UNIVERSITY OF CALIFORNIA

CHRIS JAY HOOFNAGLE
Adjunct Full Professor
School of Information
School of Law

Faculty Director
Berkeley Center for Law & Technology

University of California, Berkeley
102 South Hall
Berkeley, CA 94720-4600
Tel: 510.643.0213
choofnagle@berkeley.edu
https://hoofnagle.berkeley.edu

July 13, 2017

VIA E-Mail MyLanh.*Graves@Vermont.gov*

My-Lanh Graves
Vermont Attorney General's Office
109 State St.
Montpelier, VT 05609

Re: Data Brokers

Dear Ms. Graves,

Thank you for soliciting comments on data brokers. I wrote an early article discussing the problem of data brokers, *Big Brother's Little Helpers, How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C.J. Int'l L. & Com. Reg 595 (2003), and have worked extensively on privacy over the past 15 years. Here I provide some high-level comment on your endeavor.

Defining Data Brokers

Defining data brokers is the most challenging part of this process. Data brokers will attempt to expand the legislative definition so that it includes all companies that provide information, such that even search engines and news organizations will be "data brokers." Privacy advocates, seeking perfect solutions, will agree with a broad definition, but in the process, your regulatory effort will fail because of overbreadth.

At the same time, attempts to define data brokers by the predominance of their information selling activities (a company *primarily* engaged in the sale of personal information…) are too narrow as information-based businesses can shape their practices to evade such activity levels.

Data brokers can also evade regulation by simply renaming or reclassifying their services. For instance, until the 1990s, it was commonplace for database marketing companies to sell lists of children by age and their home addresses for advertising purposes. Data brokers would sell lists of contact information for four- to six-year old children. This practice came into scrutiny in 1996, when the longtime CNN reporter Kyra Phillips, then working for a Los Angeles television station, purchased personal information on 5,500 children from Metromail, a data broker. To purchase the children's contact information, Phillips used the name of a notorious suspected child killer. Phillips' stunt generated publicity but it did not result in new restrictions on children's information. Instead, data brokers avoided regulation by renaming their products. The same information was sold but labeled as databases of households with "presence of children."

The Attorney General should avoid "boiling the ocean" by including every information-based firm in its definition of data broker. Instead, focus on the largest companies that specialize in selling information about consumers. Because of the economics of information selling, a few large companies own this market and interventions focused on these small number of companies are more palatable from a political economy standpoint.

Data Brokers Have Destroyed Most Hopes for a Market for Privacy

There is a need for regulation of data brokers because their practices undermine the ability of consumers to use the market to protect privacy, and because data brokers undermine consumers' ability to engage in privacy self-help.

Because of their dynamics, data brokers are immune from market incentives to promote privacy. First, data brokers have no direct consumer relationship with individuals. Second, data brokers purchase information from thousands of sources (such as websites and retail stores) secretly, and thus consumers cannot avoid having information transferred to data brokers. Third, because much of their data acquisition is secret, there is no practical way for consumers to link privacy obligations made by data collectors to the data brokers that ultimately amass the information. Fourth, when direct marketers (a major customer of data brokers) buy information, they are typically seeking only marginal gains in customer sales and acquisition. This search for small increases in sales means that they tolerate consumer lists that are wildly inaccurate. After all, a 1 percent increase in sales is considered a big success in direct marketing. As a result, data brokers have strong incentives to infer "facts" about individuals and to categorize them into various lists that can be sold to direct marketers.

The lack of market discipline for privacy means that data brokers engage in some of the most aggressive data uses, with little or no obligations to data subjects. For instance, after California prohibited retailers from asking customers their home addresses during credit card transactions, data brokers created tools that allowed retailers to infer this same information by merely asking for a telephone number.[1] When requesting the phone number was prohibited, data brokers encouraged retailers to collect the ZIP code, which also could be used to identify the customer's home address.[2] These activities are entirely outside consumer control, and were designed in the words of one data broker to avoid "losing customers who feel that you're invading their privacy." It demonstrates that even where consumers seek privacy or choose not to disclose data, businesses use systems to undo their efforts to protect personal information.

