Bloomberg BNA

Electronic Commerce & Law Report™

VOL. 21, NO. 28 JULY 20, 2016

WEB SCRAPING

Craigslist Inc. has been challenging the practices of web content scrapers over the past few years. Attorneys at Proskauer Rose LLP discuss Craigslist's latest complaint as well as the ongoing debate over the legality of web scraping.

Scraping of User-Generated Content: Craigslist Files Another Suit and the Debate Continues

BY JEFFREY D. NEUBURGER AND JONATHAN P. MOLLOD

Jeffrey D. Neuburger is a partner at Proskauer Rose LLP in New York and co-chair of the firm's Technology, Media and Communications practice group. His practice is focused on intellectual property and technology-related transactions, counseling and dispute resolution. He can be contacted at jneuburger@proskauer.com.

Jonathan P. Mollod is an attorney and technology, new media and communications content editor at Proskauer in New York. He can be contacted at jmollod@proskauer.com.

Screen or web scraping is an issue that has been controversial since the days of the dot-com boom. Website owners seek to make information available to users or subscribers, but, in turn, often seek to prevent competitors and other unauthorized parties from scraping or extracting content for their own commercial purposes (as well as engaging in any activity that may annoy end users or cause the site owner to incur unwanted IT-related costs). Content aggregators and data users, on the other hand, are always thinking of new and productive uses for the data readily accessible from websites, with scraping as an obvious technical measure to access that data. With the law on scraping still not yet fully developed after all these years, and open source and other tools to harvest website data widely available online, it's an issue (and a debate) that is not going away anytime soon.

One subset of the general controversy around screen scraping involves the scraping of user-generated content posted on social media and similar sites. Craigslist has been a leader in challenging that activity. For years, Craigslist has aggressively used technology and the law to challenge unauthorized parties who were scraping, linking to, or accessing user postings for commercial purposes.

- In 2010, for example, the classified advertisement site obtained a default judgment and permanent injunction against a website operator who sold software to automate posting ads on Craigslist, gather user email addresses and otherwise circumvent the site's security measures. *See Craigslist, Inc. v. Naturemarket, Inc.*, 694 F. Supp.2d 1039 (N.D. Cal. 2010).

- In 2012, Craigslist filed a widely-watched suit against a certain aggregator that was scraping content from the Craigslist site (despite having received notice that it was no longer permitted to access the site) and offering the data to outside developers through an API. In that case, a California district court ruled, among other things, that Craigslist may, through a cease-and-desist letter and use of IP address blocking technology, revoke a specific user's authorization to access that website. Such lack of ''authorization'' could form the basis of a viable claim under the federal Computer Fraud and Abuse Act and state law counterpart. *See Craigslist, Inc. v. 3Taps Inc.*, 2013 WL 4447520 (N.D. Cal. 2013).

In 2015, Craigslist subsequently settled the *3Taps* lawsuit, with relief against various defendants that included monetary payments and a permanent injunction barring the defendants from accessing any Craigslist content, circumventing any technological measures that prohibit spidering activity or otherwise representing that they were affiliated with Craigslist.

Craigslist v. RadPad

This past April, the *3Taps* saga, in a way, was resurrected when Craigslist brought a new scraping lawsuit, and with it, the debate over whether entities that mine data from websites populated with user-generated consent should incur any civil liability. Craigslist filed a complaint against the real estate listing site RadPad, an entity that had allegedly, for a limited period, scraped Craigslist data from 3Taps that it then used on its own website. In its complaint, Craigslist claimed that after the *3Taps* litigation was settled in June 2015, RadPad or its agents began their own independent efforts to scrape Craigslist's site. Craigslist alleged that RadPad used sophisticated techniques to evade detection and scrape thousands of user postings and thereafter harvested users' contact information to send spam over the site's messaging system in an effort to entice users to switch to RadPad's services. *See Craigslist, Inc. v. RadPad, Inc.*, N.D. Cal., No. 16-1856, *complaint filed* 4/8/16. In its complaint seeking compensatory damages and injunctive relief, Craigslist brought several causes of action, including breach of contract (the site's terms prohibit scraping and spidering activity), CAN-SPAM, CFAA (and California state law equivalent), and secondary copyright infringement.

With the law on scraping still not yet fully developed after all these years, and open source and other tools to harvest website data widely available online, it's an issue that is not going away anytime soon.

