Fordham Intellectual Property, Media and Entertainment Law Journal

Volume 29 *XXIX*
Number 1

Article 2

2019

Data Scraping as a Cause of Action: Limiting Use of the CFAA and Trespass in Online Copying Cases

Kathleen C. Riley
Fordham University School of Law, kriley22@fordham.edu

Data Scraping as a Cause of Action: Limiting Use of the CFAA and Trespass in Online Copying Cases

Cover Page Footnote
J.D. Candidate, Fordham University School of Law, 2019; B.A., Oberlin College, 2010. Thank you to Mark Patterson, my advisor, as well as the others who I consulted while brainstorming and writing this note: Fordham Professors Janet Freilich, Olivier Sylvain, and Joel Reidenberg, as well as Mark Baker and Beth Bruns of Thomson Reuters. I am grateful to the dedicated staff of the Fordham IPLJ, and in particular to Senior Research & Writing Editor Sean Corrado, for their feedback and advice. I also thank Daniel C. Reich for his near-infinite patience and my family and friends for their constant support.

Data Scraping as a Cause of Action: Limiting Use of the CFAA and Trespass in Online Copying Cases

Kathleen C. Riley*

In recent years, online platforms have used claims such as the Computer Fraud and Abuse Act ("CFAA") and trespass to curb data scraping, or copying of web content accomplished using robots or web crawlers. However, as the term "data scraping" implies, the content typically copied is data or information that is not protected by intellectual property law, and the means by which the copying occurs is not considered to be hacking. Trespass and the CFAA are both concerned with authorization, but in data scraping cases, these torts are used in such a way that implies that real property norms exist on the Internet, a misleading and harmful analogy.

To correct this imbalance, the CFAA must be interpreted in its native context, that of computers, computer networks, and the Internet, and given contextual meaning. Alternatively, the CFAA should be amended. Because data scraping is fundamentally copying, copyright offers the correct means for litigating data scraping cases. This Note additionally offers proposals for creating enforceable terms of service online and for strengthening copyright to make it applicable to user-based online platforms.

---

* J.D. Candidate, Fordham University School of Law, 2019; B.A., Oberlin College, 2010. Thank you to Mark Patterson, my advisor, as well as the others who I consulted while brainstorming and writing this note: Fordham Professors Janet Freilich, Olivier Sylvain, and Joel Reidenberg, as well as Mark Baker and Beth Bruns of Thomson Reuters. I am grateful to the dedicated staff of the *Fordham IPLJ*, and in particular to Senior Research & Writing Editor Sean Corrado, for their feedback and advice. I also thank Daniel C. Reich for his near-infinite patience and my family and friends for their constant support.

INTRODUCTION 247
I. DATA SCRAPING IN CONTEXT 251
A. Data Scraping and Internet Norms 251
B. Traditional Treatment of Data by Law 261
C. Data Scraping as a Cause of Action 265
1. Trespass to Chattels 265
2. Computer Fraud and Abuse Act 266
3. Breach of Contract 272
4. Copyright 276
5. Antitrust 278
II. DATA AND PUBLIC POLICY 279
A. Balancing Exclusive Rights in Data 280
B. Real Property Metaphors and Trespass Online 285
III. SOLUTIONS 290
A. The CFAA Should Not Be Used to Penalize Data Scraping 291
1. Data Scraping Is Not Encompassed by the Contextual Meaning of "Exceeds Authorized Access" 291
2. Data Scraping Rarely Results in a "Loss" 297
3. Ultimately, the CFAA Should Be Amended to Clarify its Meaning and Add Exceptions and Preemption Provisions 299
B. Online Contracting Can Be Improved 302
C. Copyright Offers the Correct Balance of Incentives and Remedies and Should Preempt Equivalent State Law Claims 305
1. Many Online Data Scraping Cases Are Simply Post-Feist Cases Where Copying Is Enabled by Technological Means 305
2. In Cases Where the Content Being Scraped Is Expressive, We Should Allow User-Based Services to Sue On Behalf of Their Users 307
3. Fair Use Should Be Applied Broadly to Intermediary Copying Online 309
4. Copyright Preempts Equivalent State Claims Involving Copying of Data 310
CONCLUSION 310

