The Journal of Business, Entrepreneurship & the Law

Volume 12 | Issue 1 Article 8

May 2019

Making Room For Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online

Amber Zamora

Follow this and additional works at: https://digitalcommons.pepperdine.edu/jbel

 Part of the First Amendment Commons, Intellectual Property Law Commons, and the Internet Law Commons

Recommended Citation

Amber Zamora, *Making Room For Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online*, 12 J. Bus. Entrepreneurship & L. 203 (2019)
Available at: https://digitalcommons.pepperdine.edu/jbel/vol12/iss1/8

This Note is brought to you for free and open access by the School of Law at Pepperdine Digital Commons. It has been accepted for inclusion in The Journal of Business, Entrepreneurship & the Law by an authorized editor of Pepperdine Digital Commons. For more information, please contact josias.bartram@pepperdine.edu , anna.speth@pepperdine.edu.

MAKING ROOM FOR BIG DATA: WEB SCRAPING AND AN AFFIRMATIVE RIGHT TO ACCESS PUBLICLY AVAILABLE INFORMATION ONLINE

Amber Zamora

INTRODUCTION

"Big data."[1] Though many internet users are not aware of it, big data fuels innovation in today's world.[2] People use big data to further scientific discovery, diagnose and identify new diseases, learn more about

[1] Bernard Marr, *The Complete Beginner's Guide to Big Data Everyone Can Understand*, FORBES (Mar. 14, 2017), https://www.forbes.com/sites/bernardmarr/2017/03/14/the-complete-beginners-guide-to-big-data-in-2017/#18d569777365.
[2] *Id.*

the Earth and other planets, keep communities safe, and build efficiencies into our everyday lives.[3]

By some estimates, between 2012 and 2014, humans created one zettabyte of data.[4] Indeed, estimates from IBM projected that by 2010, the world's collection of information would double in size every eleven hours.[5] This wealth of information creates a massive opportunity to draw insights about everything in our world, and many private companies and public organizations have endeavored to capture this information and use it innovatively.

Successfully analyzing big data involves a three-step process: the data must first be collected, then the data must be analyzed within an optimizable model, and, finally, the data must provide informed conclusions, suggestions, or strategies for the data consumer.[6] This first step—data collection—is the subject of this paper.

One principal method to collect data is through web scraping.[7] Web scraping, also called web crawling or data scraping, "refers to the act of extracting large amounts of information from a website using automated software programs called bots."[8] These web scrapers exist in a legal gray area, not only potentially creating new uses for the collected information, but also potentially harming the host website.[9] As the number of available datasets and the hunger for new data insights grow, online platforms increasingly seek to protect their information from web scrapers.[10] Unfortunately, these online platforms do not always successfully distinguish bad from good actors, thus risking stomping out new innovation and valid competition.

This paper will explore the legality of web scraping through the lens of recent litigation between web scraper hiQ Labs and the online professional networking platform, LinkedIn. First, the paper will study the background of web scraping litigation, some challenges courts face in

---

[3] *Id.*

[4] Jonathan Shaw, *Why "Big Data" is a Big Deal*, HARVARD MAGAZINE (Mar.–Apr. 2014), https://www.harvardmagazine.com/2014/03/why-big-data-is-a-big-deal. One zettabyte is equivalent to 10^{21} bytes. *Convert Zettabyte to Gigabyte - Conversion of Measurement Units*, CONVERTUNITS, https://www.convertunits.com/from/zettabyte/to/gigabyte (last visited Oct. 23, 2018).

[5] *The Toxic Terabyte: How Data-Dumping Threatens Business Efficiency*, IBM GLOBAL TECHNOLOGY SERVS 2 (July 2006), http://www-935.ibm.com/services/no/cio/leverage/levinfo_wp_gts_thetoxic.pdf.

[6] ANGIE M. TAYLOR, ET AL., BIG DATA ANALYTICS: MEGATRENDS TO BUSINESS SUCCESS, 2017 WL 4284476 (July–Aug. 2017).

[7] Vladimir Fedak, *Big Data: What is Web Scraping and How to Use It*, TOWARDS DATA SCI. (Feb. 9, 2018), https://towardsdatascience.com/big-data-what-is-web-scraping-and-how-to-use-it-74e7e8b58fd6.

[8] ALM Media, *What Courts Have Said About the Legality of Data Scraping*, YAHOO! FIN. (July 20, 2017), https://finance.yahoo.com/news/courts-said-legality-data-scraping-090000366.html.

[9] *Id.*

[10] *Id.*

issuing consistent verdicts, and the most common claims companies make against web scrapers.

Then the paper will address three of the most common claims and identify court motivations and limitations within the doctrines. The first claims are those arising from the federal Computer Fraud and Abuse Act (CFAA). Next, the paper will investigate copyright claims and defenses that may be applicable to web scrapers. Finally, it will discuss the state law claims of trespass to chattels and of breach of contract.

Considering these doctrines, this paper will propose a legal protection for web scrapers accessing public information online to draw a comprehensive limitation on web scraping litigation, in order to protect online activity in light of First Amendment protections, anticompetition concerns, and online public policy. With such a protection in place, this paper will argue how the web as a crucible of knowledge will be preserved.

I. BACKGROUND

Courts have struggled to reach consistent resolutions in web scraping cases. One major obstacle to consistent verdicts is what this paper will term the "kitchen sink" argument—the typical argument in web scraping litigation.[11] The kitchen sink argument most often includes the following claims: (a) civil claims under the CFAA alleging that the defendant "exceed[ed] authorized access,"[12] (b) copyright infringement claims under the Digital Millennium Copyright Act or state copyright law,[13] (c) state trespass to chattels claims,[14] and (d) breach of contract claims.[15] This list, while not exhaustive, is representative of the most common claims made in web scraping litigation and will therefore serve as a foundation for our discussion.[16] With many grounds for relief, the challenge of consistent verdicts is nearly insurmountable.[17]

A second obstacle to consistent verdicts is that numerous purposes exist along a spectrum of social acceptability for a business model employing web scrapers.[18] For example, one of the most well-known web

---

[11] *Definition of Kitchen-Sink*, MERRIAM-WEBSTER (2018), https://www.merriam-webster.com/dictionary/kitchen-sink.

[12] *See, e.g.*, United States v. Nosal, 676 F.3d 854 (9th Cir. 2012); United States v. Nosal, 844 F.3d 1024, 1029 (9th Cir. 2016).

[13] *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

[14] *See, e.g.*, Register.com, Inc. v. Verio, Inc., 356 F.3d 393 (2d Cir. 2004).

[15] *See, e.g.*, *id.*

[16] For a discussion of these claims, *see* Jeffrey Kenneth Hirschey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, 29 BERKELEY TECH. L.J. 897, 908–18 (2014).

[17] *Cf.* James Snell & Nicola Menaldo, *Web Scraping in an Era of Big Data 2.0*, TECH & TELECOM ON BLOOMBERG LAW (June 8, 2016), https://www.bna.com/web-scraping-era-n57982073780/ ("[Web scraping's] legal status remains highly context-specific. And many of the most interesting legal questions emerging from this trend remain unanswered or depend on very specific factual context.").

[18] *See* Vimal Maheedharan, *A Detailed Overview of Web Crawlers*, CABOT SOLS. (Nov. 11, 2016), https://www.cabotsolutions.com/2016/11/a-detailed-overview-of-web-crawlers/.

scrapers is Google's web-indexing tool Googlebot.[19] This web scraper serves an important purpose for the users of Google search: without Googlebot, Google would not be able to rank, sort, and index search results for its users.[20] Googlebot uses an algorithm to determine which websites to crawl and how often, using links from these websites to build larger lists of websites to subsequently access and crawl.[21] On the other end of the spectrum are impersonator bots, comprising a total of twenty-four percent of overall web traffic in 2016; these bots intentionally disrupt traffic to targeted websites (called a denial-of-service, or DDOS, attack).[22] Most web scraping technologies fall somewhere along the spectrum between these two extremes,[23] making it understandably difficult to render judgments affecting the entire industry.

