Bloomberg BNA

Electronic Commerce & Law Report™

WEB SCRAPING

The legal landscape surrounding the legitimacy of web scraping continues to evolve. Attorneys from Perkins Coie revisit their prior analysis, discussing how two recent cases have found in favor of scrapers in non-competitive situations.

Web Scraping in an Era of Big Data 2.0

BY JAMES SNELL AND NICOLA MENALDO

When San Francisco rental prices surpassed those of New York last year, the San Francisco Chronicle speculated that online vacation rental services might have something to do with it. To test the theory, the newspaper gathered data.

Specifically, the Chronicle worked with another company to scrape rental listings from online services. The result: a $13 million Series A check for the company who worked with the Chronicle. The article and the startup's fundraising point to one conclusion: web scraping for analytics continues to be on the rise.

James Snell is a Partner in Perkins Coie's Privacy and Security practice. He represents and counsels clients on a wide range of complex commercial matters, including privacy and security, Internet, marketing and intellectual property litigation and matters.

Nicola Menaldo is an associate with the Perkin Coie's Commercial Litigation practice. Her litigation work centers on issues related to privacy and consumer protection.

The success of the young company that worked with the Chronicle reflects a larger trend across industries and applications toward mining the rich data available on the Internet for analytics purposes: journalists, professors, researchers, large and small businesses are all turning to the Internet to source information for big data analytics. Nate Silver, who famously predicted the outcome for 50 of the 50 states in the 2012 presidential elections, openly relies on scraping data from the Internet to generate highly-accurate and newsworthy sports and politics predictions and reporting.

Companies also use web scraping to obtain data for analytics-driven investing. Web scraping has become a central tool for statistical and scientific researching of all types. Last year, researchers even mined social media to identify correlations between tweets and heart disease.

While web scraping is increasing in use and appreciation across disciplines, its legal status remains highly context-specific. And many of the most interesting legal questions emerging from this trend remain unanswered or depend on very specific factual context.

For example, should an academic researcher who scrapes data from the web for a research paper be treated differently than a competitor who does the same thing? What if a website's terms prohibit scraping—is there any limit to what a website owner can prohibit by contract? What does it mean to prohibit using a website for ''commercial use''? Is there a difference between scraping for internal use versus using scraped material in a product offered to third parties? What about scraping by lawyers and watchdogs to evaluate claims made by public companies in their SEC filings?

In our 2013 article, *Use of Online Data in the Big Data Era: Legal Issues Raised by the Use of Web Crawling and Scraping Tools For Analytics Purposes* (with Derek Care) (18 ECLR 2466, 8/28/13), we discussed the various legal theories that website owners have used to attempt to hold web scrapers accountable for unwanted data collecting activities, as well as the various defenses available to data collectors. There, we identified five typical legal claims that arise out of web scraping activities:

(1) Copyright infringement. *See, e.g.*, *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818-22(9th Cir. 2013) (8 ECLR 721, 7/23/03) (holding that displaying low-resolution ''thumbnail'' copies of high-resolution photographs constituted reproduction of those photographs under the Copyright Act, but that the reproduction was highly transformative of, and didn't provide a substitute for, the plaintiff's high-resolution photographs, the purpose of which was primarily artistic).

(2) Breach of contract (where, for example, website terms prohibit use of a website by web crawlers, scrapers or other robots, or where they prohibit commercial use). *See, e.g.*, *Cairo Inc. v. Crossmedia Servs. Inc.*, N.D. Cal., No. 5:04-cv-04825, 4/1/05 (10 ECLR 382, 4/13/05) (discussing website's terms of use prohibiting access to defendant's websites with ''any robot, spider or other automatic device or process to monitor or copy any portion'' of the websites) (citation omitted); *Sw. Airlines Co. v. BoardFirst LLC*, 2007 BL 114340, N.D. Tex., No. 3:06-cv-00891, 9/12/07 (12 ECLR 967, 10/17/07) (holding that the defendant violated Southwest's terms of use restricting access to Southwest's website for ''personal, non-commercial purposes'' by offering a commercial service that helped Southwest's customers take advantage of the company's ''open'' seating policy and check-in process to obtain priority seating in the front of the plane).