Individuals have no right to notice of their activities, no right to access files, and no right to correct the data, although some data brokers voluntarily provide notice and limited access to consumers. The lack

[1] This practice is known as reverse enhancement. With a telephone number, e-mail address, or credit card number, the data broker can use other databases to match the consumer's home address to the telephone number, and then provide the home address to the retailer.
[2] *Pineda v. Williams-Sonoma Stores, Inc.*, 246 P.3d 612 (SCT Ca. 2011).

of rights and secrecy involved also means that data brokers can sell data to scam artists or other unseemly businesses.

Data brokers' technical abilities are impressive. Through complex data-matching capabilities, they can link individuals' online activities to their real identity, and to their offline purchasing behavior. This linkage is thought to be so privacy-invasive that the Network Advertising Initiative, a group representing online advertisers once promised to never do it without affirmative consent back in 2000. They since reneged on the promise. And now data brokers do this linkage routinely for websites. They can tell who you are before you even log-in (see discussion of hashing and data merges below).

Self Regulation in Data Brokerage Has Failed

One of the first privacy self-regulation efforts concerned data brokers: The Individual Reference Services Group (IRSG). At the FTC's nudging, IRSG proposed a self-regulatory code, the IRSG principles, for data brokers. IRSG companies sold Social Security Numbers and other information to insurers, private investigators, law enforcement, and others.

Substantively, the IRSG code was weak. It allowed companies to sell any information they wanted to "qualified subscribers," which the companies were allowed to define. The code allowed individuals to opt out of the sale of information only to the "general public." In practice, this right was illusory because data brokers did not consider any of their customers to be members of the general public. One broker stated that there was no need for it to create an opt-out mechanism because their customers were all qualified.

The IRSG was short-lived. The code seemed to be designed only to head off statutory financial privacy rules, and after these were passed in 1999, the IRSG lived on to challenge those provisions as unconstitutional. Today the organization does not even maintain a website.

Useful factors for exploring self-regulation were proposed by the UK-based National Consumer Council. That group recommended the following:

1. A self-regulatory scheme must always have clear policy objectives.
2. Self-regulation should not inhibit the scope for competition to deliver benefits for consumers.
3. A strong independent element must be involved in the scheme's design and have a controlling influence on its governance.
4. A dedicated institutional structure must be set up, separate from the existing trade and professional organisations.
5. A pragmatic approach may be inevitable (meaning that self-regulation is the best bet where there is no practical chance of regulation).
6. There should be a presumption of scepticism towards self-regulation organised on a collective basis.
7. Effective self-regulation is usually best stimulated by a credible threat of statutory intervention.
8. Self-regulation works best within some form of legal framework.

Calls to just establish self-regulation should be evaluated in light of the IRSG's history and these useful guidelines.

Data Brokers Can Characterize Personal Information Sharing as Anonymous and Thus Not Problematic

Data brokers' impressive technical abilities make it possible for them to present highly-privacy invasive activities as anonymous and non-problematic. This is misleading. As Wolfie Christl recently wrote,

> *Data companies often remove names from their extensive profiles and use hashing to convert email addresses and phone numbers into alphanumeric codes such as "e907c95ef289". This allows them to claim on their websites and in their privacy policies that they only collect, share, and use "anonymized" or "de-identified" consumer data.*
>
> *However, because most companies use the same deterministic processes to calculate these unique codes, they should be understood as pseudonyms that are, in fact, much more suitable for identifying consumers across the digital world than real names. . .*[3]

While these processes are sometime described as anonymous, Christl observes:

> *Data management platforms allow businesses in all industries to combine and link their own data on consumers, including real-time information about purchases, website visits, app usage, and email responses, with digital profiles provided by myriads third-party data providers.*