As web scraping becomes a larger part of the internet ecosystem,[24] businesses have turned to the courts to better define the industry's legal boundaries. Judges have imposed limits on these cases when possible; for example, where a plaintiff company has not suffered any harm from the web scraping,[25] in certain cases when a defendant does not have the requisite intent to harm,[26] when judges can narrowly interpret statutory requirements,[27] and where public policy permits.[28] Now, in consideration of early rulings in *hiQ Labs v. LinkedIn*, this paper will propose explicit permissions for accessing publicly available information online.

---

[19] *Googlebot*, GOOGLE (2018), https://support.google.com/webmasters/answer/182072.

[20] *How Search Works*, GOOGLE, https://www.google.com/search/howsearchworks/ (last visited Mar. 10, 2018).

[21] *Id.*

[22] Adrienne LaFrance, *The Internet is Mostly Bots*, THE ATL. (Jan. 31, 2017), https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/.

[23] *Id.*

[24] Overall, bots made up fifty-two percent of all web traffic in 2016. *Id.* This number represents a significant increase from even a year before, when web scraping traffic represented less than a quarter of overall web traffic. Paven Malhotra, et al., *What Courts Have Said About the Legality of Data Scraping; Parties Have Sought to Stop Scrapers Using a Number of Legal Bases, from the CFAA to Copyright Law*, LEGALTECH NEWS (July 20, 2017), https://advance.lexis.com/api/permalink/a7ac2880-d590-46bd-ac08-1b1aef267b4d/?context=1000516.

[25] *See, e.g.,* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

[26] *See generally* QVC, Inc. v. Resultly, LLC, 159 F. Supp. 3d 576 (E.D. Penn. 2016).

[27] For example, the Ninth and other circuits narrowly construe the CFAA "without authorization" requirement. *But see* Myra F. Din, *Breaching and Entering: When Data Scraping Should Be a Federal Computer Hacking Crime*, 81 BROOK. L. REV. 405 (2015).

[28] *See, e.g.,* Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (holding that the public benefit of utilizing data from copies of books hosted on the Google Books platform far outweighed the harm of the copying done).

II. OVERVIEW OF THE LINKEDIN LITIGATION

LinkedIn Corporation ("LinkedIn") is a social media platform for professional networking founded in 2003.[29] Microsoft acquired the company in 2016 for $26.2 billion.[30] LinkedIn has amassed a large user base—more than 450 million users in over 200 countries worldwide[31]—by allowing for customizable professional profiles, job searching, recruiting functionality, and analytics.[32]

hiQ Labs, Inc. ("hiQ") is a people-analytics company founded in 2012.[33] hiQ offers two core products to its customers, who are primarily corporate human-resources departments: Keeper, a tool that monitors employee LinkedIn profiles for changes and alerts employers which employees are at the greatest risk of being recruited away, and Skill Mapper, a service utilizing employee LinkedIn profiles to summarize the skills possessed by individual employees.[34] These products rely entirely on information scraped from user-created LinkedIn profiles designated public by the individual LinkedIn users.[35]

hiQ has been successful with this model.[36] In the six years since its founding, hiQ has raised over $12 million from venture capital firms[37] and counts among its customers various Fortune 500 companies including online auction site eBay, credit card company Capital One, and website domain host GoDaddy.[38] hiQ is also an active member of the human resources technology community, and hosts the well-attended people-analytics conference Elevate in San Francisco and New York every year.[39]

However, despite these success, hiQ's entire business model came under threat when LinkedIn sent a cease-and-desist letter on May 23,

---

[29] *Id.* at 1103.

[30] *LinkedIn*, CRUNCHBASE, https://www.crunchbase.com/organization/linkedin#section-locked-marketplace (last visited Mar. 11, 2018). For a more detailed look at the acquisition timeline, *see Schedule 14A: LinkedIn Corporation*, SEC (July 22, 2016), https://www.sec.gov/Archives/edgar/data/1271024/000104746916014430/a2229104zdefm14a.htm.

[31] *Number of LinkedIn Members from 1st Quarter 2009 to 3rd Quarter 2016 (in Millions)*, STATISTA (2018), https://www.statista.com/statistics/274050/quarterly-numbers-of-linkedin-members/#0.

[32] Mike Isaac, *A LinkedIn Timeline*, N.Y. TIMES (June 13, 2016), https://www.nytimes.com/2016/06/14/technology/a-linkedin-timeline.html

[33] *hiQ Labs*, CRUNCHBASE, https://www.crunchbase.com/organization/hiq-labs#section-overview (last visited Mar. 11, 2018).

[34] hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1104 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

[35] *Id.*

[36] *See hiQ Labs*, *supra* note 33.

[37] *Id.*

[38] Deepak Gupta, RE: Cease and Desist Letter to hiQ Labs, Inc. (2017), https://static1.squarespace.com/static/5803b57737c581885cbd0667/t/5977a69737c581c4face71ca/1501013656158/Ex.+K+-+2017-05-31+Response+Letter+to+Bajoria.pdf.

[39] *hiQ Labs Announces 2016 hiQ Elevate Event Series for People Analytics Community*, PR NEWSWIRE (Feb. 4, 2016), https://www.prnewswire.com/news-releases/hiq-labs-announces-2016-hiq-elevate-event-series-for-people-analytics-community-300215112.html.

2017.[40] The letter charged hiQ with "using processes to improperly, and without authorization, access and copy data from LinkedIn's website."[41] As support for its demand, LinkedIn cited its own user agreement, in which web scraping, copying, and using information from user profiles in any way without express permission is prohibited.[42] LinkedIn further noted that technological barriers had been put in place to limit hiQ's access to the platform and warned that circumventing these barriers would constitute a violation of both state and federal law.[43]

In response, hiQ requested a telephone meeting to attempt a resolution.[44] When those efforts failed,[45] hiQ filed a temporary restraining order to enjoin LinkedIn from denying hiQ access to its platform.[46] hiQ argued that LinkedIn's conduct was motivated by anticompetitive intent,[47] that LinkedIn lacked copyright or other exclusive interest in the data scraped by hiQ,[48] and that LinkedIn had threatened to sue under non-applicable state and federal laws.[49]

LinkedIn's response to these claims focused on four main points: hiQ's business sells LinkedIn user data to its clients without permission from those users[50]; hiQ's technology threatens the security of LinkedIn's platform and operates without regard for privacy protections promised to LinkedIn users[51]; hiQ could offer a similar product without scraping LinkedIn user profiles[52]; and LinkedIn properly denied access to the

---

[40] Drake Bennett, *The Brutal Fight to Mine Your Data and Sell It to Your Boss*, BLOOMBERG BUSINESSWEEK (Nov. 15, 2017), https://www.bloomberg.com/news/features/2017-11-15/the-brutal-fight-to-mine-your-data-and-sell-it-to-your-boss.

[41] Abhishek Bajoria, RE: Demand to Immediately Cease and Desist Unauthorized Data Scraping and other Violations of LinkedIn's User Agreement (2017), https://static1.squarespace.com/static/5803b57737c581885cbd0667/t/59721e45725e2539a60bb195/1500651078233/Letter+from+LinkedIn+to+HiQ+Labs.pdf (last visited Mar 11, 2018) (on behalf of LinkedIn Corporation, in his capacity as Senior Litigation Counsel).

[42] *Cf. User Agreement: Provision 8.2 Don'ts*, LINKEDIN (updated June 7, 2017), https://www.linkedin.com/legal/user-agreement. Provision 8.2(k) notably states that users shall not "[d]evelop, support or use software, devices, scripts, robots, or any other means or processes (including crawlers, browser plugins and add-ons, or any other technology) to scrape the [s]ervices or otherwise copy profiles and other data from the [s]ervices." *Id.*

[43] Abhishek, *supra* note 41, at 2.