(3) Violation of the Computer Fraud and Abuse Act (CFAA) or analogous state statutes. *See, e.g.*, *Craigslist Inc. v. 3Taps Inc.*, 942 F. Supp. 2d 962, 969-70(N.D. Cal. 2013) (18 ECLR 855, 5/8/13), (finding that Craigslist stated a claim against defendants that scraped its website after receiving a cease-and-desist letter prohibiting any access or use of Craigslist's website).

(4) Trespass to chattels. *See, e.g.*, *eBay Inc. v. Bidder's Edge Inc.*, 100 F. Supp. 2d 1058 (N.D. Cal. 2000) (granting motion for preliminary injunction based in part on eBay's likelihood to succeed on its trespass to chattels claim).

(5) Hot news misappropriation (confined generally to passing off another's breaking news as one's own). *See, e.g.*, *Barclays Capital Inc. v. Theflyonthewall.com, Inc.*, 650 F.3d 876 (2d Cir. 2011) (16 ECLR 1058, 6/22/11) (upholding dismissal of hot news misappropriation claim based on preemption under the Copyright Act).

We noted that the landscape relating to web crawling and scraping was still taking shape and that few courts had considered how to apply the above legal theories in purely non-competitive circumstances—where the business engaged in scraping wasn't directly competitive to the scraped website and was using the information to amass large quantities of data for analytics or other more attenuated purposes.

More than two years later, web scraping has become increasingly prevalent, but courts are only beginning to scratch the surface of how some of the theories listed above might apply in the context of big data. In this article, we discuss two recent cases involving scraping that analyze in detail liability under the CFAA: *QVC, Inc. v. Resultly, LLC* and *Fidlar Tech. v. LPS Real Estate Data Solutions Inc.* Together, these cases suggest that courts may be increasingly unwilling to protect websites from scraping activities where the websites don't take measures to protect themselves, even perhaps where significant damages result.

Recent Legal Developments in Scraping

In early web scraping cases, website owners sought relief under the CFAA for unauthorized access to protected computers. The CFAA establishes criminal liability for whoever (1) ''intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer,'' 18 U.S.C. § 1030(a)(2)(C); (2) ''intentionally accesses a protected computer without authorization, and as a result of such conduct, recklessly causes damage,'' *id.* at § 1030(a)(5)(B); and (3) ''intentionally accesses a protected computer without authorization, and as a result of such conduct, causes damage and loss,'' *id.* at § 1030(a)(5)(C).

As we discussed in our last article, courts were split on what constitutes ''unauthorized'' access for purposes of the CFAA. Most notably, in 2012, the U.S. Court of Appeals for the Ninth Circuit in *United States v. Nosal*, 676 F.3d 854 (9th Cir. 2012) (en banc) (17 ECLR 711, 4/18/12)held in an *en banc* decision that ''the phrase 'exceeds authorized access' in the CFAA doesn't extend to violations of use restrictions,'' but rather concerns ''hacking—the circumvention of technological access barriers.''

Since *Nosal*, there have been relatively fewer cases seeking to impose CFAA liability on web scrapers for violating website terms. However, website owners continue to look to the CFAA to combat unwanted scraping activities. In the last year, two important cases were considered that further clarified the reach of the CFAA in scraping cases and addressed issues beyond whether access to a protected computer was ''authorized.''

QVC v. Resultly

In *QVC Inc. v. Resultly LLC*, 99 F. Supp. 3d 525 (E.D. Pa. 2015) (20 ECLR 444, 3/25/15) a Pennsylvania district court considered whether a scraper violated the CFAA's prohibition on knowingly causing the transmission of code and intentionally causing damage. *See also* 18 U.S.C. § 1030(a)(5)(A) (imposing liability on any person who ''knowingly causes the transmission of a program, information, code, or command, and as a result of such conduct, intentionally causes damage without authorization, to a protected computer'').

Resultly was a startup company that used a web scraper to advertise products for sale that were posted on other websites. If a user wanted to buy a product displayed on Resultly's website, the user was directed through Resultly to the retailer's website to make the purchase and Resultly earned a commission through a layered affiliate marketing network.

In May 2014, Resultly began scraping the QVC website. QVC's website's terms of use didn't prohibit scraping. Soon after Resultly began scraping, QVC's servers experienced an overload that prevented consumers from making purchases on the QVC website and resulted in an alleged $2 million loss to QVC. QVC claimed that the overload was caused by the speed at which Resultly crawled its server and sued Resultly under the CFAA, seeking a preliminary injunction.