INTRODUCTION

Imagine you are an economist and want to use an online real estate database, Zillow,[1] to gather research data on housing prices. Rather than gathering data through clicking through the website and manually entering data into an excel spreadsheet, you write a program called a "bot" or "web crawler" to automatically find and copy relevant data.[2] Now, instead imagine that you are creating a new social media platform, and wish to allow users to pull data from their existing social media accounts to fill out their profiles. This task is also accomplished with user credentials, provided by the user, and a bot or web crawler. In both instances, you have just engaged in data scraping, which can violate the Computer Fraud and Abuse Act ("CFAA"), a federal anti-hacking statute.[3]

Data scraping, also termed screen scraping, web scraping, or web crawling, refers to the extraction of data from websites, often performed by programs termed "bots," "spiders," or "web crawlers."[4] While software applications may also be "scraped" for their data, online data scraping or web crawling retrieves data that is either publicly available or, in the case of social media websites, available to registered users.[5]

---

1 *See About Us*, ZILLOW, https://www.zillow.com/corp/About.htm [https://perma.cc/6ZV4-T3HB] (last visited Sept. 27, 2018).

2 *See* Frank Jennings & John Yates, *Scrapping Over Data: Are the Data Scrapers' Days Numbered?*, 4 J. INTELL. PROP. L. & PRAC. 120 (2009); Christopher Olston & Marc Najork, *Web Crawling*, 4 FOUND. & TRENDS IN INFO. RETRIEVAL 175, 176, (2010) http://infolab.stanford.edu/~olston/publications/crawling_survey.pdf [https://perma.cc/EV6A-689A].

3 Computer Fraud and Abuse Act of 1986, Pub. L. No. 99-474, 100 Stat. 1213 (codified as amended at 18 U.S.C. § 1030 (2012)).

4 *See* Jennings &Yates, *supra* note 2, at 120. Bots and web crawlers are typically used when a direct data link or application programming interface ("API") is unavailable. Web APIs often expose parts of an application's code, allowing programmers to build additional functionality on top of that code. *See Introduction to Web APIs*, MOZILLA, https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Client-side_web_APIs/Introduction [https://perma.cc/AKR8-6E3X] (last visited Sept. 17, 2018).

5 *See* Jennings & Yates, *supra* note 2, at 128.

Though web crawling and scraping of online data have been around since the 1990s and are essential to the functioning of Internet services,[6] recent cases like those of *Facebook, Inc. v. Power Ventures, Inc.*[7] and *hiQ Labs, Inc. v. LinkedIn Corp.*[8] have focused on a new problem, that of user data. Power Ventures was a social media aggregator that allowed its users to "keep track of a variety of social networking friends through a single program"[9] by using user account information to login to Facebook and "scrape," or automatically copy, users' Facebook data.[10] In December of 2008, having attracted a growing following,[11] and now financially backed by a major Silicon Valley capital venture film,[12] Power.com ran a promotion on Facebook,[13] which then had about 145 million active users,[14] asking Facebook users to invite their friends to Power.com.[15] When Facebook became aware of Power's promotional campaign, it sent a cease and desist letter to Power and asked Power to sign its developer terms of use agreement.[16] When Power did not comply, Facebook instituted an Internet Protocol (IP) address block, and Power changed IP addresses to circumvent the block and continued to run its campaign.[17]

Facebook then moved the dispute into the courts, filing an action against Power Ventures in the Northern District of California alleging that, by scraping its website, Power had, among

---

6 *See* Olston & Najork, *supra* note 2, at 180; *see also infra* Section I.A.

7 844 F.3d 1058 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017).

8 273 F. Supp. 3d 1099 (N.D. Cal. 2017), *appeal docketed*, No. 17-16783 (9th Cir. Sept. 6, 2017).

9 *See Power Ventures*, 844 F.3d at 1062.