[44] Gupta, *supra* note 38.

[45] *Id.* at 4.

[46] Plaintiff's Memorandum of Points and Authorities in Support of Renewed Ex Parte Motion for Temporary Restraining Order and Order to Show Cause RE: Preliminary Injunction, *hiQ Labs, Inc. v. LinkedIn Corporation*, 2017 WL 7715792 at 9–10 (N.D. Cal. June 22, 2017).

[47] *Id.* at 7. Indeed, hiQ alleged that the motivation behind the cease-and-desist letter was LinkedIn's development of its own data analysis products. *Id.*

[48] *Id.* at 13.

[49] *Id.* at 19.

[50] LinkedIn Corporation's Opposition to Plaintiff's Motion for a Temporary Restraining Order, hiQ Labs, Inc. v. LinkedIn Corporation, 2017 WL 7715799, at *2 (N.D. Cal. June 26, 2017).

[51] *Id.* at 3–5.

[52] *Id.* at 6.

platform with the cease-and-desist letter.[53] Additionally, LinkedIn stated that hiQ's conduct violated the Computer Fraud and Abuse Act by attempting to intentionally access LinkedIn's servers without authorization.[54]

The court issued its ruling in August 2017; in granting the preliminary injunction, the court notably held the following:

- hiQ faced the prospect of irreparable harm if LinkedIn continued to block its access to public user profiles;[55]
- there were serious doubts about LinkedIn's application of the Computer Fraud and Abuse Act to the facts at hand.[56]

Specifically, the court held that a serious ambiguity arose under the facts presented, namely, whether the CFAA applied to restriction of access to publicly available websites.[57] The court noted that the CFAA "was not intended to police traffic to publicly available websites on the Internet," and instead was originally designed for police hacking and trespass onto private and password-protected computers.[58] Further, the court referenced the Ninth Circuit's caution in *United States v. Nosal* against an overbroad interpretation of the CFAA.[59]

As of this writing, the case is currently on appeal to the Ninth Circuit, so the legal world will have to wait for a final decision on the merits of these claims.[60] However, hiQ and LinkedIn are only the most recent entrants into a murky, legal minefield for the data scraping industry. This paper will next analyze case precedent to uncover the recognized exceptions, and use cases to illustrate the myriad of claims presented and to identify areas that need further clarification. Finally, this paper will draw the conclusion that *hiQ Labs* presents a ripe opportunity to outline an exception to the CFAA for Internet users accessing publicly available information using manual or automatic methods.

III. ARGUMENT ANALYSIS: THE COMPUTER FRAUD AND ABUSE ACT

The principal controversy in the *hiQ* case is LinkedIn's potential claim under the Computer Fraud and Abuse Act (CFAA).[61] If LinkedIn were to

---

[53] *Id.* at 7.

[54] *Id.* at 8.

[55] hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1106 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

[56] *Id.* at 1108.

[57] *Id.* at 1110.

[58] *Id.*

[59] *Id.* (quoting *United States v. Nosal*, 676 F.3d 854, 858 (9th Cir. 2012)). The court further noted, "[u]nder LinkedIn's interpretation of the CFAA, a website would be free to revoke 'authorization' with respect to any person, at any time, for any reason, and invoke the CFAA for enforcement, potentially subjecting an Internet user to criminal, as well as civil, liability." *Id.*

[60] hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d at 1110.

[61] *Id.* at 1108.

bring a claim against hiQ under this theory, hiQ may be subject to both civil and criminal liability.[62] This next section will unpack a brief history of the CFAA and the essential elements of a claim under the CFAA, and then outline areas of controversy from case precedent.

A. Introduction to the CFAA

The first iteration of federal law developed to address computer crimes was passed in 1984 as the Comprehensive Crime Control Act.[63] In 1986, Congress expanded the law to include crimes against unauthorized access to computers and unauthorized use of computers and computer networks, renaming this amended law the Computer Fraud and Abuse Act.[64] Though originally intended to limit the reach of the statute to cases with a "compelling federal interest"—including cases that involve computers belonging to the federal government or some financial institutions or where the crime involves interstate conduct—subsequent amendments broadened the statute to include "those computers used in or affecting interstate or foreign commerce or communication."[65] A separate amendment also created a civil claim whereby private individuals and entities could also obtain relief under the CFAA.[66] This civil claim is a primary vehicle for private companies to obtain relief for insider and outsider unauthorized access to a web platform.

B. Elements of a Claim

While most of the claims under the CFAA address criminal activities involving government computers, private claims under the CFAA arise primarily when a plaintiff can prove that the defendant: (a) "intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer;"[67] (b) "knowingly and with intent to defraud, accesses a protected computer without authorization, or exceeds authorized access, and by means of such conduct furthers the intended fraud and obtains anything of value, unless the object of the fraud and the thing obtained consists only of the use of the computer and the value of such use is not

---

[62] 18 U.S.C. § 1030 (2008).

[63] *Prosecuting Computer Crimes*, EXEC. OFF. FOR U.S. ATT'YS OFF. OF LEGAL EDUC. 1 (Jan. 14, 2015), https://www.justice.gov/sites/default/files/criminal-ccips/legacy/2015/01/14/ccmanual.pdf.

[64] *Id.*

[65] *Id.* at 1–2. To date, the CFAA has been amended eight times: in 1988, 1989, 1990, 1994, 2001, 2002, and 2008. *Id.* at 2.

[66] *Id.* at 3.

[67] 18 U.S.C. § 1030(a)(2) (2008).

more than $5,000 in any 1-year period;"[68] or (c) "knowingly causes the transmission of a program, information, code, or command, and as a result of such conduct, intentionally causes damage without authorization, to a protected computer."[69]

In addition to these elements, a plaintiff must prove that defendant's conduct caused damage in one of the following five categories: loss of at least $5,000 in total value,[70] modification or impairment—or the potential for such—of medical care of another person,[71] physical injury,[72] any threat to public health and safety,[73] or damage that affects ten or more computers designated "protected" during a one-year period.[74]

At issue in *hiQ Labs* was how broadly to interpret the terms "access" and "without authorization."[75]

C. Rule of Lenity and Principles of Narrow Statutory Construction

One common theme in CFAA litigation is conflicting interpretations of included provisions, and courts have attempted to narrow interpretation through application of the rule of lenity.[76] The rule of lenity is defined as a principle by which courts must construe criminal laws: any ambiguities in a criminal law must be resolved in favor of the defendant.[77] This principle holds true unless doing so is clearly against the intent of the legislature.[78]

One well-known example of application of the rule of lenity to CFAA cases is *U.S. v. Nosal* ("*Nosal I"*).[79] In *Nosal I*, David Nosal, a former employee of executive search firm Korn/Ferry, convinced current Korn/Ferry employees to log into the private company system and download information from Korn/Ferry's confidential database.[80] The information included source lists, names, and contact information that

---

[68] 18 U.S.C. § 1030(a)(4) (2008).
[69] 18 U.S.C. § 1030(a)(5)(A) (2008).
[70] 18 U.S.C. § 1030(c)(4)(A)(i)(I) (2008).
[71] 18 U.S.C. § 1030(c)(4)(A)(i)(II) (2008).
[72] 18 U.S.C. § 1030(c)(4)(A)(i)(III) (2008).
[73] 18 U.S.C. § 1030(c)(4)(A)(i)(IV) (2008).
[74] 18 U.S.C. § 1030(c)(4)(A)(i)(VI) (2008).
[75] Deborah F. Buckman, *Validity, Construction, and Application of Computer Fraud and Abuse Act (18 U.S.C.A. § 1030)*, 174 A.L.R. Fed. 101 (2001).
[76] *See generally* United States v. Nosal, 676 F.3d 854 (9th Cir. 2012). The Ninth Circuit in *Nosal* encouraged Congress to clarify its intent regarding the more ambiguous provisions of the CFAA by noting that "[t]he rule of lenity requires 'penal laws . . . to be construed strictly.'" *Id.* at 863. "[W]hen choice has to be made between two readings of what conduct Congress has made a crime, it is appropriate, before we choose the harsher alternative, to require that Congress should have spoken in language that is clear and definite." *Id.* (quoting Jones v. United States, 529 U.S. 848, 858 (2000)).
[77] *Rule of Lenity*, MERRIAM-WEBSTER (2018), https://www.merriamwebster.com/legal/rule%20of%20lenity (last visited Oct. 30, 2018).
[78] *Nosal*, 676 F.3d at 857.
[79] *Id.*
[80] *Id.* at 856.

