THE NATIONAL LAW JOURNAL

NOVEMBER 2018

An ALM Publication

Practice

# Internet Data Scraping 201: A Refresher on the Basics

BY ANTHONY J. DREYER AND ANDREW GREEN

**AS THE AMOUNT (AND VALUE) OF ONLINE DATA** continues to grow exponentially, so does the practice of internet data scraping—that is, the harvesting of data from third-party websites for commercial purposes. Because scraping activity, and the efforts to stop it, have continued apace, it's necessary to stay refreshed on the topic.

## COPYRIGHT AND THE DMCA

Internet data often is protected by copyright, leading website owners to contend that scraping constitutes infringement and/or violation of the Digital Millennium Copyright Act. In recent years, the DMCA—which prohibits circumventing technological measures that effectively control access to a copyrighted work—has become a more popular enforcement tool; unlike infringement, DMCA plaintiffs need not own or hold an exclusive license to the copyrighted works at issue. To state a DMCA violation, plaintiffs generally must allege that they implemented technological barriers within their website, but a scraper evaded those barriers by some technological workaround.

For example, in *Ticketmaster v. Prestige Entertainment*, a DMCA claim was properly alleged where the defendant used automated software "bots" to bypass CAPTCHA controls and purchase tickets. By contrast, in *CouponCabin v. Savings.com* in the U.S. District Court for the Northern District of Indiana in 2016, a DMCA claim was dismissed where access to a website did not require the application of "information or a process or treatment," such as a password. Under such circumstances, merely blocking defendant's servers—prompting defendant to utilize different servers for scraping—did not support a DMCA claim because the website remained available to any server that had not been specifically blocked.

TIPPAPATT/SHUTTERSTOCK.COM

## COMPUTER FRAUD AND ABUSE ACT

Web content owners also have asserted claims under the CFAA, which prohibits obtaining information from a protected computer and causing damage by (i) intentionally accessing the computer without authorization or (ii) exceeding authorized access. CFAA claims generally proceed where a website owner affirmatively rescinded a defendant's authorization to access its website, but the defendant nevertheless continued scraping the site.

In *Facebook v. Power Ventures*, for example, a CFAA violation was affirmed with respect to the scraping of data after Facebook had sent the defendant a cease-and-desist letter and attempted to block future access to Facebook's website; the scraping of Facebook's website prior to the express rescission of permission, however, was not "without authorization" for CFAA purposes.

Other recent cases suggest that the CFAA does not prohibit scraping from publicly available portions of a website, because the scraping from a generally accessible website is merely a particular use of information that users otherwise are entitled to see. In *hiQ Labs v. LinkedIn*, a data analytics company obtained a preliminary injunction against attempts to block it from accessing and scraping publicly available user data from LinkedIn's website. In granting the injunction, the court contrasted hiQ's scraping of publicly available information with password-protected data in *Facebook*.

### TRESPASS TO CHATTELS

Unauthorized access to a computer system also can give rise to a common-law claim for trespass to chattels. Determination of whether access is unauthorized generally is the same as for a CFAA claim, although trespass claims are not limited to the scraping of private, password-protected information. Trespass claims often turn on whether scraping caused actual damages, such as impairing a website's functionality. This is a fact-dependent inquiry that often is not resolved at the motion to dismiss stage.

### BREACH OF CONTRACT

Because scraping often violates a website's terms of use, breach of contract is another common claim in this area. Terms of use typically are conveyed to website users through either a "clickwrap" or "browsewrap" agreement, with the former generally being enforceable and the latter's enforceability often dependent upon a fact-specific inquiry about the location, accessibility and defendant's awareness of the terms. Plaintiffs seeking to enforce a browsewrap agreement typically must demonstrate that the user had actual or constructive knowledge of the terms.

At least one court has looked to a data scraper's practices on its own website to enforce a browsewrap agreement against the scraper. In *DHI Group v. Kent*, a breach of contract claim survived a motion to dismiss because defendant's operation of "a similar site with a similar browsewrap agreement" constituted constructive notice of plaintiff's terms. This finding of notice was limited, however, to instances where "both parties are sophisticated businesses that use browsewrap agreements."

### ADDITIONAL OBSERVATIONS

Additionally, statutory claims may be available under applicable state law. (E.g., Texas' Harmful Access by a Computer Act; California's Unfair Competition Law.) For their part, some data scrapers are fighting back by challenging attempts to block or restrict their access to websites as tortious interference (*Fidlar*) or violating antitrust laws.

Ultimately, parties that engage in scraping should ensure their activities are consistent with website authorizations and terms of use. Website owners, in turn, should have terms of use—in clickwrap, where possible—that clearly notify third parties of prohibited scraping and/or circumvention activities. For an additional layer of protection, website owners should place their most valuable data behind password protections.

*Anthony J. Dreyer is a partner, and Andrew Green is an associate, with Skadden, Arps, Slate, Meagher & Flom.*