Proskauer

Court Rules That Scraping of Public Data by Competitor Constitutes Trade Secret Misappropriation

New Media and Technology Law Blog on **September 16, 2024**

In an ongoing dispute commenced in 2016, the Eleventh Circuit for the second time in the lifetime of the litigation considered trade secret misappropriation and related copyright claims in a scraping case between direct competitors.

The case involved plaintiff Compulife Software, Inc. ("Plaintiff" or "Compulife") – in the business of generating life insurance quotes on the internet – and a group of Compulife competitors and others ("Defendants") who allegedly misappropriated Plaintiff's admittedly publicly available insurance quotes.

The appeals court affirmed the lower court judgment finding the Defendants liable for trade secret misappropriation when they orchestrated a "scraping attack" of Compulife's website that acquired millions of variable-dependent insurance quotes in a usable format to set up a competing service. (*Compulife Software, Inc. v. Newman*, No. 21-14074 (11th Cir. Aug. 1, 2024)). The decision affirmed the lower court holding that Defendants acquired Compulife's trade secret via "improper means" when it scraped so much of the database from Compulife's site that they posed a competitive threat to Compulife. In doing so, the court acknowledged that manually accessing the life insurance quote information from the Plaintiff's publicly web-accessible database would generally not constitute the improper acquisition of trade secret information. The court noted, however, that "even if individual quotes that are publicly available lack trade secret status, the whole compilation of them (which would be nearly impossible for a human to obtain through the website without scraping) can still be a trade secret."

Importantly, the court narrowed the holding of its decision carefully, to avoid casting a taint on the general practice of scraping:

> “Here, Compulife alleges that the defendants used ‘scraping’ to acquire its trade secret—i.e., its database of insurance quotes. It is important to note that scraping and related technologies (like crawling) may be *perfectly legitimate*. […] But the defendants in this case did not take innocent screenshots of a publicly available site; instead, they copied the order of Compulife’s copyrighted code and used that code to commit a scraping attack that acquired millions of variable-dependent insurance quotes. If they had not formatted and ordered their code exactly as Compulife did, they would not have been able to get the millions of quotes that they got.”

In another part of the ruling, the circuit court also reversed the lower court’s dismissal of Plaintiff’s copyright claims based on Defendant’s unauthorized access and copying of code from Plaintiff’s licensed insurance quote software, remanding the claim for reconsideration of whether Plaintiff’s “arrangement” of the data elements in its source code was protectable.

The decision is an important addition to the emerging case law around scraping for at several reasons: (1) there are very few cases where viable trade secrets claims are pled following instances of data scraping; (2) the Circuit Court entertained the argument over whether the plaintiff’s “arrangement” of the data elements in its source code was protectable; and (3) the court found culpability for a group of defendants despite various levels of participation in the activity at issue. But what makes this holding particularly remarkable is the finding of a trade secret misappropriation, despite the public nature of the data being scraped.

We've previously written extensively about the landmark Ninth Circuit opinion in the *hiQ* case, where the appeals court found that hiQ "raised at least serious questions" that its scraping of public LinkedIn member profile data, even after having had its access revoked and blocked by LinkedIn, does not violate the CFAA. Despite the pro-scraping posture of that opinion, we advised then that even though the path for data scraping involving public websites may have cleared considerably with respect to the CFAA, it is by no means an open road. This latest Eleventh Circuit decision highlights this cautionary principle, with the court declining to foreclose a finding of trade secret theft merely because Compulife's site was publicly accessible, but instead centering its analysis on how much of the database itself Defendants scraped and how that amounted to a protected portion of Compulife's trade secret that could not have feasibly been compiled without a bot. This complemented the lower court's reasoning that the open nature of Compulife's site wasn't necessarily determinative in this case: "I reject Defendants' argument that Compulife cannot establish misappropriation due to its failure to restrict use at Term4Sale.com prior to the scraping attack."

The facts of this case, discussed in greater detail below, may at times read like a corporate espionage thriller and the holding is certainly tied to the circumstances of the dispute and the nature of the parties' relationship, yet the *Compulife* decision changes the calculus for certain scraping activities of public websites in the future and should be evaluated along with other reasoned business and legal decisions.

Background and Proceedings

The fact pattern of this case is extensive and the case has been there and back again from the Eleventh Circuit. As we outlined in our prior post on this case,Compulife and the defendants are direct competitors in the niche industry of generating life-insurance quotes. Compulife obtains monthly rate tables from insurance companies (often ahead of their public release), but does not own this information.. It compiles this information into a database in a confidential manner and encrypts the database to prevent reverse engineering. Compulife then licenses access to the database for a fee under various licensing schemes, including allowing licensees to embed a "web quoter" on their websites for customers (which pulls data from a remote server). In addition, Compulife maintains a website at www.term4sale.com that allows visitors to obtain life insurance quotes at no cost. Term4Sale.com generates life insurance quotes using Compulife's web-based HTML code, host-based software, and database of information.