would aid Nosal in building a competing business.[81] Though these employees technically were authorized to access the database, the government indicted them, along with Nosal, for exceeding authorized access with the intent to defraud.[82]

Nosal filed to dismiss the CFAA counts based on a narrow construction of the phrase "exceeds authorized access," and the lower court held in his favor.[83] The government appealed, arguing that that the phrase "exceeds authorized access" applied to individuals who had unfettered access to a computer or database but used the information gleaned from the computer for an untoward purpose.[84]

The Ninth Circuit disagreed with this interpretation of the statute, noting that the government's interpretation "would transform the CFAA from an anti-hacking statute into an expansive misappropriation statute."[85] The court analyzed legislative history for guidance, finding that the motivation behind the statute was to prevent computer hacking and feared that a broad construction of the provisions would "make criminals of large groups of people who would have little reason to suspect they are committing a federal crime."[86]

Notably, the court asserted that online behavior required utilizing "one computer to send commands to other computers at remote locations," and suggested that the many agreements and policies underlying this interaction were only vaguely understood by members of the public using the internet every day.[87] As an example, the court pointed to the 2007 through 2012 iteration of the Google Terms of Service which forbade minors from utilizing the Google search function.[88] Any minors who used Google Search during this time period, the court noted, would be subject to criminal liability under the government's interpretation of the CFAA.[89]

---

[81] *Id.*

[82] *Id.*

[83] *Id.* In issuing its decision, the lower court cited *LVRC Holdings LLC v. Brekka*, 581 F.3d 1127 (9th Cir. 2009), in which the Ninth Circuit applied narrow constructions of the phrases "without authorization" and "exceeds authorized access" in the CFAA to a similar fact pattern involving employee access to an employer database. *Id.*

[84] *Nosal*, 676 F.3d at 856.

[85] *Id.* at 857. The Ninth Circuit further noted that "[i]f Congress meant to expand the scope of criminal liability to everyone who uses a computer in violation of computer use restrictions—which may well include everyone who uses a computer—we would expect it to use language better suited to that purpose." *Id.*

[86] *Id.* at 859. "Minds have wandered since the beginning of time and the computer gives employees new ways to procrastinate, by g-chatting with friends, playing games, shopping or watching sports highlights. Such activities are routinely prohibited by many computer-use policies, although employees are seldom disciplined for occasional use of work computers for personal purposes. Nevertheless, under the broad interpretation of the CFAA, such minor dalliances would become federal crimes." *Id.* at 860.

[87] *Id.*

[88] *Id.* at 861.

[89] *Id.*

The court was not reassured when the government insisted it would not prosecute minor claims,[90] instead suggesting that the government would favor conservative application of the law only until an attractive target came around.[91]

Finally, the court resisted applications of the CFAA that might criminalize a "broad range of day-to-day activity."[92] By creating a crime of conduct that is highly subjective, the court reasoned that judges and juries would struggle to apply the law consistently.[93] It also suggested that failure to be mindful of the effects of broad interpretations of the CFAA on ordinary citizens had led other circuits away from Congress's original intent.[94] The Court concluded that "exceeds authorized access" did not extend to violations of a company's use restrictions.[95]

The Court revisited *Nosal* in 2016.[96] While *Nosal I* had considered the phrase "exceeds authorized access," *Nosal II* considered the reach of "'knowingly and with intent to defraud' accessing a computer 'without authorization'" provision of the CFAA, namely, whether it applied to Nosal as a former employee who accessed the Korn/Ferry computer database through alternate means after his credentials were revoked.[97] The Court did not hesitate to find Nosal liable under this provision of the CFAA, noting that the *mens rea* requirements would prevent any innocent or well-meaning citizens from liability.[98]

The Court distinguished the two opinions by reiterating its rule from *Nosal I*: the CFAA could not be broadly construed to create liability for unauthorized use of corporate information or violations of corporate fiduciary duties.[99] However, since the provision at issue in *Nosal II* involved both intent and merely access, the Court did not struggle to apply liability to Nosal and his associates.[100]

The rule from *Nosal II* is problematic in web scraping cases because it affirms the position that companies are permitted to revoke

---

[90] *Id.* at 862. The Court cited *United States v. Stevens*, 559 U.S. 460 (2010), in which the Supreme Court stated, "[w]e would not uphold an unconstitutional statute merely because the Government promised to use it responsibly." *Id.*

[91] *Id. Cf.* United States v. Drew, 259 F.R.D. 449 (C.D. Cal. 2009) (holding that a mother who posed as a teenage boy on Myspace and bullied her daughter's classmate was guilty under 18 U.S.C. § 1030(a)(2)(C) for violating Myspace's Terms of Service).

[92] *Nosal*, 676 F.3d at 862 (quoting United States v. Kozminski, 487 U.S. 931 (1988)).

[93] *Id.*

[94] *Id. See, e.g.,* United States v. Rodriguez, 628 F.3d 1258 (11th Cir. 2010) (holding that defendant violated the CFAA when he used his access to the Social Security database to look up old romantic partners, friends, and acquaintances in violation of agency policy); United States v. John, 597 F.3d 263 (5th Cir. 2010) (holding that defendant exceeded authorized access when she used her bank login credentials to access account information for corporate clients in order to execute a fraud); Int'l Airport Ctrs, LLC v. Citrin, 440 F.3d 418 (7th Cir. 2006) (holding that a defendant that had installed and utilized a secure-erase software on an employer-provided laptop had violated the CFAA).

[95] *Nosal*, 676 F.3d at 863.

[96] United States v. Nosal, 844 F.3d 1024, 1029 (9th Cir. 2016).

[97] *Id.* at 1029.

[98] *Id.*

[99] *Id.* at 1034.

[100] *Id.*

access to their platforms for any reason.[101] The dissent in *Nosal II* warned that a broad interpretation of the "access" provision would undo the work done in *Nosal I* to protect the daily activities of unsuspecting citizens.[102] The dissent further encouraged the majority to limit application of the CFAA to the original purpose of the law: to stop hackers.[103]

The essential difference between *Nosal I*, *Nosal II*, and *hiQ Labs* is the nature of the information at issue: the information in *Nosal I & II* was protected by a password authentication system, while the information in *hiQ* was publicly available.[104] Indeed, the information on the public LinkedIn pages is regularly indexed by Googlebot and other search engines, discussed above.[105] Attorney and Harvard Law School constitutional scholar Laurence Tribe drew the following real-world analogy: libraries once contained books with a physical card attached, listing the book's borrowing history.[106] If the government were to try to prevent a company from indexing borrowing history of library books using the CFAA in order to ensure it remained the exclusive distributor of such information, in Professor Tribe's view, that conduct would be clearly unconstitutional.[107]

Another notable difference is that David Nosal perpetrated all of the misdeeds at issue in *Nosal I and Nosal II* despite a signed non-compete agreement between Nosal and former employer Korn/Ferry.[108] This bilateral agreement involved negotiation between Nosal and Korn/Ferry and required Nosal to refrain from any action—whether including confidential information or not—that would result in Nosal entering the market until a year after his separation from Korn/Ferry.[109] Unlike Nosal, *hiQ* was under no agreement or understanding with LinkedIn; it did not negotiate a bilateral agreement; instead, only LinkedIn's various terms of service and use attempted to control access and use of its platform.[110] These agreements, said the *Nosal I* court, are often "lengthy, opaque,

---

[101] *Cf.* Christine D. Galbraith, *Access Denied: Improper Use of the Computer Fraud and Abuse Act to Control Information on Publicly Accessible Internet Websites*, 63 MD. L.R. 320, 338 (2004).

