Hedge Fund Law Report
An Acuris company

hflawreport.com

January 21, 2021

TECHNOLOGY

To Scrape or Not to Scrape: The Potential Legal Implications of Using Web Scraping for Market Research

By Douglas A. Rappaport, Peter I. Altman and Kelly Handschumacher, *Akin Gump Strauss Hauer & Feld LLP*

Through the use of automated processes performed by software, a web scraper visits a website and attempts to gather relevant data or information that may be provided on the site, such as consumer product reviews or social media profile data. That information may be readily accessible through a simple Google search, or it may require access through a click-through terms-of-service agreement or a firewall. Some investment advisers interested in web scraping conduct that activity in-house, while others may look to outside vendors to accumulate the information.

The law regarding web scraping, however, is still developing and implicates a large number of statutory regimes and areas of common law. For example, web-scraping activity may implicate federal statutes, such as the Computer Fraud and Abuse Act (CFAA), Digital Millennium Copyright Act (DMCA) and insider trading laws; state blue sky laws; privacy laws; and common law claims, such as breach of contract, fraud and trespass to chattels. This article provides an overview of the evolving area of web-scraping law and practical guidance to investment advisers considering web scraping.

Potential Sources of Liability for Unlawful Web Scraping

Computer Fraud and Abuse Act

The CFAA is a criminal statute that also provides a private right of action that is commonly invoked in web-scraping cases. Among other possible violations, the CFAA proscribes "intentionally access[ing] a computer without authorization or exceed[ing] authorized access, and thereby obtain[ing] . . . information from any protected computer." Courts have disagreed, however, on what constitutes access without authorization or exceeding authorization.

In *hiQ Labs, Inc. v. LinkedIn Corp.* (*hiQ*), the Ninth Circuit affirmed the district court's order granting a preliminary injunction barring LinkedIn from blocking hiQ from accessing and scraping information from publicly available LinkedIn member profiles for use in hiQ's data-analysis products. LinkedIn did not require a password or other authentication to access the public profile data. LinkedIn's terms of use, however, prohibited web scraping, and LinkedIn sent a cease-and-desist letter to hiQ demanding that it stop.

The Ninth Circuit held that hiQ had shown a likelihood of success on the merits of its claim that a user's act of accessing data made available by the owner to the general public does not constitute access "without authorization" under the CFAA. When reaching its decision, the court opined that a person may violate the CFAA's prohibition on accessing a computer "without authorization" when he or she "circumvents a computer's generally applicable rules regarding access permissions, such as username and password requirements, to gain access to a computer." The court, however, further held that it is "likely that when a computer network generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA." LinkedIn filed a petition for a writ of certiorari in March 2020, which is currently pending before the U.S. Supreme Court.

In a previous litigation under the CFAA – Facebook, Inc. v. Power Ventures, Inc. (*Power Ventures*) – the Ninth Circuit held that a user accesses a computer "without authorization" under that statute when he or she continues to circumvent technological measures employed by the operator to block that user's access. *Power Ventures* involved the blocking of a user's IP address to prevent access to password-protected information after issuing a cease-and-desist letter. In contrast to *hiQ*, the court in *Power Ventures* found that Facebook had "'tried to limit and control access to its website' as to the purposes for which [defendant] sought to use it," specifying that "Facebook require[d] its users to register with a unique username and password." The court emphasized that, although the defendant in *Power Ventures* was "gathering user data that was protected by Facebook's username and password authentication system, the data hiQ was scraping was available to anyone with a web browser."

Although *hiQ* and other litigation out of the Ninth Circuit have dominated the recent headlines in the space, the U.S. Supreme Court has agreed to hear a CFAA case that will likely impact the law of web scraping. Specifically, the Supreme Court granted certiorari in April 2020 in U.S. v. Van Buren (*Van Buren*), an Eleventh Circuit decision that addressed the question of whether a person who is authorized to access information on a computer for certain purposes violates the CFAA if he or she accesses the same information for an improper purpose. Van Buren was a Georgia police officer who was convicted of violating the CFAA and honest-services wire fraud based on his use of the Georgia Crime Information Center (GCIC) database to obtain information on a particular person in exchange for money.

