Alert

November 23, 2021

COZEN O'CONNOR

Web Scrapers And Their Targets Beware. Regulators Are Zeroing In On Privacy Implications

This Alert has been updated to include additional information about Vermont's claims against Clearview.

Selling web users' personal data is big business — with a projected worth of $400 billion by 2025. In industries as diverse as health insurance and automobile manufacturing, companies that collect and aggregate user data to sell it (often referred to as data brokers)[1] are becoming a powerful force. Often, data brokers utilize web scraping to collect this data, frequently without the data subjects' or owners' knowledge or consent. And while certain states have enacted laws to give consumers more control over their data while bringing a modicum of transparency to this industry, the patchwork of applicable laws and regulations can be confusing. Companies doing business with data brokers need to understand the risks associated with web scraping, including the fact that state attorneys general (AGs) are talking about these issues. Below, we provide a primer on web scraping as well as an overview of the legal and regulatory challenges.

What Is Web Scraping?

Web scraping, in general, is the process of collecting data from websites by automated methods. It can occur either with permission from the company or person hosting the website or, as in most cases, without any knowledge or authorization from the website owner. Despite its ubiquity, this topic has received very little public attention from regulators in the last 10 years. In the 2014 Federal Trade Commission Report on Data Brokers,[2] for example, web scraping is only mentioned in passing, despite the well-known use of web crawlers by data brokers at the time. Part of the reason for this lack of attention is the opaque nature of web scraping. Companies often create web crawlers and unleash them on public websites — no contract required, no permission requested. However, as regulators zero in on the privacy implications of collecting and disclosing user data, we predict web scraping will begin to draw their increasing attention and questions.

What Laws Could Potentially Regulate the Use of Scraped Data?

Regulatory enforcement actions related to web scraping are still relatively rare, which means there are few "bright line" rules. Most civil litigation on the issue has been brought under the Computer Fraud and Abuse Act (CFAA), 18 U.S.C. § 1030 — a federal anti-hacking law. To prevail, a plaintiff must establish that the defendant "intentionally accessed" a protected computer "without authorization" or otherwise exceeded his or her "authorized access." The CFAA does not define the term "without authorization" though. And, as a result, the courts' application of the CFAA to data scraping has been anything but uniform. Some courts have interpreted the CFAA to not prohibit accessing (or scraping) public information from websites, so long as it can be done without bypassing a "permission requirement" such as a password.[3] Other courts have construed the CFAA more broadly, and have found that a person "exceeds authorized access" if he or she accesses information for an improper or unauthorized purpose.[4] Thus, for publicly available information (i.e., information accessible without the need to enter log-in credentials), web scraping may not violate the CFAA — at least not right now in certain courts.

Meghan Stoppel, CIPP/US

Member

mstoppel@cozen.com
Phone: (720) 479-3880
Fax: (303) 625-4901

Related Practice Areas

- State Attorneys General
- Technology, Privacy & Data Security

However, the CFAA is only one of many regulations capable of reaching both web scrapers and the companies from whom they mine data. As is often the case, when federal laws are nonexistent or limited in their reach, state unfair and deceptive trade practice (UDAP) laws can be applied by state AGs. In addition, certain states, including Vermont and California, have enacted laws specific to data brokers to supplement their UDAP statutes.[5] Accordingly, state AGs may be better positioned than federal agencies or private litigants to regulate web scraping, since state UDAP laws could be interpreted to extend to both authorized and unauthorized scraping, in addition to conduct that may be covered by state data brokers laws or conduct otherwise lawful under the CFAA.

In Vermont, Clearview Is Contesting State AG's Allegations

In Vermont, for example, Attorney General T.J. Donovan sued Clearview AI, Inc. (Clearview) in March of 2020 for scraping photos online, in violation of the state's UDAP law and the state's first-in-the-nation data broker law. In that case, Vermont alleged Clearview used scraping to amass a database of three billion photographs to improve its facial recognition technology. Clearview, Vermont alleged, did this without consent and in violation of the certain websites' terms of use (alleged unfair acts under Vermont's UDAP law). The state also alleged Clearview made a series of deceptive statements to consumers and fraudulently acquired brokered personal information in violation of Vermont's Fraudulent Acquisition of Data Law (FADL). The case is still pending after Vermont defeated the majority of Clearview's motion to dismiss in September 2020. The court dismissed the AG's fraudulent acquisition of data claim, and one claim regarding deceptive statements.

While Vermont is unique with its data broker law, other state AGs also have focused on data brokers and their practices in the past few years. For example, in 2018, New York Attorney General Eric Schneiderman and Massachusetts Attorney General Maura Healey announced a joint investigation into the activities of political consulting firm Cambridge Analytica for allegedly scraping 50 million Facebook profiles. More recently, in a round of California Consumer Privacy Act (CCPA) regulations issued in 2020, the California Attorney General exempted certain data brokers from the statute's notice requirements, in seeming recognition that businesses without a direct relationship with the consumer "cannot feasibly provide a notice 'at or before the point of collection.'"[6] Against this backdrop, these and other state AGs appear poised to use their UDAP statutes against both data brokers and companies that allow web scraping to occur without proper disclosure and consent.

Why Should Companies Whose Data Is Being Scraped Care?