[102] Nosal, 844 F.3d at 1051.

[103] *Id.* The dissent further stated: "[w]e would not convict a man for breaking and entering if he had been invited in by a houseguest, even if the homeowner objected." *Id.*

[104] hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1108 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

[105] Allison Frankel, *hiQ v. LinkedIn: Does First Amendment Limit Application of Computer Fraud Law?*, REUTERS (Aug. 1, 2017), https://www.reuters.com/article/us-otc-linkedin/hiq-v-linkedin-does-first-amendment-limit-application-of-computer-fraud-law-idUSKBN1AH59X.

[106] *Id.*

[107] *Id.*

[108] Nosal, 844 F.3d at 1030.

[109] *Id.*

[110] *hiQ Labs, Inc.*, 273 F. Supp. at 1104.

subject to change and seldom read," and it is problematic to apply potentially criminal liability to such one-sided rule-making.[111]

IV. ARGUMENT ANALYSIS: COPYRIGHT

Though not directly at issue in *hiQ Labs*, many web platforms assert copyright claims during web scraping litigation.[112] This is generally because web scraping by its nature involves copying, one of the principal ways to infringe a copyright.[113] In asserting copyright claims, internet platforms can run afoul of the essential elements of asserting a claim: they must be able to assert ownership, they must be attempting protection of expression, not ideas, and they must be able to negate a fair use defense.[114]

A. Introduction to Copyright

Copyright law aims to protect authors of original works for a fixed amount of time after the work is created.[115] Copyright owners have limited monopolies on creative works and are imbued with the exclusive rights to reproduce the works, prepare derivative works, distribute copies, and perform and display the works publicly.[116] Copyrights automatically attach to creative works once fixed in a medium and can be then transferred to others through license or other contracted means.[117]

B. Elements of a Copyright Claim

An action under copyright law exists when a plaintiff can prove: (a) ownership of a valid copyright and (b) copying the original elements of the work.[118] There can be no valid copyright in mere facts, though the presentation of those facts may be subject to copyright.[119] In addition, web

---

[111] United States v. Nosal, 676 F.3d 854, 860 (9th Cir. 2012); *see also* Tom Towers, *Thousands Sign Up for Community Service After Failing to Read Terms and Conditions*, METRO (July 14, 2017, 11:12pm), http://metro.co.uk/2017/07/14/thousands-sign-up-for-community-service-after-failing-to-read-terms-and-conditions-6781034/.

[112] *See, e.g.,* Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962 (N.D. Cal. 2013); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2013).

[113] Anthony J. Dreyer and Jamie Stockton, *Internet 'Data Scraping': A Primer for Counseling Clients*, N.Y. L.J., LITIG. (July 15, 2013), https://www.skadden.com/insights/publications/2014/01/internet-data-scraping-a-primer-for-counseling-cli.

[114] *Id.*

[115] *Copyright Basics*, U.S. COPYRIGHT OFFICE (Sept. 2017), https://www.copyright.gov/circs/circ01.pdf.

[116] *Id.* at 2.

[117] *Id.* at 2–3.

[118] Feist Publ'n, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991).

[119] *Id.*

platforms may assert valid copyrights over their users' created works if the terms of service include an automatic transfer provision.[120]

A defendant's principal protection in copyright law is a fair use defense.[121] Courts weigh a number of factors in determining if a use is fair: (a) the purpose and character of the use, (b) nature of the protected work, (c) amount of the work used, and (d) the market value of the use.[122] A number of commercial web scrapers have had success asserting a fair use defense to a web platform's copyright claims.[123]

C. Fair Use: Authors Guild v. Google

One seminal example of application of the fair use defense in context of the Internet is *Authors Guild v. Google.*[124]

Google was founded in 1998 with the goal of building a search engine that would organize web pages on the internet and use links between websites to determine the importance of individual pages online.[125] The company experienced explosive popularity and growth and today boasts 60,000 employees, a catalog of hundreds of products, and a robust Google Search at the core of its activity.[126]

Authors Guild is a collective of writers founded in 1912.[127] As a collective, Authors Guild advocates for writers in a variety of arenas including copyright, contracts, and free speech, in service of its mission to "protect the rights of all authors, whether engaged in literary, dramatic, artistic, or musical competition, and to advice and assist all such authors."[128] It counts among its members famous authors and advocates, including Theodore Roosevelt, who joined Authors Guild as Vice President after signing the Copyright Act into law in 1909.[129]

This case arose after Google launched its Google Books product in 2004, a product that—in its original conception—would provide a

---

[120] Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962 (N.D. Cal. 2013); Kelly v. Arriba Soft Corp., 336 F.3d 811, 971 (9th Cir. 2013).

[121] Dreyer et. al, *supra* note 113.

[122] *Id.*

[123] *See, e.g.,* Ticketmaster Corp. v. Tickets.com, Inc., 2003 WL 21406289 (C.D. Cal. 2003) (holding that a data aggregator that scraped plaintiff's ticket purchasing platform in order to acquire event information was protected from plaintiff's copyright claim by a fair use defense, even though use was for a commercial purpose and only slightly transformative when source code was downloaded, final display was only of plaintiff's aggregated non-copyrightable information, and defendant's final product did not damage the market value of plaintiff's product).

[124] Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

[125] *From the Garage to the Googleplex*, GOOGLE, https://www.google.com/intl/en/about/our-story/ (last visited Mar. 5, 2018).

[126] *Id.*

[127] *Who We Are*, THE AUTHORS GUILD (last visited Mar. 5, 2018), https://www.authorsguild.org/who-we-are/.

[128] *Id.*

[129] *Id.*

digital platform to libraries and would be able to supply data for language-based research projects.[130] In service of this project, Google scanned and indexed over 20 million books into a digital, readable format.[131] In digitized format, Google permitted researchers to conduct research on fluctuations in subject matter interest over time, word frequencies, and linguistic changes over time.[132] Additional functionality permitted general users limited snapshots of text, called "snippets."[133] Finally, Google permitted the libraries that provided source material to access full-length and complete digital copies of the books they provided.[134]

In 2005, Authors Guild organized a class action suit against Google on behalf of rights-owning authors.[135] After several iterations of proposed and rejected settlements, the *Authors Guild* case made its way to the Second Circuit.[136] The Court focused on the historic application of the fair use doctrine in guiding its decision, highlighting the original creation of fair use to permit unauthorized copying if such copying promoted "the Progress of Science and the useful Arts."[137]

In considering the first factor of a fair use analysis, the Court asked whether Google Books was a sufficiently transformative use of the copyrighted work.[138] It further clarified that "transformative" use required "justification for the taking," including commenting on or criticizing the original work.[139] In *Authors Guild*, the Court held that creating a searchable database including text from the copied books was a highly transformative use, though Google permitted viewing of a snippet of the text and Google itself is a for-profit entity, a characteristic that generally weighs against a finding of fair use.[140]

Notably, the Court quoted the Supreme Court's statement on commercial fair use in *Campbell v. Acuff-Rose Music, Inc.*, which rejected the argument that commercial fair use was by its nature invalid: "Congress could not have intended such a broad presumption against commercial fair

---

[130] Authors Guild v. Google, Inc., 804 F.3d 202, 207 (2d Cir. 2015) To access Google Books in its present form, visit https://books.google.com. The original concept for Google Books was developed in 1996 by Google co-founders Sergey Brin and Larry Page who wanted to create a digital library where people could browse large collections of digital book copies and analyze a single book's relevance and usefulness through analysis of connections with other written works. *Google Books History*, GOOGLE BOOKS, https://books.google.com/intl/en/googlebooks/about/history.html (last visited Mar. 5, 2018).