The central question in *Resultly* was whether Resultly intended to cause damage to QVC when it scraped its website. The court found that the relevant section of the CFAA required that a plaintiff allege and prove that the defendant ''*both* knowingly transmit[ed] a code *and* intend[ed] to cause damage to the plaintiff's computer.''

The court also found that in order to prove that a defendant intended to cause damage to a computer, the evidence had to show that it was the defendant's ''conscious objective'' to cause the damage. In other words, it wasn't enough under the CFAA to show that the defendant was technologically sophisticated and *should* have known that damage would be caused—the defendant had to want to cause damage.

In the case of Resultly, the court decided that Resultly didn't intend to cause any damage to QVC's server, and therefore QVC was unlikely to succeed in proving that Resultly violated the CFAA.

To reach the conclusion that Resultly didn't intend to cause damage to QVC's servers, the *Resultly* court considered a number of factors that we discussed in our 2013 article. First, the court rejected QVC's argument that Resultly's ''crawl rate'' of up to 40,000 hits per minute showed that Resultly intended to harm QVC's server. Resultly's procedure for crawling websites was to comply with any throttling requirements set forth in a website's robots.txt specification. (Robots.txt is a voluntary specification that a website owner can use to notify crawlers of any limitations the website owner wishes to impose on scraping.)

If a website didn't have a robots.txt specification, then Resultly crawled the website as fast as the server could respond to its requests. It had never had a problem using this system with other retailers. Here, because QVC had failed to implement a robots.txt specification that addressed crawl rate, Resultly crawled the QVC website as fast as the QVC server could respond to its request, which resulted in a high crawl rate.

The court determined that the crawl rate was insufficient to show that Resultly intended to harm QVC's servers because Resultly's procedure had never caused a problem in the past and because QVC could have specified a slower crawl rate, but didn't do so.

Second, the court also rejected QVC's contention that Resultly's failure to identify its user agent identifier as a bot indicated that Resultly intended to do harm. On this point, the court accepted Resultly's evidence that this misidentification had been a mistake that was corrected when another large retailer alerted it to the problem. *Id.* at 541.

The court ultimately determined that the evidence showed that ''if Resultly knew it would have damaged QVC's computer, it wouldn't have engaged in the conduct.'' Interestingly, the court was persuaded that Resultly didn't intend to cause damage based largely on what the court viewed as the non-competitive nature of Resultly's scraping activities.

The court emphasized that ''Resultly was not QVC's competitor, a disgruntled QVC employee, or an unhappy QVC customer aiming to cause damage to QVC's server'' and that Resultly's business depended on the QVC website running smoothly as well as QVC allowing Resultly to crawl its site. Based on this evidence, the court concluded that Resultly couldn't have intended to damage QVC's website.

Additionally, the court also noted that QVC used a third-party server, Akamai, to cache content and that Resultly's scraping activity was directed at Akamai's servers and not QVC's. For this reason too, the court determined that Resultly couldn't have intended to damage QVC's server.

Fidlar Technologies v. LPS Real Estate Data Solutions, Inc.

Most recently, the U.S. Court of Appeals for the Seventh Circuit addressed in *Fidlar Tech. v. LPS Real Estate Data Solutions Inc.*, 810 F.3d 1075, 1078-79(7th Cir. 2016) whether a data analytics company that accessed a technology company's online real estate data violated the CFAA.

In *Fidlar*, the plaintiff, Fidlar Technologies, had developed software for county offices to manage public land records. Fidlar licensed software to the counties, and the counties contracted with users for access to their records through the Internet.

The defendant, LPS Real Estate Data Solutions (LPS), was a data analytics company that developed a web harvester to download county records *en masse* through Fidlar's system. Fidlar sued LPS for trespass to chattels and violation of the CFAA for harvesting the data. On appeal, the Seventh Circuit affirmed the lower court's decision in favor of LPS.

The central issues on appeal were (a) whether LPS intended to defraud Fidlar for purposes of CFAA Section 1030(a)(4) (prohibiting any person from ''knowingly and with intent to defraud, access[ing] a protected computer without authorization, or exceed[ing] authorized access, and by means of such conduct further[ing] the intended fraud and obtain[ing] anything of value'') and (b) whether LPS caused damage to Fidlar.

