NEWMEYER & DILLION LLP

DATA SCRAPING: THEFT OR FAIR GAME?

By: Scott Satkin

Friday, November 30, 2018

Scott Satkin

Contact

949.271.7204
scott.satkin@ndlf.com

Practice Areas

Business Litigation
Business Transactions
Insurance Law
Privacy & Data Security
Real Estate Litigation
Real Estate Transactions

In the modern information age, data has become one of society's most valuable commodities. More and more information is being put out into the world all the time, and businesses, and even entire industries, centered on collecting and then selling data have sprung up. Others have seen more success in making data freely available to attract views and clicks and earning money indirectly, such as through ad revenue. This business model is the foundation of the social networking phenomenon, as well as many other data services. But what happens when another business collects, or "scrapes," that freely available data and then uses it in a way that the original aggregator disapproves of, or that may even be harmful to their business? Does doing the initial work to create or collect data entitle a company to control how that data is used, or should it be expected that this type of behavior will occur when information is made freely available on the internet? Courts have struggled to answer this question and have reached differing conclusions in different cases. Companies planning to make data publicly available online, or to use data made available by others, would do well to closely examine the cases that have been decided so far if they don't want their business disrupted by lengthy and costly litigation.

What does the Law Say?

The primary legal authority for legal disputes over data scraping is a federal statute called the Computer Fraud and Abuse Act (CFAA). Among many other things, the CFAA prohibits access without authorization or use that exceeds authorized access of a protected computer. The question then, is whether data scraping is authorized because data aggregators put information out on the internet where anyone can access it, or unauthorized, because a service may tell a particular data scraper to stop using their information in certain ways.

Cases That Have Gone Both Ways

In one instance, a company called 3Taps allegedly scraped the entirety of popular classified ads website Craigslist and made a copy of the site, including all of the ads posted there. Craigslist tried to get 3Taps to stop using their data, but 3Taps ignored and bypassed their attempts. Craigslist filed a lawsuit, alleging that 3Taps had violated the CFAA by accessing its website without authorization. 3Taps moved to dismiss the case, but the court found that while they likely had authorization to use Craigslist's data initially, Craigslist's argument that authorization had been revoked was reasonable, allowing the suit to proceed.

In another case, a company called hiQ Labs operated a service providing data to companies on their workforces, having obtained this data from popular professional social networking site LinkedIn. Like Craigslist, LinkedIn attempted to prevent hiQ from accessing the data on their website. In response, hiQ filed suit, alleging that LinkedIn's attempt to restrict their access amounted to unfair competition under state law, while LinkedIn counterclaimed that hiQ was violating the CFAA by accessing their site without authorization. When hiQ requested a preliminary injunction forcing LinkedIn to lift its

restrictions and allow access to its site, the court reasoned that the intent of the CFAA was to prevent malicious hacking rather than unwelcome use of public data, and granted the injunction, analogizing the situation to a business trying to claim that a customer was trespassing.

What Does this Mean for My Business?

Businesses, then, are left with a conundrum – can I make use of data that has been made publicly available for my own service, or will I be exposing myself to liability? Can I sue a company that is using data from my website against my wishes, or will it be a waste of my time to try? The law in this area is far from settled – in addition to the conflict between the Craigslist and hiQ decisions, both were only at the lowest level of courts, and neither involved the final resolution of the case. Moreover, the hiQ case, which came second, mentioned the Craigslist case but did not attempt to explain why the results differed despite the extremely similar facts.

There are, however, some lessons that can be taken away from these cases:

1. For those looking to use publicly available data, it may be wise to take a lesson from 3Taps in what not to do. 3Taps essentially copied the entirety of the Craigslist website. While the court did not explicitly rely on this fact in its decision, it likely made Craigslist more motivated to take action and made 3Taps a less sympathetic litigant once the case came before a judge. **Businesses looking to use publicly available data aggregated by another website may wish to distinguish their service as much as possible from the original source of their information.** For those looking to be as safe as possible, the surest route would be to seek permission from the source to use it before launching your own service, though this route does run the risk that your prospective source of information will charge a hefty price or even refuse permission altogether.
2. For those looking to prevent unwanted exploitation of their data, the most effective route **would be to restrict access in such a way that courts are less likely to consider the data publicly available.** In particular, the court in the hiQ case suggested that giving desired users passwords to access data while refusing them to those that are unwanted is likely a sure way to make unwanted access fall into the legal category of "unauthorized" under the CFAA. It would be hard to argue that hacking past a username and password requirement does not constitute unauthorized access. Of course, this also runs the risk of making a service less convenient to consumers, who may then be less likely to use it. Businesses will have to strike a balance between keeping out unwanted competitors, while not driving away necessary customers.

With conflict brewing in the lower courts, it is likely only a matter of time before appellate courts begin to weigh in on the issue of what, if anything, counts as unauthorized access to publicly available data, and provide some clearer guidance. Until then, businesses on both sides of the data scraping divide will have to tread lightly if they do not want to lose out in legal disputes.

Scott Satkin is an associate in the firm's Privacy & Data Security practice, and supports the team in advising clients on cyber related matters, including policies and procedures that can protect their day-to-day operations. For more information on how Scott can help, contact him at scott.satkin@ndlf.com.

ABOUT NEWMEYER & DILLION LLP

For almost 35 years, Newmeyer & Dillion has delivered creative and outstanding legal solutions and trial results for a wide array of clients. With over 70 attorneys practicing in all aspects of business law, privacy & data security, employment, real estate, construction, insurance law and trial work, Newmeyer & Dillion delivers legal services tailored to meet each client's needs. Headquartered in Newport Beach, California, with offices in Walnut Creek, California and Las Vegas, Nevada, Newmeyer & Dillion attorneys are recognized by The Best Lawyers in America©, and Super Lawyers as top tier and some of the best lawyers in California, and have been given Martindale-Hubbell Peer Review's AV Preeminent® highest rating.