10 *See id.* at 1063.

11 *Id.* at 1062.

12 *See The Man Who Stood Up to Facebook*, NPR (Oct. 13, 2016, 4:52 PM), https://www.npr.org/sections/alltechconsidered/2016/10/13/497820170/the-man-who-stood-up-to-facebook [https://perma.cc/5H5M-8UX4].

13 *Power Ventures*, 844 F.3d at 1063.

14 Ami Sedghi, *Facebook: 10 years of social networking, in numbers*, GUARDIAN (Feb. 4, 2014, 9:38 AM), https://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics [https://perma.cc/QNA8-PCSJ].

15 *Power Ventures*, 844 F.3d at 1063.

16 *Id.*

17 *Id*.

other things, infringed its copyrights and violated the CFAA.[18] Facebook specifically asserted that by accessing Facebook in violation of its terms of use, Power had accessed the website "without authorization" or "in excess of authorization" in violation of the CFAA.[19] In its decision issued in 2016, the Ninth Circuit disagreed with this reasoning, but still found that Power had violated the CFAA by virtue of continuing to scrape Facebook's website after receiving the cease and desist letter.[20] However, as a result of the lawsuit, Power had ceased operating in 2011.[21]

Determining why Facebook filed its suit requires looking beyond its stated rationale, and considering the value Facebook places on its exclusive control of user data. In its amended complaint, Facebook stated that it was "dedicated to protecting the privacy and security of its users" and accused Power of "interfering with its relationships with its users."[22] It is difficult to accept Facebook's stated concern with privacy at face value; Facebook has been widely criticized for its failure to protect the privacy of its users[23] and its founder once described privacy as a disappearing social norm.[24] In addition, Facebook itself uses web scraping to create previews of articles shared by users.[25] Facebook

---

18 First Amended Complaint at 1, Facebook, Inc. v. Power Ventures, Inc., 252 F. Supp. 3d (N.D. Cal. 2017), ECF no. 9 [hereinafter *First Amended Complaint of Facebook*].

19 *Id.* at 18.

20 *See Power Ventures*, 844 F.3d at 1067–68.

21 *See id.* at 1063.

22 *First Amended Complaint of Facebook*, *supra* note 18, at 2–3.

23 *See e.g.*, Facebook, Inc.: Analysis of Proposed Consent Order to Aid Public Comment, 76 Fed. Reg. 75,883, 75,884 (proposed Dec. 5, 2011) (proposed consent agreement); *see* Julia Angwin et al., *Facebook Doesn't Tell Users Everything It Really Knows About Them,* PROPUBLICA (Dec. 27, 2016, 9:00 AM), https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them [https://perma.cc/VX5N-CD4K]. In contrast, Power Ventures presented itself as a champion of user privacy. *See* Amended Answer and Counterclaims of Defendants Power Ventures, Inc. and Steve Vachani at 1, Facebook, Inc. v. Power Ventures, Inc., 844 F. Supp. 2d 1025 (N.D. Cal. 2012), ECF no. 54 ("Power believes in a borderless Internet where users have the right to own and control their own data.").

24 *See* Bobbie Johnson, *Privacy no longer a social norm, says Facebook founder*, GUARDIAN (Jan. 10, 2010 9:58 PM), https://www.theguardian.com/technology/2010/jan/11/facebook-privacy [https://perma.cc/XH4M-EXCG].

25 *See Facebook's New Link Previews: What You Need to Know About Creating Your Own*, MEETEDGAR (Feb. 26, 2018), https://meetedgar.com/blog/facebooks-new-link-previews-need-know-2018/ [https://perma.cc/C4V3-TWTS].

also allows third-party developers to access Facebook users' data through its developer application programming interfaces ("APIs"), and has been criticized for the ability of these developers to access extensive amounts of user data contrary to users' expectations of privacy.[26] Power's scraping of Facebook's website also does not precisely appear to have harmed Facebook's servers, as Facebook based its claim of a CFAA loss in the employee time spent discovering and attempting to block Power's activity.[27] Facebook's behavior suggests that it views its exclusive control of user data as key to maintaining its competitive advantage in its core business, advertising.[28]