These practices, highly dependent on data brokers, are confirmed by Facebook's former product manager for advertising, Antonio García Martínez. In his book CHAOS MONKEYS, Martinez wrote:

> *. . . Facebook and companies like Acxiom and Datalogix have compared personal data (with none sharing actual data with the other, again via the miracle of hashing), and joined the universal FB user ID to the analogous IDs inside Acxiom, Datalogix, and Epsilon.*
>
> *[. . .]*
>
> *Facebook, Google, and others have achieved the holy grail of all marketers: a high-fidelity, persistent, and immutable pseudonym for every consumer online. Even better, they've joined that to your real-world persona. . .*

To make this more explicit, Martinez continues:

> *That personal information is stored in a database, along with the browser cookies that corresponds to it, forming a bridge from real-world you to the browser version of you. It's probably in hashed form, but that's just privacy theater; if everyone agrees on the same hash function, it doesn't matter how it's stored."*
>
> *That join, between a cookie and personal information, is then sold and resold a bazillion times a day to whoever is willing to pay for it. . .*

Simply put, companies have misled the public by saying that they only share hashed or de-identified data. These datasets are easily re-identified, and so in effect, this is really personal information sharing.

---

[3] Wolfie Christl, Corporate Surveillance in Everyday Life, Cracked Labs, June 2017, http://crackedlabs.org/en/corporate-surveillance/.

Overcoming the Political Economy Problem

The political economy of data brokers makes them nearly impossible to regulate. Many kinds of businesses – including all lawmakers – purchase services from data brokers and they can do so secretly. In particular, the financial services industry has become reliant on data brokers and thus lobbies aggressively to keep data brokers free of privacy restrictions. Further complications come from the definition of data broker, as mentioned above. Companies that use data brokers can point to increased efficiencies and profitability. In competition with vague notions of "privacy," data brokers will always win in the legislative arena. Especially when their lobbyists can point to lawmakers' own use of these services and credibly claim that voter targeting and fund-raising would suffer from regulatory intervention.

Yet, the Attorney General should not be deterred. Even baby steps in the area could bring more transparency and accountability to the field. The Attorney General should consider:

- A focus on the small number of companies that are the quintessential data brokers. Avoid expansive definitions that sweep in search engines, consumer reporting activities, social networks, and the like.
- Be skeptical of self-regulatory alternatives, as the leading data broker group, the IRSG, does not even exist anymore, and when it existed, there is no evidence it did anything besides mount court challenges to regulation. If self-regulation is adopted, evaluate it against the National Consumer Council standards (see above).
- Approaches that focus on reducing the secrecy of data brokers' information collection:
    - A requirement that companies that provide data to data brokers specifically disclose this fact to consumers.
    - An ability for consumers to opt out/opt in to transfer of data from retailers and other companies they do business with to data brokers.
    - A requirement that clients of data brokers include "how did you get my information" disclosures on direct mail and other personalized advertisements.
- Approaches that focus on how data brokers use data:
    - A requirement that consumers can ask for an audit or disclosure log to determine how data about them has been sold and to whom.
    - A requirement that consumers can learn of the various "categories" they have been binned into and why.
- Approaches that reduce the deceptiveness of data brokerage
    - Prohibit as deceptive the practice of calling data "anonymous" or deidentified where data brokers and their clients can re-identify or otherwise link the data to individuals.
    - Prohibit data brokers from re-identifying datasets that were collected under promises of privacy or anonymity.

    - Prohibit data brokers from coaching clients to collect data in misleading ways (for an example, look at the reverse enhancement battle in California, where data brokers coached clients to collect zip codes because consumers would not realize that zip codes were personally identifiable).
- Approaches that internalize the externalities of data brokerage:
    - Data brokers could give notice of—or even share liability—where a client of a data broker uses personal information to swindle consumers.
    - Data brokers should be required to screen clients' use of their services, and refuse to provide services where the client is engaged in deceptive marketing, stalking, or other illegal behavior.

Respectfully submitted,

Chris Jay Hoofnagle