New York Law Journal
SERVING THE BENCH AND BAR SINCE 1888
WWW.NYLJ.COM
An ALM Publication
VOLUME 258—NO. 49
MONDAY, SEPTEMBER 11, 2017

ANTITRUST TRADE AND PRACTICE

Expert Analysis

Big Data, Web 'Scraping' and Competition Law: The Debate Continues

By **Shepard Goldfein** And **James Keyte**

We live in a world—and deal with markets—increasingly driven by data. Consumers and companies throughout the globe generate massive amounts of data at any given moment. Internet searches, mobile phone clicks, website profile information, e-commerce transactions and basically any other action that can be quantified digitally make up the basis of "Big Data." This data can in turn be analyzed and studied to inform competitive decision-making and increase the accuracy of market predictions.

Big Data is a complex issue—different firms and individuals have different access to different sources of data, and those firms and individuals want to use that data in different ways.

This complexity means that the legality of some methods of culling and using Big Data remains unclear. But as Big Data's presence and importance to market success continues to grow, it will become increasingly necessary to consider its effect on antitrust analyses. As FTC Chairwoman Edith Ramirez explained in her Keynote Remarks at the Fordham Competition Law Institute's Annual Conference on International Antitrust Law and Policy in 2016, "[t]here is no question that the aggregation of data may have important implications for competition."[1] And EU Competition Commissioner Margrethe Vestager has vowed to "keep a close eye on how companies use data."[2] Some antitrust enforcement action is beginning to back up these claims. In June, the European Commission hit Google with a record €2.4 billion fine for abusing its market dominance in Internet searches to illegally benefit its own shopping-comparison service.[3]

Web "scraping" is one method of accumulating data that has sparked recent legal debate, both antitrust and otherwise. Web scraping is an automated process that firms can use to efficiently collect large amounts of targeted data from different websites. This scraping of information inherently involves accessing a site that is hosted by another company. In some cases the information that is scraped is "private" information of that site's users. Thus, legal challenges to web scraping have involved privacy claims and claims under the federal Computer Fraud and Abuse Act (CFAA), in addition to antitrust claims about the need to collect public data to be able to compete freely. The most recent legal decision to involve web scraping is *hiQ Labs v. LinkedIn*, No. 17-CV-03301-EMC, 2017 WL 3473663 (N.D. Cal. Aug. 14, 2017), from the Northern District of California. The case signals a shift in the way courts may be viewing attempts to restrict data scraping, giving web scrapers some arrows in their legal quiver to fight back against recent opinions condemning scraping.

> The 'hiQ' opinion has the potential to be a game-changer in the Big Data field.

SHEPARD GOLDFEIN *and* JAMES KEYTE *are partners at Skadden, Arps, Slate, Meagher & Flom.* DREW KABBES, *an associate at the firm, assisted in the preparation of this column.*

It also demonstrates the importance of considering conventional antitrust principles when acting in the realm of Big Data, highlighting data as a competitive necessity in today's world.

The dispute in that case arose when professional social networking site LinkedIn attempted to stop hiQ, an HR data analytics company, from scraping publicly available user data from LinkedIn's site. Id. at *1-*2. hiQ's business model is based on analyzing this scraped public data to provide its client businesses with more accurate information about its workforce, including workforce skillsets and the likelihood any of its workers are actively looking for a new job. In May 2017, LinkedIn sent hiQ a letter demanding it cease automatically collecting data from LinkedIn public profiles. LinkedIn claimed the scraping violated the CFAA, the Digital Millennium Copyright Act, and state trespass law. LinkedIn also explained it had implemented technical procedures to block hiQ from accessing its data. hiQ then filed suit for a declaration that it has not and will not violate laws through its scraping. hiQ also asserted affirmative rights of access to the publicly available data based in part on California's Unfair Competition Law, and eventually moved for a preliminary injunction to maintain its access.

Beginning its analysis of the motion for a preliminary injunction, the court found that the balance of hardships heavily favored hiQ. See id. at *2-4. hiQ's business model is based on the use of the publicly available LinkedIn data. Simply put, hiQ cannot exist without LinkedIn access. LinkedIn's hardship claim, on the other hand, relied on asserting privacy concerns for its users, arguing that the integrity of its privacy policy—including a "do not broadcast" feature employed by some users to prevent LinkedIn from alerting other users about profile updates or changes—is threatened by hiQ's actions. This claim did not sway the court, as it found that users could apply the "do not broadcast" feature for a number of reasons other than privacy concerns and the fine-print privacy policy was unlikely to reflect actual user privacy expectations for a public online profile. And the court notably seemed suspicious of LinkedIn's concern for user privacy, explaining that LinkedIn has even championed its own "Recruiter" product—a nascent competitor of hiQ—as able to track any user's LinkedIn activity.