On the first issue, the court rejected Fidlar's theory that evidence that LPS avoided printing fees by using its web crawler indicated intent to defraud. In reaching this conclusion, the court relied on a number of highly context-specific facts, including that LPS didn't understand that downloading documents regularly incurred a ''printing'' fee and that LPS paid the same full-subscription amount to counties that didn't charge for printing as those that did, so its intent was to download documents quickly, not to avoid fees.

However, the court also emphasized that Fidlar's terms of use didn't prohibit using a web crawler to access county records and that two of LPS's competitors were also using third-party programs to acquire county records through Fidlar's software. Since Fidlar didn't prohibit other companies from gaining automated access to its records, despite having knowledge of their activities, it appeared to the court that even Fidlar viewed LPS's activities as permissible. For all of these reasons, the court determined that no reasonable jury could find that LPS had intended to defraud Fidlar.

On the second point, the court found that LPS caused no damage to Fidlar. First, it rejected Fidlar's contention that, by not tracking the documents it accessed and downloaded, LPS was accessing Fidlar's computer in an unauthorized manner, thus causing damage. Second, the court also dismissed the theory that LPS caused damage to Fidlar's ''system'' by avoiding tracking, because the CFAA only protects computers, not systems.

Lessons Learned from *Resultly* and *Fidlar*

Both *Resultly* and *Fidlar* addressed whether a website owner can successfully sue a scraper for CFAA violations in the absence of a contract or other restriction (e.g., robots.txt) on scraping website content. In both cases, the courts determined that the CFAA theories posited by the plaintiffs were unconvincing.

It should be noted that the courts in these cases confined their analyses to the plaintiffs' CFAA claims, and it may well be that website owners in future cases are able to assert more successful claims under different theories of liability, even in the absence of clear contractual restrictions. However, one potential trend from these and previous cases is that courts are less willing to rule in favor of website owners where there is no enforceable contract prohibiting scraping, particularly where the scraping isn't competitive.

Resultly also emphasized the website owner's failure to employ the voluntary robots.txt protocol, which would have prevented the alleged damage had QVC required a slower crawl rate for unknown scrapers. Although the robots.txt specification is voluntary, it, like a website's terms of use, puts a scraper on notice of what the website owner intends to prohibit and allow. Indeed, robots.txt instructions can be more easily understood and digested by scrapers than a website's terms of use.

Thus, as the case law begins to address less competitive and more analytical uses for scraping, it will be interesting to see whether courts in those cases will find relevant whether a website owner has used the robots.txt protocol, and what instructions were given, in determining whether a scraper violated law.

Conclusion

The legal landscape relating to web crawling and scraping is still taking shape, and courts are still at the nascent stage of considering claims based on crawling or scraping for analytics purposes. Moreover, whether scraping or crawling for analytics purposes raises legal concerns is a highly fact-specific inquiry.

Nonetheless, the cases to date, including the two recent cases discussed above, suggest a number of issues that should be considered both by website owners and by those who seek to perform analytics using data gathered from web-based sources, including:

> **a) the language of the terms of use or service**, and whether such terms address access to the website through automated means, use of any data collected through such means and use of the website for other than the user's personal, non-commercial use;
>
> **b) the enforceability of the terms of use**, for example, whether they are presented to the user through a clickwrap mechanism that requires the user to indicate his or her assent to those terms as opposed to a browsewrap agreement, or on a terms of use page that can be reached through a conspicuous link on every other page on the website and which indicates that any use of the website is subject to the user's agreement to those terms;
>
> **c) use of technological tools** to deter unwanted crawling or scraping or to specify crawl rates, including but not limited to the robots.txt protocol;
>
> **d) whether access to the website is protected** such that a claim under the CFAA or California's Penal Section 502 may be alleged;
>
> **e) whether data on the website content is protected by copyright**; and
>
> **f) whether the website owner will license or authorize uses of content**.

It is inevitable that the uses of crawling and scraping for analytics purposes will continue to develop and that the courts will continue to grapple with the facts and legal theories applicable to instances of crawling and scraping. While the law continues to develop in this area, both website owners and scrapers should remain aware of signposts that have been identified in prior cases and be vigilant about staying abreast of future developments.

This article is made available by the lawyer or law firm publisher for educational purposes only as well as to give you general information and a general understanding of the law, not to provide specific legal advice. By reading this article you understand that there is no attorney client relationship between you and the article author. This article should not be used as a substitute for competent legal advice from a licensed professional attorney in your state. © 2016 Perkins Coie LLP.