[131] Authors Guild v. Google, Inc., 804 F.3d at 208. This was first accomplished by scanning the books in the library collections of Harvard, the University of Michigan, the New York Public Library, Oxford, and Stanford. *Google Books History*, *supra* note 130.

[132] *Authors Guild*, 804 F.3d at 209.

[133] *Id.* at 210.

[134] *Id.*

[135] *Id.*

[136] *Id.* at 212.

[137] *Id.* (quoting *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994)).

[138] *Authors Guild*, 804 F.3d at 214.

[139] *Id.*

[140] *Id.* at 217.

uses, as nearly all of the illustrative uses listed in the preamble paragraph of § 107 are generally conducted for profit in this country."[141]

The Court focused the rest of its analysis on the third and fourth factors of fair use, namely, whether Google copied a substantial portion of the original work and whether that copy created a substitute or competing product to the original.[142] Here, it focused on the difficulty in accessing a smooth flow of text via the Google Books Snippets, and the fact that much of the book remains inaccessible in this viewable format.[143] The impossibility of accessing the complete work, the choppiness of the snippet view, and the difficulty in accessing a smooth flow of text led the Court to the conclusion that there was no significant risk that Google Books would devalue the author's rights.[144]

There are a few notable conclusions from this case that bear repeating in the context of hiQ *Labs*. First, fair use is not a defense limited to not-for-profit uses of works potentially subject to copyright protection.[145] This means that purveyors of commercial products, including data aggregators like hiQ, have a powerful tool in their arsenal. Second, notable factors weighing in favor of Google are also applicable to hiQ and other data aggregators. These web scrapers are collecting data and putting that data to a new use, thereby transforming the data from one form into an entirely new form for a new set of users — for example, utilizing a public resume as a signal of flight risk. Just as Google did not attempt to make entire copies of books available online for free, hiQ is not using copies of LinkedIn data to create a competing professional networking platform.[146]

D. Limits to Copyright: C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Advanced Media, L.P.

A second important aspect of copyright law involves deciding exactly what material is subject to copyright protection. This was the core

---

[141] *Id.* at 219 (quoting *Campbell*, 510 U.S. at 584). The preamble explicitly mentions the application of fair use to reproduction for purposes of criticism, comment, news reporting, teaching, scholarship, and research. 17 U.S.C. § 107 (1992).

[142] *Authors Guild*, 804 F.3d at 221.

[143] *Id.* at 222.

[144] *Id.* at 224. Notably, the Court stated: "[e]ven if the snippet reveals some authorial expression, because of the brevity of a single snippet and the cumbersome, disjointed, and incomplete nature of the aggregation of snippets made available through snippet view, we think it would be a rare case in which the searcher's interest *in the protected aspect* of the author's work would be satisfied by what is available from snippet view, and rarer still . . . that snippet view could provide a significant substitute for the purchase of the author's book." *Id.* at 224–25 (emphasis in original).

[145] *See* Dan Cohen, *What the Google Books Victory Means for Readers*, THE ATL. (Oct. 22, 2015), https://www.theatlantic.com/technology/archive/2015/10/what-the-google-books-victory-means-for-readers-and-libraries/411910/.

[146] *See generally* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

issue in *C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Advanced Media, L.P.*[147]

C.B.C. Distribution and Marketing, Inc. (CBC) is one of over three hundred businesses running online fantasy baseball leagues each season.[148] These fantasy leagues operate by connecting competing participants who build rosters of individual professional athletes.[149] The winners are determined by using those players' statistics to determine point allocation; whichever team's players had the best statistics is considered the winner for a specified period.[150] These fantasy leagues garnered significant popularity and participation: as of 2005, the number of participants was estimated to be around six million people.[151]

For almost ten years, between 1995 and 2004, CBC licensed relevant statistics from the Major League Baseball Players Association (MLBPA).[152] These license agreements included access to "the names, nicknames, likenesses, signatures, pictures, playing records, and/or biographical data of each player."[153] After the expiration of these agreements, MLB developed its own fantasy leagues featured on its website (through its "MLB Advanced Media" branch) and did not offer CBC a renewed license to the players' statistics.[154]

In response, CBC filed for a declaratory judgment that continuing to offer a fantasy baseball product would not infringe any rights of the baseball players.[155] In its answer to CBC's claim, MLB responded with its own breach of contract claim, alleging that by bringing suit CBC was violating its contractual agreement not to challenge MLB's title to publicity rights, and its agreement to cease further use of the statistics after the expiration of the contract.[156]

Though the majority of the *CBC* opinion focuses on the state-law publicity claim, the Court of Appeals for the Eighth Circuit provided relevant copyright guidance in its holding on the breach of contract claim.[157] It held that "the information used in CBC's fantasy baseball games is all readily available in the public domain, and it would be strange

---

[147] C.B.C. Distribution and Mktg, Inc. v. Major League Baseball Advanced Media, L.P., 505 F.3d 818 (8th Cir. 2007).

[148] Jeff Douglas, *Fantasy Leagues Allowed to Use MLB Stats*, WASH. POST (Aug. 8, 2006), http://www.washingtonpost.com/wp-dyn/content/article/2006/08/08/AR2006080800991.html.

[149] Matthew G. Massari, *When Fantasy Meets Reality: The Clash Between On-line Fantasy Sports Providers and Intellectual Property Rights*, 19 HARV. J. OF L. & TECH. 443, 444–45 (2006).

[150] *Id.*

[151] *Intellectual Property - Eighth Circuit Holds that the First Amendment Protects Online Fantasy Baseball Providers' Use of Baseball Statistics in the Public Domain*, 121 HARVARD L. REV. 1439 (2008). As of 2017, this number has risen to 59.3 million. *Industry Demographics: Actionable Insights and Insightful Data*, Fantasy Sports Trade Association (last visited Mar. 5, 2018).

[152] *CBC Distribution and Mktg., Inc.*, 505 F.3d at 821.

[153] *Id.*

[154] *Id.*

[155] *Id.*

[156] *Id.* at 823.

[157] *Id.*

law that a person would not have a first amendment right to use information that is available to everyone."[158] From this, along with other factors, it found there was no title at issue because there could be no exclusivity rights in the statistics.[159]

Though this case may not seem a direct analogue to web scraping litigation, it adds some relevant precedent to our understanding of the rights associated with information that is publicly available. Works that are in the public domain can include works that have expired or unrenewed copyrights, works that have been placed in the public domain intentionally by the author, or works that cannot make a claim to copyright.[160] Included in the latter category are facts and theories, including the statistics at issue in *CBC* and, likely, the employment history represented on public LinkedIn profiles.

V. ARGUMENT ANALYSIS: TRESPASS TO CHATTELS AND BREACH OF CONTRACT

Web platforms seeking relief against web scrapers frequently make various state law claims. One common state law claim in the argument is trespass to chattels.[161] This claim, borrowed from real property law, can succeed when a plaintiff can show (a) "the defendant intentionally and without authorization interfered with plaintiff's possessory interest in the computer system," and (b) "the defendant's unauthorized use proximately resulted in damage to the plaintiff."[162]

Another common state law claim is breach of contract. To be successful, commonly a plaintiff must prove this claim by showing (a) the existence of a contract, (b) fulfillment of performance requirements by the plaintiff, (c) breach by the defendant, and (d) damages suffered by the plaintiff.[163]

---

[158] *Id.* at 823–24.

[159] *Id.*

[160] *Welcome to the Public Domain*, STANFORD UNIV. LIBRARIES, https://fairuse.stanford.edu/overview/public-domain/welcome/#facts_and_theories (last visited Oct. 25, 2018).

[161] *See* Snell & Menaldo, *supra* note 17.

[162] Perry J. Viscounty, et al., *Spiders, Crawlers and Bots, Oh My: The Basics of Website Scraping*, INTELLECTUAL PROP. TODAY 31 (Oct. 2012), https://www.lw.com/thoughtLeadership/basics-of-web-scraping-IP.