On appeal, Van Buren argued that he did not exceed authorized access within the meaning of the CFAA because he was authorized as a police officer to access the GCIC database and distinguished his improper use of the GCIC database from his legitimate right to access it. The Eleventh Circuit affirmed Van Buren's CFAA conviction, arguably creating a split with the Ninth and Second Circuits over the interpretation of the statute's provision regarding conduct that "exceeds authorized access." The Supreme Court's decision in *Van Buren* could bear on whether scraping data that one is authorized to access for certain purposes – such as browsing as a potential customer or participating as a member of a social media network – but not authorized to access for web-scraping purposes, constitutes a breach of the CFAA.

In another important 2020 decision, a D.C. federal district court held that the CFAA's access provision did not criminalize the violation of a consumer website's terms of service. In *Sandvig v. Barr*, academic researchers brought a pre-enforcement challenge, alleging that the CFAA violated the First Amendment right to free speech by criminalizing certain terms-of-service violations related to employment websites. On the parties' cross-motions for summary judgment, the court held that it need not address the First Amendment issue because the CFAA did not criminalize the violation of a consumer website's terms of service. In so holding, the court observed that the majority of courts that have examined whether violating the terms of service of consumer websites constitutes a criminal CFAA violation have found no liability.

In short, although the scope of the CFAA's access provision is unsettled, significant authority suggests that the scraping of publicly available information, such as from LinkedIn member profiles, does not violate the CFAA. Likewise, it suggests that violation of a website's terms of use alone, without more, may not violate the CFAA.

Copyright/DMCA

In addition, the operator of a website that is the target of web scraping may bring a claim for copyright infringement against the user of the web-scraping device by proving:

1. its ownership of a valid copyright; and
2. the user's copying of the original elements of the work in question.

Copyrightable work is further protected by the DMCA, which provides that "[n]o person shall circumvent a technological measure that effectively controls access to a work protected under this title." At least one federal court has held that a party faces liability under Section 1201(a)(1)(A) of the DMCA when it uses bots to circumvent security measures that control nonhuman access to copyrighted material on a webpage.

It is also worth noting the general copyright principle that, although compilations of facts can be protected by copyright, authors may not copyright their ideas or the facts they narrate. Accordingly, if the data scraped are purely facts without a creative component, then there is no copyright claim.

Privacy Statutes

Web scraping may also implicate the privacy statutes of states and other jurisdictions. For example, the E.U.'s General Data Protection Regulation and the California Consumer Privacy Act of 2018 grant consumers a variety of rights and protections with respect to their personal information. Web-scraping activity that compiles personally identifiable information could implicate a variety of privacy statutes – and potentially subject a web scraper to government and private litigation.

See our two-part series on the GDPR: "Impact" (Feb. 21, 2019); and "Compliance" (Feb. 28, 2019). See also "A Roadmap to Understanding and Complying With the California Consumer Privacy Act" (Nov. 14, 2019).

Insider Trading

Under certain circumstances, web scraping could also potentially violate federal insider trading law or state blue sky laws. For example, using affirmative misrepresentations to obtain

material nonpublic information through web scraping and then trading based on that information could potentially constitute insider trading.

In addition, the Second Circuit held in SEC *v.* Dorozhko that trading on information obtained through computer hacking could be insider trading if the information was obtained by "deceptive" means, such as misrepresenting one's identity. For more on that case, see "Tips and Warnings for Navigating the Big Data Minefield" (Jul. 13, 2017).

The law in this area is unsettled, and it remains to be seen how strict an approach regulators and law enforcement may take when deciding what constitutes a breach of duty or deception in the web-scraping context.

See "Best Practices for Using Alternative Data: Mitigating Regulatory and Other Risks (Part Two of Two)" (Feb. 13, 2020); and our two-part series on mitigating insider trading risk: "Relevant Laws and Regulations; Internal Controls; Restricted Lists; Confidentiality Agreements; Personal Trading; Testing; and Training" (Sep. 27, 2018); and "Expert Networks, Political Intelligence, Meetings With Management, Data Rooms, Information Barriers and Office Sharing" (Oct. 11, 2018).

