New York Law Journal

A NEW YORK LAW JOURNAL SPECIAL SECTION

Litigation

An ALM Publication

WWW.NYLJ.COM

MONDAY, JULY 15, 2013

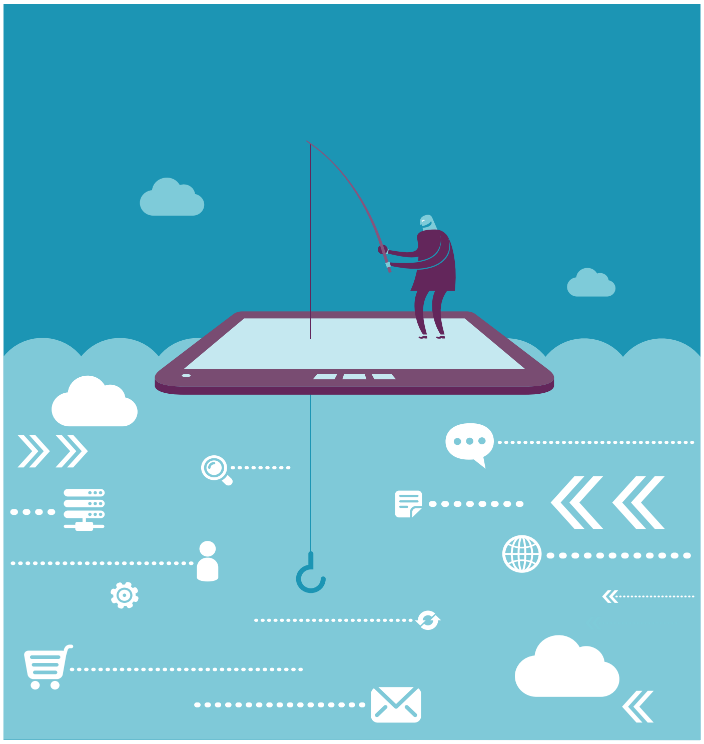
Internet 'Data Scraping'

A Primer for Counseling Clients

BY ANTHONY J. DREYER
AND JAMIE STOCKTON

The proliferation of Internet access and mobile devices has led to an exponential explosion of content on the Web, creating a vast repository of "publicly available" information. This includes not only news, business, and financial information, but also personal data, movie and restaurant reviews, concert ticket sales, flight information, and a virtually endless array of other categories. This same technological explosion, however, has made it far easier for third parties to extract this data for commercial sale and use—and to do so for free and without authorization. This data extraction, commonly referred to as "scraping," "crawling," or "spidering" (collectively "scraping"),[1] creates legal issues and concerns for both sides of this issue—those who want to scrape, and those who want to protect against scraping of their websites.

This article provides a primer on the legal framework surrounding scraping, addressing both the grounds for potential claims against scrapers, and ways to avoid liability for scraping. The common theories of liability arising from scraping are copyright infringement, trespass to chattels, breach of contract, and violation of the Computer Fraud and Abuse Act (CFAA). This article discusses the leading cases applying these legal theories to website scraping, and concludes that the most effective way to create potential claims against scrapers is through carefully drafted prohibitions in a website's terms of use. Conversely, the most effective way to defend against a claim of unauthorized scraping is to abide by such terms of use, or to establish that scraping constitutes a fair use and does not overburden the servers of the website being scraped.



Copyright Infringement

Scraping inherently involves copying, and therefore one of the most obvious claims against scrapers is copyright infringement. However, such claims are often open to attack on several grounds. First, in order to have standing to bring a claim for copyright infringement, the owner (or exclusive licensee) of the website being scraped must also be the owner of the copyrightable content that is the subject of the claim.[2] This can pose a barrier to bringing a lawsuit if, for example, the content at issue is user-generated (such as videos or reviews), and the rights in the

ANTHONY J. DREYER *is a partner, and* JAMIE STOCKTON *is an associate, with Skadden, Arps, Slate, Meagher & Flom.* BRITTANY BETTMAN, *a summer associate, assisted in the preparation of this article.*

content have not been transferred to the website owner.