[163] *Survey of the Fifty (50) States and District of Columbia Elements of a Breach of Contract Claim*, N.Y. LITIG. GUIDE (2017), http://www.nylitguide.com/survey-50-states-breach-contract-claim/.

A. Register.com, Inc. v. Verio, Inc.

An early example of web scraping litigation can be found in *Register.com v. Verio*.[164] In 2004, Register.com ("Register") was one of more than fifty companies permitted to issue domain names by the Internet Corporation for Assigned Names and Numbers ("ICANN").[165] Register issued domain names to applicants, who in turn provided contact information, including name, address, telephone number, and e-mail address.[166] As a part of its association with ICANN, Register was required to maintain this information and provide access to the information for free public review.[167] Register attached a legend when queries were made to the database, prohibiting use of the information for mass solicitations.[168]

Defendant Verio sold web site design, development, and operation services, some of which competed directly with Register's own development services.[169] In an effort to increase sales, Verio developed a web scraper that submitted queries to Register's public database of domain registrants on a daily basis, extracted the contact information, and then—in violation of Register's legend—sent sales materials to the registrants via email.[170]

Register took a number of affirmative steps to prevent Verio's access to its platform, including sending cease and desist letters, changing the legend attached to the public database in order to prohibit access for the purpose of mass solicitations, and sending follow up demands when those methods failed to elicit a positive response from Verio.[171]

For its part, Verio was forced to concede to the court that it was aware of the restrictions attached to its access of the domain registrant database.[172] The Court noted that Verio had notice of Register's terms over the period in which Verio was making queries to the database; though the legend appeared after the query was made, since Verio was making

---

[164] *See, e.g.*, Register.com, Inc. v. Verio, Inc., 356 F.3d 393 (2d Cir. 2004).

[165] *Id.* at 395. To learn more about ICANN, visit https://www.icann.org/get-started.

[166] *Register.com*, *Inc.*, 356 F.3d at 395.

[167] *Id.* Databases containing registrant information are called WHOIS databases, and they are intended to provide information on the real people behind any single internet presence, in order to, among other things, determine if a domain is available for sale, contact network administrators with technical problems, identify the party behind a domain name, contact a domain registrant for negotiating a transaction, and investigate wrongdoing online. *WHOIS Primer*, ICANN WHOIS (last updated July 2017), https://whois.icann.org/en/primer.

[168] *Register.com, Inc.*, 356 F.3d at 395.

[169] *Id.* at 396.

[170] *Id.* at 397. The legend Register applied to its query responses initially only prohibited use of WHOIS information to send solicitation emails, but, as time went on, Register changed this legend to "bar mass solicitation via direct mail, electronic mail, or by telephone." *Id.* at 398. While Verio initially contacted domain registrants via email, it eventually stopped this practice in favor of mail and telephone solicitation. *Id.*

[171] *Id.*

[172] *Id.*

multiple queries on a daily basis, it had ample notice of the terms it was violating.[173]

The Court disagreed with Verio's claim that it was not bound by Register's terms because it did not affirmatively agree to be bound.[174] The Court compared the facts in the case to circumstances under which an individual visiting an apple stand takes an apple and then sees that the apples are being offered for fifty cents; the visitor may not be liable under a contract theory for the price of the first apple, the Court reasoned, but would be liable on subsequent visits to the apple stand.[175] Under the same reasoning, Verio would not be liable for misuse under Register's terms for its first visit to the registrant database, but it would be liable for breach of contract if, with knowledge of Register's terms, it still made an unauthorized use of the information.[176]

On the trespass to chattels claim, the Court focused on the fact that web scraping accounted for "a significant portion of the capacity of Register's computer systems."[177] The Court further reasoned that permitted web scraping by Verio might have the effect of encouraging others to do the same, thus allowing for the potential that web scrapers en masse could incapacitate Register's systems.[178] In sum, the Court held for Register on multiple theories.[179]

While this case could be easily interpreted to deal a swift blow to the web-scraping industry, many aspects of the data analytics industry have changed since this 2004 ruling. For example, though Register was able to prove that Verio's automated queries made up a substantial portion of its server capacity,[180] improvements in server capacity industry-wide made it impossible for LinkedIn to allege any type of harm—financial or technological—resulting from five years of web scraping from hiQ.[181]

In light of this ever-shrinking real-world impact of web scraping, courts should consider restricting application of trespass to chattels claims altogether and applying the doctrine only in instances of malicious bot

---

[173] *Id.* at 402.
[174] *Id.* at 403.
[175] *Id.* at 401.
[176] *Id.* at 402.
[177] *Id.* at 404.
[178] *Id.*
[179] *Id. But see* Tarra Zynda, *Ticketmaster Corp. v. Tickets.com, Inc.: Preserving Minimum Requirements of Contract on the Internet*, 19 BERKELEY TECH. L.J. 495 (2004) (discussing another trespass to chattels case, *Ticketmaster Corp. v. Tickets.com*, 2003 WL 21406289 (C.D. Cal. 2003), in which the court dismissed the trespass to chattels claim after finding that the plaintiff failed to show actual damage to its property).
[180] *Id.* at 438.
[181] hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1108 (N.D. Cal. Aug. 14, 2017), *appeal filed*.

activity intended to overload servers.[182] In this regard, there is helpful case precedent that can assist in crafting the guardrails for such a rule. For example, in *eBay, Inc. v. Bidder's Edge, Inc.*, data aggregator Bidder's Edge was held liable for trespass to chattels when it crawled online auction website eBay up to 100,000 times daily, totaling up to 1.53% of the total requests received by eBay during the period of its web crawling.[183] The court did not require any physical damage to eBay's computer system, instead finding that such a significant appropriation of eBay's systems was sufficient for a trespass claim.[184] Indeed, the court found that the sheer number of calls made to eBay's site "exceeded the 'scope of consent' granted by eBay even though the website was publicly accessible."[185]

On the other end of the spectrum lies *Intel Corporation v. Hamidi*.[186] Intel Corporation (Intel) brought a claim against former Intel engineer Hamidi after Hamidi sent six mass emails to active Intel employees criticizing Intel and its employment practices.[187] These six emails, sent over a 21-month period, were sent to as many as 35,000 people.[188] The court held that cognizable trespass to chattels claims involving interference required "some additional harm to the personal property or the possessor's interests" and that such minimal use as that perpetrated by Hamidi did not impair the system in any way.[189] The court also stated that successful trespass to chattels claims alleging interference with electronic systems often involve the defendant over-burdening the system and making it unavailable to others, a situation not presented in *Intel*.[190]

These cases serve as helpful guidance for determining the role of real damage in online trespass to chattels claims. Thus, in an attempt to narrow the field of possible litigants, requiring cognizable state law trespass to chattels claims to prove significant usage of a websites computer system would limit the number of claims made under this law. So, given technological advances over time, web scraping would become less and less actionable under a trespass to chattels theory.

The breach of a contract presents a different and more challenging barrier to web scraping. Unlike the view held in *Nosal I*,[191] many courts

---

[182] *Cf.* Jane K. Winn, *Crafting a License to Know from a Privilege to Access*, 79 WASH. L. REV. 285, 306 (2004) (arguing that there should be an implied license and privilege to access information on the Internet if done so in a way that approximates individual access).

[183] eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal. 2000).

[184] *Id.* at 1071.

[185] Viscounty, *supra* note 162.

[186] Intel Corp. v. Hamidi, 71 P.3d 296 (Cal. 2003).

[187] *Id.* at 301.

[188] *Id.*

[189] *Id.* at 302.

[190] *Id.* at 304, *see also* CompuServe, Inc. v. Cyber Promotions, Inc., 962 F. Supp. 1015 (S.D. Ohio 1997).