Before assessing hiQ's likelihood of success, the court needed to address LinkedIn's contention that the CFAA condemned hiQ's scraping regardless of an alleged state right to acess. Id. at *4-*8. The CFAA creates federal civil and criminal liability for anyone who "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) Other recent cases applying this "unauthorized access" language had found that parties accessing computers after an express revocation of permission violated the CFAA. See e.g., *Facebook v. Power Ventures*, 844 F.3d 1058 (9th Cir. 2016); *United States v. Nosal*, 844 F.3d 1024 (9th Cir. 2016); *Craigslist v. 3Taps*, 942 F. Supp. 2d 962 (N.D. Cal. 2013). But the court distinguished hiQ's conduct here—most of those cases had involved unauthorized access of password-protected information or computers, while hiQ was simply accessing otherwise publicly-available information hosted by LinkedIn servers. The court also reasoned that outlawing such access would "effectuat[e] the digital equivalence of Medusa," turning defendants to stone for viewing public information. hiQ Labs, 2017 WL 3473663 at *6. This could lead to perverse consequences, effectively allowing websites to criminalize access to its public site on the basis of discrimination or anticompetitive intent. The court found that proscribing unauthorized access to otherwise public information could not have been what Congress intended when enacting the CFAA to address computer "trespass," and thus concluded it seriously doubted hiQ was violating the Act.

Turning to hiQ's claims of an affirmative right to access LinkedIn's publicly-available profiles, the court first analyzed hiQ's argument that LinkedIn violated hiQ's California constitutional free speech protections by blocking its site access. Id. at *10-*12. California law ensures that free expression can take precedence over the rights of private property owners, and hiQ argued that LinkedIn is the type of publicly-accessible forum where these rights of free expression should be protected. The court refused to make such a leap, finding the analogy "imperfect" and lacking sufficient precedent to support hiQ's assertions. Separately, the court also found no basis for hiQ's claim that

LinkedIn should be estopped from denying it access to the site, on the basis it "promised" its users their information would be truly public to all.

But the court ultimately did find that hiQ had raised sufficiently serious issues on the merits of at least one claim—its California UCL claim—and granted the preliminary injunction. Id. at *11-*12. California's UCL "broadly prohibits any 'unlawful, unfair or fraudulent business act or practices.'" Id. at *11 (quoting Cal. Bus. & Prof. Code §§17200 et seq.). Unfair practices under the UCL are not limited to actions that would violate the federal antitrust laws, but include any conduct that "violates the policy or spirit" of an antitrust law. Id. The court agreed that hiQ had raised serious issues on its claim that LinkedIn had violated the spirit of the antitrust laws by (1) unfairly leveraging dominance in the professional networking market to gain an anticompetitive advantage in the data analytics market, and (2) denying hiQ an "essential facility" for competitive viability in the data analytics market.

The court noted that hiQ had plausibly asserted that LinkedIn is the dominant player in the professional networking market. The court then again detailed LinkedIn's new "Recruiter" product—a direct competitor of hiQ's services that LinkedIn released around the same time it cut off hiQ's access to LinkedIn's public data. LinkedIn's dominance coupled with this recent expansion convinced the court that hiQ had established a plausible inference that LinkedIn's actions were motivated by an intent to eliminate hiQ as a competitor in the professional data analytics market. And LinkedIn's own actions undermined its claims that it was only acting in the interest of its users' privacy: Its Recruiter product made the same data available to third parties and LinkedIn has even claimed in previous litigation a right to harvest information which its users choose to make public.

One can expect dominant social media platforms to continue to push the boundaries on allowable methods to limit data scraping (e.g., CAPTCHA systems, log-in requirements, rate limiting, etc.).

Eventually, of course, the court could conclude that hiQ has failed to prove its antitrust-styled claims under the UCL, or that LinkedIn truly is merely acting out of concern for its users' privacy. But permitting the preliminary injunction to issue at all—let alone based on antitrust concerns—is a dramatic change of course in dealing with information on major social networks. In *Craigslist v. 3Taps*, 942 F. Supp. 2d 962 (N.D. Cal. 2013), for example, Craigslist convinced the court that companies scraping data from its public listings after receiving cease-and-desist letters were accessing its system without authorization in violation of the CFAA.[4] And just last year, the Ninth Circuit in *Facebook v. Power Ventures*, 844 F.3d 1058 (9th Cir. 2016), concluded that scraping data from Facebook profiles with consent from users—but not Facebook itself—constituted a CFAA violation. To be sure, successful antitrust claims against the major social network players have generally been few and far between. One can also expect dominant social media platforms to continue to push the boundaries on allowable methods to limit data scraping (e.g., CAPTCHA systems, log-in requirements, rate limiting, etc.), but with a cert petition in *Power Ventures* currently before the Supreme Court, the *hiQ* opinion has the potential to be a game-changer in the Big Data field—and at the very least something for practitioners to keep their eyes on no matter who they represent in the e-commerce space.

1. Edith Ramirez, "Keynote Remarks of 43rd Annual Conference on International Antitrust Law and Policy, Fordham Competition Law Institute: Deconstructing the Antitrust Implications of Big Data" (Sept. 22, 2016).

2 Margrethe Vestager, Speech: "EDPS-BEUC Conference on Big Data" (Sept. 29, 2016).

3 European Commission, "Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service" (June 27, 2017).

4. Skadden Arps represented 3Taps in the matter.