Second, copyright law does not protect ideas, but rather only tangible expression.[3] Thus, the scraping of general factual data does not give rise to a viable claim for copyright infringement. For example, in *Ticketmaster v. Tickets.com*, the court rejected an infringement claim because the material being extracted—factual information regarding concerts and URLs—was not copyrightable.[4]

Third, even if the information copied by the scraper is protectable under copyright law, the defendant may be able to rely upon the "fair use" defense. Under the Copyright Act, courts are to consider the following factors to determine if a use is a fair use: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.[5] For example, in *Kelly v. Arriba Soft*, the court held that the use of scraping software by a search engine to reproduce images in thumbnail form was not a sustainable basis for a claim of copyright infringement, because the thumbnail images created from the full-size scraped images were "transformative" and qualified as a fair use of the images.[6]

Trespass to Chattels

A trespass to chattels is defined as intentionally dispossessing another of a chattel or using or intermeddling with a chattel in the possession of another.[7] This legal theory applies to the Internet inasmuch as a website proprietor has a "fundamental property right to exclude others from its computer system[.]"[8] Moreover, even if a website is publicly accessible, its servers are private property, and the proprietor may therefore grant conditional access to users, including prohibitions against scraping.[9]

For example, in *Bidder's Edge*, the court held that excessive scraping can support a claim for trespass to chattels if it taxes the plaintiff's computer system in such a way that would substantially impair it, and, if so, an injunction may be granted.[10] Specifically, the court held that there was a viable trespass cause of action due to the excessive scraping of eBay's website at the rate of 80,000-100,000 times per day.[11]

Scraping **inherently involves copying**, and therefore one of the most obvious claims against scrapers is **copyright infringement**.

Similarly, in *Register.com v. Verio*, the Court of Appeals for the Second Circuit held that Verio's use of search robots consumed a significant portion of the capacity of Register's computer system, and that Verio was therefore engaged in a trespass.[12] The court reasoned that if it were to allow these queries, then it was "highly probable" that other companies would begin to do the same, which would likely result in Register's system being "overtaxed and [it] would crash."[13] However, in *Ticketmaster*, the court held that the use of scrapers to extract data was not a trespass to chattels, because there was no evidence that the scraping caused any tangible interference with the operation of Ticketmaster's system.[14]

Breach of Contract

Courts have held that a viable method of preventing scraping is to include prohibitions against scraping in the website's terms of use.[15] Such restrictions are generally conveyed to website users through a "clickwrap" or "browsewrap" agreement.

A clickwrap agreement is an online agreement that requires the user to consent to terms and conditions by affirmatively clicking a dialogue box agreeing to the terms before the user can proceed to use a website.[16] Clickwrap agreements are generally enforceable, due to the user's clear manifestation of assent, so long as the terms do not violate other basic contract principles (e.g., unconscionability).[17]

For example, in *Bidder's Edge,* the court took note of the fact that the user agreement at the time, to which users were required to click "I Accept," expressly prohibited "any robot, spider, other automatic device, or manual process to monitor or copy our web pages or the content contained herein without our prior expressed written permission."[18] The court stated that these terms of use constituted a limited license, and that actions not permitted by this license were restricted.[19]

Browsewrap agreements, on the other hand, involve the posting of a link to terms and conditions on a website for users to read, but do not require users to affirmatively manifest assent to the terms and conditions—instead, user consent is implied by continued use of the website.[20]

The enforceability of such agreements requires a fact-specific inquiry, and turns largely upon the location and accessibility of the terms of use.[21] According to the *Specht* court, "[r]easonably conspicuous notice of the existence of contract terms and unambiguous manifestation of assent to those terms by consumers are essential if electronic bargaining is to have integrity and credibility."[22]