Facebook is not alone in pursuing the CFAA as a means of eliminating competitors whose business models rely on data scraping. Others, including eBay,[29] LinkedIn,[30] Craigslist,[31] and Ticketmaster[32] have attempted similar legal strategies.[33] As a result, this Note considers when U.S. law should protect online data as proprietary and how courts should handle online data scraping cases. It argues that the use of the CFAA and trespass claims in data scraping cases are premised on a basic misunderstanding of how users access content online and

---

[26] *See* Elizabeth Dwoskin & Tony Romm, *Facebook's rules for accessing user data lured more than just Cambridge Analytica*, WASH. POST (Mar. 19, 2018), https://www.washingtonpost.com/business/economy/facebooks-rules-for-accessing-user-data-lured-more-than-just-cambridge-analytica/2018/03/19/31f6979c-658e-43d6-a71f-afdd8bf1308b_story.html [https://perma.cc/MY2B-QSBL]; Kevin Granville, *Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html [https://perma.cc/TLV2-LNL2]. Cambridge Analytica, which siphoned data from 50 million Facebook users in 2014 and 2015, did so through a survey app downloaded by 270,000 users and enabled by one of Facebook's developer APIs. Dwoskin & Romm, *supra*; Granville *supra*.

[27] *See* Facebook Inc.'s Supplemental Brief at 10, Facebook, Inc. v. Power Ventures, Inc., 252 F.Supp. 3d 765 (N.D. Cal. 2017) ECF No. 292.

[28] *See* Reuters, *Facebook Now Has an Almost Advertising-Only Business Model*, FORTUNE (May 5, 2017), http://fortune.com/2017/05/05/facebook-digital-advertising-business-model/ [https://perma.cc/4S29-KH3Z].

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

[30] *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017).

[31] *See* Craigslist, Inc. v. 3Taps, Inc., 964 F. Supp. 2d 1178 (N.D. Cal. 2013).

[32] *See* Ticketmaster Corp. v. Tickets.com, Inc., No. 99CV-7654, 2000 WL 1887522, at *3 (C.D. Cal. Aug. 10, 2000), *aff'd*, 2 F. App'x 741 (9th Cir. 2001).

[33] *See generally infra* Section I.C.

overbroad constructions of both claims. Rather than being interpreted using real world norms, CFAA terms like "entitle[ment]" and "authorization" should instead be interpreted in their native context, that of computers, computer security, and the Internet.

Part I discusses the context in which data scraping occurs, the traditional treatment of data by intellectual property law, and how various causes of action have played out in data scraping cases. Part II offers a framework for evaluating public policies around data, and discusses the conflict that occurs when the Internet interacts with traditional property law. Part III proposes a number of solutions, including a contextual interpretation of the CFAA and amendment.

I. DATA SCRAPING IN CONTEXT

This Part places data scraping against a backdrop of other web technologies and laws around data. Section I.A discusses online data scraping, copying, and hacking within the broader context of Internet norms and standards. Section I.B discusses the lack of protections for data under U.S. intellectual property law and attempts to protect data through other legal means, such as contract law. Section I.C explains how trespass to chattels, the CFAA, breach of contract, copyright, and antitrust claims have played out in data scraping cases.

A. Data Scraping and Internet Norms

Copying is essential to the functioning of the Internet.[34] When a user streams a song on Spotify or watches a movie on Netflix, a copy of the song or film is temporarily stored on the computer's random access memory (RAM),[35] or cached on or near the central processing unit (CPU).[36] Caching refers to the storing of data for

---

[34] Mark A. Lemley, *The Law and Economics of Internet Norms*, 73 CHI.-KENT L. REV. 1257, 1278 (1998).

[35] *See* Columbia Pictures Indus. v. Bunnell, No. CV 06-1093FMCJCX, 2007 WL 2080419, at *2 (C.D. Cal. May 29, 2007) (explaining the functioning of RAM).