Common Law Claims

In addition to the boundaries imposed by the statutes discussed above, a plaintiff could seek to invoke various common law remedies in an attempt to stem or curtail web scraping. For instance, some website operators have attempted to assert claims for breach of contract against alleged web scrapers.

Courts, however, have held that defendants must be on notice of a website's terms of service for the terms to be enforced against them. For this reason, a "clickwrap" agreement to the terms of service – requiring the user to consent to terms through an affirmative click before being granted access – is more likely to give rise to an enforceable contract than a "browsewrap" agreement, where a link to the terms of use is posted on the site and user consent is merely implied by continued use of the website.

Other potential claims include fraud and unjust enrichment. Website operators could potentially argue that web-scraping activity that simulates organic human use of the website or takes deceptive actions to avoid web-scraper restrictions is a fraudulent misrepresentation on which the user of the web scraper intends the website to rely. In addition, when there is no enforceable contract that governs, a website might bring a claim against a web scraper for unjust enrichment, arguing that the user of a web scraper is unjustly profiting from data extracted from the website to the website's detriment.

Finally, aggrieved website operators have asserted claims for trespass to chattels as a possible means to curtail web scraping. Although courts have held that trespass to chattels is a potentially viable theory for recovering from injury due to web scraping, they have generally stressed the requirement of showing physical harm to the host computer or computer network (the chattel). For example, in *hi*Q, the court said that "it may be that web scraping exceeding the scope of the website owner's consent gives rise to a common law tort claim for trespass to chattels, at least when it causes demonstrable harm." At least one court – in *Ticketmaster Corp. v. Tickets.*

Com, Inc. – has held that the use of a web scraper to gather information from a public website, without more, was insufficient to show the physical injury or dispossession of the host computer or computer network required in a trespass action. On the other hand, the Second Circuit has held that a threat of irreparable injury to computer servers existed where defendant's web-scraping robots "consumed a significant portion of the capacity of [a website's] computer systems."

See "A Fund Manager's Roadmap to Big Data: MNPI, Web Scraping and Data Quality (Part Two of Three)" (Jan. 18, 2018).

Advice for Investment Managers That May Engage in Web Scraping

On the whole, the law on web scraping is still developing, and only further court decisions and legal pronouncements will thoroughly define its parameters. In light of the evolving legal landscape, the following is a high-level, non-exhaustive list of practical takeaways for investment managers that may engage in web scraping:

- Consider whether any web-scraping practices are potentially deceptive. Do not use any techniques or strategies that could be perceived to constitute affirmative misrepresentations.
- Review the website's terms of use and robot.txt files before consenting to web-scraping data collection activity.
- Monitor and consider any actions a website takes to restrict web scraping, such as the use of CAPTCHAs, rate limits and blocking of IP addresses.
- Consider the appropriateness of IP structure and use of names and passwords.
- Avoid adversely impacting a website's physical operation, which could lead to a claim for trespass to chattels or similar claims.
- Avoid impacting the availability of inventory for goods or services to customers.
- Avoid collecting personally identifiable information.
- Consider whether any data to be scraped is protected by copyright.
- Be prepared to stop if asked to do so through a cease-and-desist letter or otherwise.
- Conduct thorough due diligence on any data vendors regarding their web-scraping practices.
- Stay up to date on evolving law in this area.

See "Report Calls for Transparency and Development of Standards in the Alternative Data Market" (Apr. 30, 2020); and "Best Practices for Private Fund Advisers to Manage the Risks of Big Data and Web Scraping" (Jun. 15, 2017).

Douglas A. Rappaport is a partner in the New York office of Akin Gump. His practice focuses on counseling investment funds and other clients on a broad range of regulatory, securities, compliance and general business issues; litigating complex commercial and securities disputes; and assisting clients with regulatory investigations and reviews.

Peter I. Altman is a partner in the Los Angeles office of Akin Gump. He handles white collar and other enforcement and regulatory matters;

securities class action litigation; and internal investigations. His clients include investment management firms; public and private companies; and individuals. Altman served as Senior Counsel in the SEC's Division of Enforcement in Los Angeles, as well as a member of the Division's selective Market Abuse Unit.

Kelly Handschumacher is an associate in the New York office of Akin Gump. She represents public and private companies, as well as investment funds. She also represents individuals in government investigations.