore than they can control its use.

The rapid increase in the number of available sources of data is especially troubling when it involves the capture of analog information, said Susan Graham, a computer science professor at the University of California, Berkeley (UC Berkeley). That is because the devices that capture this information—surveillance cameras and various kinds of sensors—for valid technical reasons sweep up more data than is needed for the immediate purpose. For example, she said, dramatic improvements in high-resolution digital cameras and facial recognition software will soon make it possible to identify people gathering at news events or in cars passing traffic law-enforcement cameras. With such abilities come obvious possibilities for abuse, she said.

The digitization of vast amounts of previously analog data, plus advancements in the algorithms behind data analytics, have enabled a dramatic leap in the ability of data brokers to track and understand the day-to-day activities of individuals, Graham said. "You can put together very disparate pieces of information and create a profile of a person. It's not necessarily correct, but you can do it very quickly, and you can do it with information that the individual may not have provided. That's a fairly profound shift."

Much of this shift has been hidden from the public, according to a year-long investigation of data brokers by the U.S. Senate Committee on Commerce, Science, and Transportation. "The refusal by several major data broker companies to provide the committee complete responses regarding data sources and customers only reinforces the aura of secrecy surrounding the industry," the committee said in its final report in late 2013.

Flow of consumer data through data brokers to third-party users



Source: GAO.

IMAGE COURTESY OF YAHOO FINANCE

Industry Response

The Direct Marketing Association, which calls itself "the world's largest trade association for data-driven marketers," said its members operate entirely within the law. Asked if the actions of data brokers nevertheless raise legitimate privacy concerns, DMA senior vice president Stephanie Miller responded, "No. Marketing data used responsibly for marketing purposes is a consumer benefit, providing much of the value in our data-driven

lifestyles." She cited an independent study that showed the "data-driven marketing economy" contributed $156 billion to the U.S. economy and supported 675,000 jobs in 2012. Miller also said a number of data brokers, as well as vendors such as Google and Yahoo, already offer consumers ways to opt out of targeted marketing.

Indeed, the FTC cited a number of benefits that flow from data-driven marketing. "Data broker products help to prevent fraud, improve product offerings, and deliver tailored advertisements to consumers," the FTC said in its report. The brokers also foster competition by enabling small businesses to pitch innovative products to consumers they could not otherwise reach. The agency also acknowledged the existence of laws that protect the data of consumers in certain industries, such as finance and health care.

For example, the Fair Credit Reporting Act (FCRA) of 1970 regulates and restricts the use of consumer data when such data may be used for "eligibility" determinations in credit, employment, insurance and housing. Last April, the FTC settled with two data brokers for selling data in violation of the FCRA. The FTC found the companies had sold data to employers and landlords without ensuring the accuracy of the data, or that the buyers had legitimate reasons for wanting the information, as required by law.

David LeDuc, a senior director at the Software & Information Industry Association, said he is not sympathetic to those calling for legislative reforms because they find what data brokers do "creepy." "We oppose the imposition of unnecessary barriers to the collection and use of data," he said. "The focus should be on real harm, not crystal ball gazing about what makes certain people uncomfortable."

Solutions

The use of personal information by data brokers and others has become so pervasive that limiting its collection by existing means has become unworkable. Asking consumers to "opt out" of data collection at myriad companies they have never heard of is unrealistic, and the existing online "notice and consent" forms—in which users "agree" to the collection and use of personal data—are ineffective because they are mostly ignored by consumers.

Instead, said UC Berkeley's Graham, government policy as well as technology should focus more on the use of personal data, and less on its collection. Her views and those of 10 other information technologists, businesspeople, and policy experts appear in a report submitted to President Obama last May, "Big Data and Privacy: A Technological Perspective." In the report, a working group of the President's Council of Advisors on Science and Technology (PCAST) argued policies that focus on data "collection, storage, applications, and analysis" are not scalable, as it becomes increasingly difficult to ascertain what personal information may be latent in a particular dataset or its fusion with other data. Also, PCAST said, policies limiting collection and retention will become increasingly unenforceable by other than draconian means.

To control use, individuals might choose a privacy preference profile offered by third parties, PCAST said. For example, "Jane" might choose one offered by the American Civil Liberties Union that gives special weight to privacy, while "John" might prefer one from *Consumer Reports* that emphasizes economic value to the consumer. Market forces or government regulation would compel the users of personal data to conform to the profiles.

Technology is already moving to enable that kind of control. For example, commercial privacy systems have been developed by firms such as Booz Allen Hamilton and IBM and are in use by a few government agencies, financial services firms, and pharmaceutical firms. These systems are based on the Trusted Data Format (TDF) for file-level tagging and security.

Expecting consumers to understand and specify their privacy preferences may be unrealistic at present.

Expecting consumers to understand and specify their privacy preferences may be unrealistic at present, said Mark Gorenberg, managing director of Zetta Venture Partners and a member of the PCAST working group. "You need to create a market for this," he said. "You'd see products and systems and cloud-based services with usage-based components in them." If major vendors such as Google and Amazon began offering file tagging and tracking services, perhaps based on TDF, then the public might come to use them, he said. If not, Congress would have to decide whether to mandate some kind of usage-based controls, he said.

"It would be a mistake to stop collecting data which, when combined with other data, gives much better results to the public," Gorenberg said. "We are at a golden age of being able to use data to get better results for companies and individuals."

Graham sees the glass as half-full. "Things are becoming technically feasible today that would have been crazy 10 years ago—for the good guys and the bad guys." C

Further Reading

U.S. Federal Trade Commission, Data brokers: a call for transparency and accountability May 2014
http://1.usa.gov/1kXR5g0

Fertik, Michael
The Rich See a Different Internet Than the Poor, *Scientific American*, Jan. 15, 2013
http://bit.ly/1qtmpQt

Executive Office of the President, Big data and privacy: a technological perspective, President's Council of Advisors on Science and Technology, Big Data and Privacy Working Group, May 2014 http://1.usa.gov/1ro5aq5

Russell, Matthew A.
Mining the social web, 2nd edition, O'Reilly Media, October 2013
http://oreil.ly/1ljlIwX

U.S. Senate Committee on Commerce, Science, and Transportation, *A review of the data broker industry: collection, use, and sale of consumer data for marketing purposes*, Staff report for Chairman Jay Rockefeller, Dec. 18, 2013
http://1.usa.gov/1vlVESn

Gary Anthes is a technology writer and editor based in Arlington, VA.