SEPTEMBER 11, 2017

Legaltech news

ONLINE

An ALM Publication

The Great Scrape: How Website Owners and Data Scrapers Can Avoid Litigation

The proliferation of data scraping increases the chance of litigation, which means both sides should become educated on ways to avoid it.

BY PAVEN MALHOTRA AND ANDREA NILL SANCHEZ, KEKER, VAN NEST & PETERS

Data scraping—extracting large amounts of information from a website using automated software programs called "bots"—has been a growing subject of costly litigation. But it doesn't have to be. The first article of this two-part series outlined the legal claims companies have brought against scrapers. This second part will identify the measures that websites can implement to discourage data scraping and the precautions scrapers can take to avoid litigation.

Website Owners

The best preventative measure a website owner can take against data scraping is to insist its users affirmatively agree not to engage in the practice. This can be accomplished through a "clickwrap" agreement, which requires website visitors to check a box assenting to terms of use before being able to access any content. Website users who chose to violate the agreement can be held legally accountable for breaking their promise not to scrape.

"Browsewrap" agreements that only appear somewhere on the website are much less reliable. Several courts considering data scraping claims have held that a website user must have had actual or constructive knowledge of the site's conditions. If a website neither prompts users to review its terms of use nor prominently displays them, it will be difficult to establish either of those requirements.

Companies can also adopt technical measures to tell users that they do not want their information to be scraped.

One such measure is the robots.txt protocol, a piece of code embedded in a webpage that tells bots which portions of the website should and should not be accessed. Scrapers can still choose to ignore the protocol's instructions, but many will abide by the robots.txt restrictions.

Having a robots.txt specification is also important because some courts will interpret the protocol's absence to constitute an implied license for scrapers. In *QVC, Inc. v. Resultly*, for example, QVC lost its case against start-up Resultly in part because it failed to set up a robots.txt crawl rate specification. Even though Resultly's scraping ended up crashing QVC's servers, the court reasoned that "Resultly crawled the QVC website in the same manner as it crawled any other website that did not provide a robots.txt file specifying a crawl delay." Although not all courts will assume that the lack of a robots.txt protocol excluding web scrapers authorizes website access, it never hurts to have one.

Finally, companies can take steps to block scrapers altogether. One option is to block the IP addresses of known scrapers. Another, more aggressive action is to issue cease and desist letters to scrapers explicitly demanding that they stop extracting information from the company's website. If a scraper actively skirts these restrictions, it serves as compelling evidence of legal wrongdoing.

Data Scrapers

Conversely, there are a number of measures that data scrapers can take to avoid litigation and mitigate their liability.

For starters, data scrapers should strive to benefit consumers and promote access to information. While using scraping tools to hack sensitive financial and personal data is obviously illegal, scrapers can also get in trouble for scraping information as part of a scheme to hinder competition.

Ultimately, even data scrapers with noble purposes may face legal liability. The best way to avoid the threat of litigation is to respect a website's efforts to deter data scraping. This means abiding by a website's terms of use and following the robots.txt specifications. Certainly if a company escalates a situation by blocking IP addresses or serving cease and desist letters, a responsible data scraper should immediately back off.

Data scrapers are also better off if they temper the scale of their activities. Otherwise, overzealous data scrapers who try to retrieve as much data as possible as quickly as possible risk compromising a website's functionality, which could translate to concrete legal harm.

For example, in *eBay v. Bidder's Edge*, eBay obtained preliminary injunctive relief against Bidder's Edge, an auction aggregation company that accessed eBay's site approximately 100,000 times a day. The court determined that Bidder's Edge used "valuable bandwidth and capacity … necessarily compromising eBay's ability to use that capacity for its own purposes." The court also noted that if Bidder's Edge activity was left unchecked, it would only encourage others to engage in similar conduct that would ultimately reduce eBay's system performance and potentially lead to system unavailability or data loss.

The growing proliferation of data scraping will undoubtedly increase litigation around the practice. But the more steps companies take to guard their data, the less likely they are to get scraped in the first place and the more likely they are to win in court if it does happen.