[36] *See* United States v. Winkler, 639 F.3d 692, 695 (5th Cir. 2011) (explaining that "a video file is copied to a temporary internet cache when the user takes an affirmative action such as clicking on the video in order to play it. Thus . . . a video file differs as a

potential future requests to reduce the need for duplicate data transfers.[37] Web browsers cache data from websites, making short-term copies of the websites' content and front-end, or user-facing, code.[38] While storing a work in RAM has been held by courts to constitute the copying required for copyright infringement, paradoxically caching has been held to be transformative use and thus not infringement.[39] Copying is also essential to search. The Googlebot, a web crawler,[40] "fetch[es]" web pages and notes new websites and changes to existing websites to create the Google index.[41] Google also caches these websites, copying them in their entirety, and keeping a backup of the website's content in case it becomes unavailable.[42] To create Google Books, Google scanned

---

technological matter from a still photo displayed on a web site, which is downloaded automatically to an internet cache when the web page it is displayed on is loaded.").

[37] *See* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1169 (9th Cir. 2007) (explaining that "[l]ocal caching by the browsers of individual users is noncommercial, transformative, and no more than necessary to achieve the objectives of decreasing network latency and minimizing unnecessary bandwidth usage (essential to the [I]nternet).").

[38] *See Winkler*, 639 F.3d at 695–96 (defining internet caching as "where internet browser software automatically saved the content of visited websites for the purpose of reducing page-loading time if the user revisits the site" and discussing whether caching of an internet file constitutes "knowing receipt of electronic child pornography."); *see* Dig. Assurance Certification, LLC v. Pendolino, No. 6:17-cv-72-ORL-4TBS, 2017 WL 4342316, at 6 n.5 (M.D. Fla. Sept. 29, 2017) (defining a temporary file cache).

[39] *Compare* Quantum Sys. Integrators, Inc. v. Sprint Nextel Corp., 338 F. App'x 329, 337 (4th Cir. 2009) (holding that temporary storage in RAM was copying sufficient for copyright infringement) *and* DocMagic, Inc. v. Ellie Mae, Inc., 745 F. Supp. 2d 1119, 1148 (N.D. Cal. 2010) ("The Ninth Circuit has held that loading the program into the computer's RAM constitutes an act of 'copying' for the purposes of copyright law.") *with Perfect 10*, 508 F.3d at 1162 (holding that caching was not copying required to allege copyright infringement) *and* Field v. Google Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006) (holding that Google search's caching of copyrighted works was fair use).

[40] Also termed robots or spiders; a web crawler is a piece of software, and like other Internet bots, performs automated tasks. The terms web crawling and data scraping are used here interchangeably.

[41] *See Googlebot – Search Console Help*, GOOGLE, https://support.google.com /webmasters/answer/182072?hl=en [https://perma.cc/HK9V-NK46] (last visited Feb. 26, 2018). Google offers an opt-out mechanism from inclusion in Google search via robots.txt files. *See Learn About Robots.txt Files – Search Console Help*, GOOGLE, https://support.google.com/webmasters/answer/6062608?hl=en [https://perma.cc/MLA4-TXNY] (last visited Sept. 18, 2018).

[42] *See About Cached Links*, GOOGLE, https://support.google.com/websearch/answer /1687222?hl=en&ref_topic=3036132 [https://perma.cc/3Q7F-9WNP] (last visited Feb. 26, 2018).

and made machine-readable more than 20 million books, thus making copies of those books in the process.[43]

The Googlebot's web crawling activity is also a form of data scraping. Online data scraping, sometimes also termed web crawling, is common and has been around since the dawn of the Internet.[44] Data scraping is generally accomplished using bots, which—like the Googlebot—are software programmed to complete clearly defined, automated tasks.[45] Bots are so common online that they are thought to constitute the majority of Internet traffic.[46]

Like other bots, scraping bots are often programmed to be "polite," meaning that they will note and follow robot exclusion headers and robot.txt files, which indicate a website host's preference regarding the presence of bots, and limit their rate of requests so as to not impose a burden on the servers of the websites they crawl.[47] In effect, crawlers act like faster versions of the web browsers used by ordinary human users, making HTTP[48] requests to specific web addresses, or URLs.[49] One difference between an ordinary user and a data scraping bot is that a bot will not entirely render HTML,[50] only looking at plain text, whereas a human user will allow a website to "load," rendering the HTML.[51]

---

43 *See* Authors Guild v. Google, Inc., 804 F.3d 202, 208 (2d Cir. 2015).