In response, RadPad filed its answer in May, arguing that Craigslist was attempting to exclude RadPad from accessing publicly-available information that would allow it to compete in the classified ad market for real estate rentals. In essence, RadPad is evoking the chorus of Weird Al Yankovic's parody song, ''Craigslist''—''Craigslist. I'm on Craigslist, baby, come on''—in suggesting that Craigslist and similar sites exert too much control over posted user-generated content and that it should be fair game for outside developers to use the postings for related services. In its counterclaim, RadPad claimed that, in its efforts to block RadPad, Craigslist has prevented e-mail messages containing the word ''RadPad'' from being delivered to landlords in response to Craigslist listings, an act that, it alleged, constitutes unfair competition.

The Ongoing Debate

Outside of the *RadPad* dispute, the debate over web scraping continues. Web services hosting valuable user-generated content or other data typically wish to exercise control over which parties can access and use it for commercial purposes. A carefully drafted website terms of use is helpful in specifying what is and isn't permitted on a site and, in appropriate circumstances, may provide a basis for a claim of unauthorized access and misuse of content. As such, social media platforms and other operators generally include restrictions on gaining unauthorized access to the site's systems, using automated software to submit or retrieve data, using sites for commercial purposes and bypassing the site's technical protection measures. In addition, site owners also employ technical measures, such as IP address blocks, CAPTCHA challenge screens, as well as industry standard robots.txt protocols to specify crawl delays and otherwise communicate the site's limitations on scraping and bot activity.

Others argue that, given the ease of employing standard technical measures to address scraping, the website owner should be under some obligation to use at least minimal technical efforts to limit scraping or signal to potential scrapers that it is not allowed. These might include throttling mechanisms that limit scraping activity or a crawl delay that uses the standard robots.txt script protocol. For example, in *QVC Inc. v. Re-*

sultly LLC, 99 F. Supp.3d 525 (E.D. Pa. 2015), the court denied the plaintiff's request for a preliminary injunction based on the CFAA against the maker of a shopping app whose scraping efforts unintentionally resulted in QVC suffering a website outage. The court reasoned that the defendant Resultly, a non-competitor, had assumed that QVC would handle any unwanted website requests by its bot via a crawl delay, a measure which QVC had not implemented. In denying QVC's preliminary injunction request, the court also found no threat of continuing harm since QVC has the ''undisputed ability to protect itself against any future outages caused by unknown bots.'' Interestingly, in further proceedings, the court allowed QVC's CFAA claim to go forward based upon the theory that authorization to access its site was rescinded because the defendant Resultly had signed a certain affiliate marketing agreement that may have placed it on notice about certain prohibitions against scraping QVC's site. *See QVC Inc. v. Resultly LLC*, E.D. Pa., No. 14-06714, 2/11/16.

Website operators have taken note of the recent cases in this area in an attempt to strengthen potential legal claims stemming from unwanted third-party access. In fact, in a notable decision from this month, the Ninth Circuit ruled that a violation of the terms of use of a website, without more, can't be the basis for liability under the CFAA. *See Facebook, Inc. v. Power Ventures, Inc.*, 9th Cir., No. 13-17102, 7/12/16. Instead of relying solely on restrictive terms of use, website operators are increasingly using technical barriers to prevent unwanted access and sending cease and desist letters to inform third parties that their access is revoked and thereafter ''unauthorized.''

Indeed, it seems possible that courts may follow the Ninth Circuit's holding in *Power Ventures* and adopt the principle that those who have permission to freely access a website can't act ''without authorization'' unless and until their authorization to access the website is specifically revoked. In a recent scraping dispute, an Indiana court ruled that CouponCabin, an online deal website, could proceed with CFAA and contract claims against entities that allegedly scraped its data, based, in part, on CouponCabin's sending cease and desist letters to almost all the defendants and having implemented technological blocks against cloud providers that hosted the defendants' traffic. *See Couponcabin LLC v. Savings.com, Inc.*, N.D. Ind., 2016 WL 3181826, 6/8/16 (''CFAA liability may exist in certain situations where a party's authorization to access electronic data—including publicly accessible electronic data—has been affirmatively rescinded or revoked'').

Under this informal scraping ''takedown'' regime, courts may be even more receptive to allowing ''unauthorized access'' and similar claims to go forward, given concrete evidence that an outside entity accused of unauthorized scraping had received notice that its access has been rescinded, particularly if technical barriers are erected to hinder future access. As a result, courts can make better reasoned evaluations of the CFAA ''unauthorized access'' issue and the law surrounding scraping just might become more settled.