For example, in *Hines* the court held that the browsewrap agreement was not enforceable, because in this case the plaintiff had no actual or constructive notice of the terms and conditions of use.[23] However, in *Southwest Airlines v. BoardFirst*, where there was evidence that defendant had actual knowledge of Southwest's terms and conditions, but nevertheless continued to use Southwest's website in violation of those terms, the court held that the browsewrap agreement was an enforceable contract.[24]

Terms of use may also be binding where the terms are reasonably known to the user—even in circumstances in which the terms are not known to the user before the first use of the website. For example, in *Register.com*, the user was made aware of the terms of use only after first accessing the information provided on the website.[25] The court held that while the terms of use were technically neither a clickwrap nor a browsewrap agreement, because they were only displayed after the user accessed the information on the website, the restrictions therein were nevertheless enforceable, because the user accessed the website repeatedly and therefore was on notice during subsequent visits.[26]

In sum, while statements of assent such as "I agree," which are often elicited through clickwrap agreements, are preferable and unequivocally reflect a manifestation of assent, the user need not necessarily state the magic words "I agree" (or some similar formulation).[27] However, "the website user must have had actual or constructive knowledge of the site's terms and conditions, and have manifested assent to them" in some manner, implicit or explicit.[28]

Violation of the CFAA

The CFAA is a federal statute that provides liability for anyone who "intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains…information from any protected computer."[29] The CFAA also requires that there be a minimum amount of damages of at least $5,000 over a one-year period.[30] Similar to the breach of contract cases discussed above, CFAA cases often hinge upon whether a user had actual or constructive knowledge of the restrictive terms of a website's terms of use (i.e., knowledge that the scraping was "unauthorized").

For example, in *Southwest Airlines v. Farechase*, defendants scraped fare, route, and scheduling information from Southwest.com.[31] The court denied a motion to dismiss the CFAA claim because Southwest alleged (i) damages of at least $5,000, and (ii) that it had put defendant on actual notice that scraping was prohibited.[32]

However, in *Cvent*, even though the terms of use stated that competitors were prohibited from accessing and utilizing the information on the website, the court held that there was no violation of the CFAA.[33] The court concluded that the terms of use were not sufficiently visible because the link was "buried" at the bottom of the first page, in extremely fine print, and users had to scroll down to see it, thereby rendering them insufficient protection for the site.[34]

Conclusion and Proposed Terms of Use

In conclusion, scraping may be permissible under U.S. law if the content at issue is not subject to copyright protection, if the scraping does not unduly burden the website's servers, and if the website's terms of use do not prohibit scraping or if assent to such terms has not been manifested.

However, if the client's goal is to reduce or protect against scraping, and to establish a potential basis for liability, the website's terms of use should contain language to the following effect, and users should be put on reasonable notice of such terms. This language is, of course, merely provided as an example:

> By accessing this website, you accept without limitation or qualification, and agree to be bound and abide by, the following terms and conditions (Terms of Use). [CLIENT] may revise and update these Terms of Use from time to time in its sole discretion. Your continued use of this website following the posting of revised Terms of Use means that you accept and agree to any and all changes to the Terms of Use. You may use this website only for lawful purposes and in accordance with these Terms of Use, and you agree not to: (i) use this website in any manner that could disable, overburden, damage, or impair this website, or interfere with any other use of this website, including, but not limited to, any user's ability to engage in real-time activities through this website; (ii) use any robot, spider or other automatic device, process or means to access this website for any purpose, including to monitor or copy any of the material on this website; (iii) use any manual process to monitor or copy any of the material on this website, or to engage in any other unauthorized purpose without the express prior written consent of [CLIENT]; (iv) otherwise use any device, software or routine that interferes with the proper working of this website; or (v) otherwise attempt to interfere with the proper working of this website.