[191] United States v. Nosal, 676 F.3d 854, 860 (9th Cir. 2012) ("Significant notice problems arise if we allow criminal liability to turn on the vagaries of private policies that are lengthy, opaque, subject to change and seldom read.").

agree with the *Register* rationale that notice of Register's terms of use bound Verio to the agreement.[192] The notable problem in recognizing rights under this theory is well-discussed in legal scholarship: by enforcing these unilateral terms, courts remove the concept of assent and mutual agreement from online interactions.[193] By adopting *Nosal I*'s reasoning, and enforcing only mutual agreements, courts can empower businesses online to make use of the wealth of information on the internet to build new insights into the human experience.

VI. BUILDING A SOLUTION

The cases presented above support the conclusion that courts are hesitant to apply liability when companies copy information from other sources and make a transformative use of that information. Courts are even more unwilling to apply liability in instances where a negative ruling could make criminals of average, well-intentioned citizens, which is the risk of broad interpretations of the CFAA. When the data that is copied is public and factual, courts become concerned about implicating constitutional rights with negative verdicts.

In order to narrow the field of potential defendants, this paper proposes an explicit carve-out rule for publicly available information. This limitation on web scraping claims would serve to limit the pool of defendants to true bad actors and allow the activities of data aggregators to continue. Since these aggregators generally pool, analyze, and create new ideas from web scraping, they are a benefit to the public at large, and, in light of technological advances permitting web platforms to handle large volumes of traffic, they are a minimal burden to the information-generating platforms.[194] Further, this carve-out rule would prevent large internet companies from using the courts for anti-competitive purposes.

---

[192] Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 403 (2d Cir. 2004) ("[The defendant] was offered access to information subject to terms of which [it was] well aware. [Its] choice was either to accept the offer of contract, taking the information subject to the terms of the offer, or, if the terms were not acceptable, to decline to take the benefits"); *see also* Mark A. Lemley, *Terms of Use*, 91 Minn. L. Rev. 459, 459 (2006) ("Ten years ago, courts required affirmative evidence of agreement to form a contract. No court had enforced a 'shrinkwrap' license, much less treated a unilateral statement of preferences as a binding agreement. Today, by contrast, more and more courts and commentators seem willing to accept the idea that if a business writes a document and calls it a contract, courts will enforce it as a contract even if no one agrees to it.").

[193] Lemley, *supra* note 192; *see also* Woodrow Hartzog, *Website Design as Contract*, 60 AM. U. L. REV. 1635 (2011), *and* Juliet Moringiello and William L. Reynolds*, From Lord Coke to Internet Privacy: The Past, Present, and Future of the Law of Electronic Contracting*, 72 MD. L. REV. 452 (2013).

[194] Maureen A. O'Rourke, *Property Rights and Competition on the Internet: In Search of an Appropriate Analogy*, 16 BERKELEY TECH. L.J. 561 (2001). O'Rourke references the *eBay, Inc. v. Bidder's Edge, Inc.*, 100 F. Supp. 2d 1058 (N.D. Cal. 2000) decision detailed above, noting the questionable rationale behind a decision that permitted some data scrapers to continue to index the

Such a rule is not without precedent in the American court system.[195] Underlying this proposed carve-out rule is a First Amendment justification.[196] In *Virginia State Board of Pharmacy v. Virginia Citizens Consumer Council, Inc.*, the Supreme Court stated that a "[p]urely factual matter of public interest may claim [First Amendment] protection."[197] The Court focused on a public policy perspective in this conclusion, stating that a "consumer's interest in the free flow of commercial information . . . may be as keen, if not keener by far, than his interest in the day's most urgent political debate."[198]

Additionally, borrowing from the utilitarian perspective underpinning copyright law, data scrapers making a transformative use of publicly available information online may be making a fair use of that material.[199] Courts should use a modified balancing test for application to data scraping specifically, to determine (a) if the copied material is sufficiently creative, or if the material was primarily factual in nature, (b) whether the primary website gained some significant financial benefit from keeping the information publicly available that might negate its complaint against a scraper, (c) whether the copied material was utilized by a bad actor to take away market share or otherwise cripple the business of the primary website or if it was utilized in the creation of a new product, and (d) whether there is a suggestion that anticompetitive motives are the driver of the complaint.

Further, the future of competition online demands limitations to litigation motivated by anticompetitive intent.[200] As many data scrapers are smaller companies than the websites targeted for crawling, a rule permitting access to publicly available factual information would efficiently limit risk for these smaller companies when they would not otherwise be able to afford lengthy antitrust litigation.[201]

Finally, a rule permitting new companies to grow from a seed based on information pulled from other websites would honor the traditions of openness upon which the Internet was built.[202] Indeed, as the Organization for Economic Co-operation and Development (OECD) stated, "[t]he Internet is fundamentally designed to be open and global,

---

auction site, while preventing other scrapers from crawling the site premised on a theory of system burden. O'Rourke, *supra* note 194 at 601.

[195] Galbraith, *supra* note 101, at 365–66.

[196] Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, Inc., 425 U.S. 748 (1976).

[197] *Id.*

[198] *Id.* at 763.

[199] *Cf.* Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

[200] O'Rourke, *supra* note 194.

[201] *Id.* at 612.

[202] *Id.* at 616.

which has enabled it to be an engine of economic growth and innovation."[203]

Thus, via new interpretations of case law or Congressional action, a legal carve-out rule protecting data scraping in limited instances would have a net benefit on the online economy.[204] This final proposal would protect data scraping from litigation if the following four conditions are met:

- The data scraper acted as a good citizen of the web, and did not seek to overburden the targeted website;[205]
- The information copied was publicly available and not behind a password authentication barrier;
- The information copied was primarily factual in nature, and the taking did not infringe on the rights—including copyrights—of another; and
- The information was used to create a transformative product and was not used to steal market share from the target website by luring away users or creating a substantially similar product.

Given these limitations, courts could better distinguish between lawsuits against bad and good actors and lawsuits motivated by anti-competitive intent and those motivated by good intentions. Further, such limitations would protect good-actor data scraping which serves to collate, analyze, and create new knowledge from knowledge gleaned from the web.

CONCLUSION

Big data is an industry experiencing explosive growth, with some estimates suggesting that it will be a $203 billion industry by the year 2020.[206] This industry is fueled by data collection; the more that data can be collected and analyzed, the more insights can be generated in order to serve business and individual consumers eager to make informed decisions about their daily lives. Because of its massive size and growth,

---

[203] *Economic and Social Benefits of Internet Openness: 2016 Ministerial Meeting on the Digital Economy Background Report*, OECD DIGITAL ECONOMY PAPERS No. 257, 5 (OECD Publishing, 2016), https://www.oecd-ilibrary.org/docserver/5jlwqf2r97g5-en.pdf?expires=1538476591&id=id&accname=guest&checksum=10DF190F63C23F9160EC1970F6DA3FFA.

[204] Though a faster response may be reached via new court interpretation, Congressional action leading to preemption of the broad slate of legal remedies currently available to online entities would be a more comprehensive protection for well-meaning data scrapers.

[205] *Googlebot*, *supra* note 19.

[206] Gil Press, *6 Predictions for the $203 Billion Big Data Analytics Market*, FORBES (Jan. 20, 2017), https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#4b8f7f502083.

big data should not be left to exist so wholly in a world of legal ambiguity. This new industry deserves a rethinking of old and inapplicable case law in order to better balance the interests of refining and generating positive insights from online information with the corporeal rights of website owners.

hiQ Labs is only one example of a business built in the age of big data, but it is an important symbol of well-intentioned web scraping and the challenges facing data collection in legal limbo. While only time can tell what a higher court will conclude about the facts of *hiQ Labs v. LinkedIn*, a decision in favor of hiQ would issue a signal of the court's recognition of the value of data analytics for the modern age.

In creating a legal protection for good actor web scrapers collecting information from publicly available sources, courts can focus their time and resources toward punishing true bad actors. This legal protection also honors the function and original purpose of the Internet as an open place built on sharing. It is through sharing information, after all, that many of the great insights of our age are discovered, new rights are created and enforced, and people throughout the world are connected.