There are several reasons why companies should be concerned that their data is being scraped, or could be. First, if a company's data is being mined without authorization, then the company's ability to control that data is slipping away. Second, and more importantly, most companies are under increasing pressure, and regulatory obligation, to know how and with whom they are sharing their customers' data. Under recently enacted state privacy laws, consumers in certain states already, or will soon, have both the right to receive notice that their personal information is being collected and the ability to request that a company delete their personal information. This means companies need to understand whether the "scraped" data is protected as personal information under those state privacy laws and whether they have any ability, or legal obligation, to control access to that data. Of paramount importance, regardless of whether data is being scraped *with or without* authorization, businesses need to fully understand how their customers' data is being used, should they be questioned by consumers or state AGs.

Even with Authorization, Scraping May Violate State Privacy Laws

The scraping of consumer information by a third party from a website *with* permission may violate state privacy laws already in place in California and to be implemented in Virginia or Colorado. In those cases, deletion requests from consumers would need to be relayed to and implemented by the third parties engaged in the scraping. In California, this arrangement may also qualify as a "sale" of the consumer's personal information, which would trigger additional obligations under the CCPA (such as posting a "Do Not Sell My Personal Information" link on a company's website). Entities collecting information on Vermont and California residents may also have an obligation to

register as "data brokers"[7] with the Vermont or California AGs depending on what they are doing with the data and whether they have a "direct relationship" with the consumer. In Illinois, if the scraped data includes biometric information of consumers, that state's Biometric Information Privacy Act (BIPA)[8] may also be implicated. Therefore, if companies are selling or sharing consumers' personal information to web scrapers, due diligence at the outset (and regular monitoring of the web scraper) will be critical to ensure all parties are complying with applicable state privacy and data broker laws. The costs associated with managing these risks should be carefully weighed against any benefit derived from these partnerships.

Companies Need To Monitor for Scraping Without Authorization

If the web scraping is occurring without authorization, companies should consider taking steps to monitor and prevent this sort of activity — especially if the scraped information could include "personal information" as defined by state privacy laws. While "publicly available information" is often exempt from the reach of state privacy laws, this term is construed quite narrowly under most state laws. It does not necessarily mean anything publicly accessible on a website.[9] If unauthorized web scraping may be occurring, companies should pay special attention to their privacy policies and other consumer-facing communications. In addition to potentially violating state privacy laws, any misrepresentation or omission in such statements creates an opportunity for state AGs to allege violations of state UDAP laws.

Targets Should Act Quickly to Minimize Risk

Companies who are concerned that their customers' information is being scraped should take action immediately, both to protect their customers' privacy and minimize their own regulatory exposure. If an internal investigation confirms the company has been the target of web scraping, companies should consult with legal counsel and consider reporting the matter to the U.S. Department of Justice and/or the appropriate state AG for investigation under either the CFAA or state UDAP laws. At the same time, businesses should consider whether to pursue civil litigation against data brokers utilizing their consumer information without permission. Craigslist, for example, has repeatedly filed data scraping cases against third parties suspected of violating their terms of use in this manner.[10] And of course, all companies should regularly review and update their terms of use to ensure that there are no loopholes that will allow data brokers to scrape data without permission.

Data scraping was a much-discussed topic at the October 2021 National Association of Attorneys General Eastern Regional Meeting: "The Surveillance Economy: How Attorneys General Protect Privacy, Safety, and Equality in the Information Age," with most speakers agreeing there needs to be more conversation and transparency around the information data brokers are collecting and how they are using it. As more states consider (and eventually enact) state privacy laws giving consumers increasing control over their data, and imposing corresponding obligations on businesses, we predict an uptick in enforcement.

---

[1] The Federal Trade Commission defines data brokers as "Companies whose primary business is collecting personal information about consumers from a variety of sources and aggregating, analyzing, and sharing that information, or information derived from it, for purposes such as marketing products, verifying an individual's identity, or detecting fraud."

[2] Available here.

[3] *See e.g., hiQ Labs, Inc. v. LinkedIn Corp.*, 938 F. 3d 985 (9th Cir. 2019).

[4] *See e.g., United States v. John*, 597 F.3d 263, 271–72 (5th Cir. 2010); *Int'l Airport Ctrs., L.L.C. v. Citrin*, 440 F.3d 418, 420–21 (7th Cir. 2006); *EF Cultural Travel BV v. Explorica, Inc.*, 274 F.3d 577, 581–84 (1st Cir. 2001).

[5] *See* 9 V.S.A. § 2430 *et seq*; Cal Civ. Code §§ 1798.99.80 *et seq.*

[6] Final Statement of Reasons, Office of the Attorney General, p. 11 (2020).

[7] *See* 9 V.S.A. § 2430(4) and Cal. Civ. Code §§ 1798.99.80.

[8] *See* 740 Ill. Comp. Stat. 14/15(b)(3).

[9] Under the CCPA, "Personal information" does not include publicly available information. For these purposes, "publicly available" means information that is lawfully

made available from federal, state, or local government records, if any conditions associated with such information. “Publicly available” does not mean biometric information collected by a business about a consumer without the consumer's knowledge. Information is not “publicly available” if that data is used for a purpose that is not compatible with the purpose for which the data is maintained and made available in the government records or for which it is publicly maintained. “Publicly available” does not include consumer information that is de-identified or aggregate consumer information.” California Civil Code § 1798.140(o)(2).

[10] *See Craigslist, Inc. v. 3Taps, Inc.*, 2013 WL 1819999 (N.D. Cal. Apr. 30, 2013); *Craigslist, Inc. v. RadPad, Inc.*, No. 16-1856 (N.D. Cal. filed Apr. 8, 2016); *Craigslist, Inc. v. Instamotor, Inc.*, No. 17-02449 (N.D. Cal. filed April 28, 2017).