Plaintiffs eventually discovered that Defendants had, allegedly under false pretenses, obtained access to and copied Compulife's HTML code from its web-based software onto their own websites, even though such parties did not have Compulife's permission or authority to access Compulife's database or copy Compulife's HTML code. Subsequently, the Defendants hired a third party to scrape data from Compulife's Term4Sale website. In order to produce the proper commands for the scraping, portions of Compulife's HTML code were required—specifically, the correct names for certain database variables—without which the commands would not have provided accurate results. This third party individual purportedly created a partial copy of Compulife's database. While a human user could permissibly enter query after query into Compulife's database to generate quotes, the bot was able to scrape millions of quotes. Compulife alleged that the defendants then used the scraped data as the basis for generating quotes on their own websites, resulting in a purported decline in sales for Compulife. The Term4Sale website did not include a user agreement at the time of the scraping, but one was added afterward.

In the 2020 Eleventh Circuit opinion, the appeals court rejected the lower court's holdings following a bench trial that the publicly-available nature of the insurance quotes negated the trade secret claim (Circuit Court: "Even granting that individual quotes themselves are not entitled to protection as trade secrets, the magistrate judge failed to consider the important possibility that so much of the Transformative Database was taken—in a bit-by-bit fashion—that a protected portion of the trade secret was acquired"). Furthermore, the court reasoned that the nature of scraping itself, even on a publicly available database, may yet constitute an "improper means" to acquire trade secret information under state trade secret law ("So, while manually accessing quotes from Compulife's database is unlikely ever to constitute improper means, using a bot to collect an otherwise infeasible amount of data may well be…"). Ultimately, back in 2020 the appeals court remanded for a trial on Compulife's claims for copyright infringement and misappropriation of trade secrets.

After a second bench trial, the Florida district court ruled in favor of Compulife on the trade secret claim, but dismissed the copyright infringement claim, finding that most of Compulife's code was not protectable, and that the protectable parts were not substantially similar to the Defendants' code. Based on those rulings, the district court granted Compulife injunctive relief and entered judgment and an award of damages against all defendants jointly and severally. Both parties appealed.

Copyright Claim

While Compulife argued that its HTML code is "original creative authorship entitled to copyright protection" and the *arrangement* of the various components of the code (e.g., insurance data classifications such as state, birthday, insurance type) is creative and protectable, Defendants countered that its copying of Compulife's HTML code was not actionable as the merger doctrine makes such common organizational elements non-copyrightable. The lower court found that while Compulife presented evidence that Defendants factually copied the variables and parameters from its HTML code, Defendants succeeded in proving that the majority of the program's copied elements were unprotectable and that the remaining protectable portions of the copied code were not qualitatively significant to advance a cognizable infringement claim. On appeal, the Eleventh Circuit reversed the lower court and remanded for reconsideration of the copyright claim stating that the "arrangement" of the code may be protectable and the lower court erred in failing to consider the copyrightability of the code's arrangement when it examined which parts of the copied code were protectable (and could constitute a claim for actionable copying) and which parts were not.

Trade Secret/Scraping Claims

The appeals court agreed with the lower court's ruling on this claim:

> "A fair reading of this record supports the district court's finding that the defendants obtained so much of the database that they posed a competitive threat to Compulife. Faced with this record, the district court did not err in concluding that Compulife's trade secret was acquired by its competitors.

Final Considerations

Taken together, the Eleventh Circuit has laid out an intriguing theory of trade secret misappropriation via scraping. The latest ruling doubles down on its initial 2020 holding that recognized the individual Compulife quotes themselves are not entitled to protection as trade secrets because they are publicly available, but that the sheer amount of data taken amounted to a protected portion of Compulife's trade secret and posed a competitive threat to Compulife. It's a concept reminiscent of a copyright in a compilation, where an original and independent selection, coordination and arrangement of data might deserve copyright protection, even if the underlying facts are not protectable. The court's conclusion is almost intuitive – at some point, the amount of data scraped from a proprietary database open for ordinary public use greatly diverges from what a human could retrieve through ordinary clicks and reaches such a point where a court might consider such actions to be actionable misappropriation. Still, this is an evolving area of law and another court may decide that the public availability of certain data would preclude a finding that scraping such a data compilation constitutes misappropriation.

View original.

Related Professionals

- **Jeffrey D. Neuburger**