••••••••••••●●●••••••••••••

[1] See *EF Cultural Travel BV v. Zefer*, 318 F.3d 58, 60 (1st Cir. 2003) ("A scraper, also called a 'robot' or 'bot,' is nothing more than a computer program that accesses information contained in a succession of webpages stored on the accessed computer"); *eBay v. Bidder's Edge*, 100 F. Supp. 2d 1058, 1060 (N.D. Cal. 2000). While it is possible to embed instructions on websites that inform the scraping software whether scraping is permitted (called "robot.txt" files), compliance with such instructions is voluntary. See *Bidder's Edge*, 100 F. Supp. 2d at 1061.

[2] See, e.g., *Nautical Solutions Mktg. v. Boats.com*, No. 8:02-CV-760, 2004 WL 783121, at *2-3 (M.D. Fla. April 1, 2004) (denying post-trial motion for declaration of copyright infringement, because, inter alia, the website that was being scraped did not own the copyright to the data and images that were being copied).

[3] See *Feist Publ'ns v. Rural Tel. Serv.*, 499 U.S. 340 (1991).

[4] See *Ticketmaster v. Tickets.com*, No. 99-CV-7654, 2003 WL 21406289, at *4-6 (C.D. Cal. March 7, 2003); see also *Nautical Solutions*, 2004 WL 783121, at *2-3 (reaching similar result for scraping of information regarding the sale of yachts).

[5] See 17 U.S.C. §107.

[6] *Kelly v. Arriba Soft*, 336 F.3d 811, 819 (9th Cir. 2003). An in-depth discussion of the nuances of the fair use doctrine is outside the scope of this article. For a discussion of fair use, see Melville B. Nimmer, 4 Nimmer on Copyright §13.05 (Lexis 2013).

[7] See Restatement (Second) of Torts §218 (Westlaw 2012); see also *Bidder's Edge*, 100 F. Supp. 2d at 1069.

[8] *Bidder's Edge*, 100 F. Supp. 2d at 1067.

[9] Id. at 1070.

[10] Id. at 1071-72.

[11] Id. at 1071.

[12] *Register.com v. Verio*, 356 F.3d 393, 404-05 (2d Cir. 2004).

[13] Id. at 404.

[14] *Ticketmaster*, 2003 WL 21406289, at *3.

[15] See, e.g., *Bidder's Edge*, 100 F. Supp. 2d at 1067; *Zefer*, 318 F. 3d at 62.

[16] See *Specht v. Netscape Commc'ns*, 306 F.3d 17, 22 n.4 (2d Cir. 2002); *Hines v. Overstock.com*, 668 F. Supp. 2d 362, 366-67 (E.D.N.Y. 2009).

[17] See *Specht*, 306 F.3d at 22 n.4.

[18] *Bidder's Edge*, 100 F. Supp. 2d at 1060.

[19] Id. at 1067.

[20] See *Specht*, 306 F.3d at 25.

[21] See, e.g., *Specht*, 306 F.3d at 35; *Hines*, 668 F. Supp. 2d at 367.

[22] See *Specht*, 306 F.3d at 35 (finding a browsewrap agreement unenforceable).

[23] See *Hines*, 668 F. Supp. 2d at 367.

[24] *Sw. Airlines v. BoardFirst*, No. 3:06-CV-0891, 2007 WL 4823761, at *7 (N.D. Texas Sept. 12, 2007).

[25] *Register.com*, 356 F.3d at 401-04.

[26] Id.

[27] See id. at 402-03.

[28] *Cvent v. Eventbrite*, 739 F. Supp. 2d 927, 937 (E.D. Va. 2010); see also *Hines*, 668 F. Supp. 2d at 367.

[29] 18 U.S.C. §1030(a)(4); see also 18 U.S.C. §1030(g) (providing for civil liability and a private right of action).

[30] See 18 U.S.C. §1030(a)(4).

[31] *Sw. Airlines v. Farechase*, 318 F. Supp. 2d 435, 440 (N.D. Tex. 2004).

[32] Id. at 439-40; see also *Zefer*, 318 F.3d at 62-63 (upholding a preliminary injunction issued under the CFAA where defendant had knowledge that scraping was unauthorized).

[33] *Cvent*, 739 F. Supp. 2d at 932-34.

[34] Id.