44 *See* Olston & Najork, *supra* note 2, at 180.

45 *See* Adrienne LaFrance, *The Internet is Mostly Bots*, ATLANTIC (Jan. 31, 2017), https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/ [https://perma.cc/CN8E-6U6D].

46 *Id.*

47 *See* Olston & Najork, *supra* note 2, at 180-81; eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1061 (N.D. Cal. 2000) ("Programmers who wish to comply with the Robot Exclusion Standard design their robots to read a particular data file, 'robots.txt,' and to comply with the control directives it contains.").

48 *See* HTTP, or HyperText Transfer Protocol, is a protocol by which web clients, like browsers and bots, communicate with web servers. S*ee HTTP*, MOZILLA, https://developer.mozilla.org/en-US/docs/Web/HTTP [https://perma.cc/S2A4-WFG2] (last updated (Sept. 2, 2018, 1:10 AM)).

49 *See* Olston & Najork, *supra* note 2, at 184. A URL, or Uniform Resource Locator, is an address for a web resource, like an .html file or .jpg file. S*ee What is a URL?*, MOZILLA, https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_URL [https://perma.cc/2TXU-ZW8T] (last updated May 23, 2018, 2:41 PM).

50 *See* HTML, or HyperText Markup Language, is the very basic, simple layer of a website that describes and defines its content and layout. S*ee HTML*, MOZILLA, https://

Web crawling and data scraping are often beneficial to both services and users. Mint.com offers an example of a relatively uncontroversial use of data scraping, and one that is arguably beneficial to users.[52] To use the service, users give Mint.com their account login information for banks and other financial institutions, and if a direct data transfer or API is unavailable, Mint.com logs in on behalf of the user and scrapes the relevant data from the website, and uses the scraped data to show users a comprehensive view of their finances.[53] When a Facebook user posts a link to an article, the user also experiences a beneficial use of data scraping, as Facebook scrapes information from the article to create a preview that appears on the user's newsfeed.[54] Other examples of data scraping abound. Online retailers such as Amazon and Walmart often use bots to check their competitors' prices, a form of data scraping.[55] Uber is known to have scraped its

---

developer.mozilla.org/en-US/docs/Web/HTML [https://perma.cc/2SZN-P3QV] (last updated July 11, 2018, 7:07 AM).

[51] *See* Hartley Brody, *I Don't Need No Stinking API: Web Scraping for Fun and Profit*, HARTLEY BRODY (Feb. 3, 2017), https://blog.hartleybrody.com/web-scraping/ [https://perma.cc/RGX6-23JG].

[52] *See* Penny Crosman, *The Truth Behind the Hubbub Over Screen Scraping*, AM. BANKER (Nov. 12, 2015), https://www.americanbanker.com/news/the-truth-behind-the-hubbub-over-screen-scraping [https://perma.cc/4UTN-DV5E].

[53] *Id.* This is very similar on a technical level to the service provided by Power.com. *See* Facebook, Inc. v. Power Ventures, Inc., No. 08-cv-5780-LHK, 2013 WL 5372341, at *1 (N.D. Cal. Sept. 25, 2013). Banks have been accused of occasionally blocking Mint.com and similar services, and are sometimes said to have done so because they do not want to compete with financial aggregators. *See* Ethan Wolff-Mann, *Big Banks Are Attacking Personal Finance Apps Like Mint*, MONEY (Nov. 9, 2015), http://time.com/money/4101961/banks-attack-mint-aggregators/ [https://perma.cc/5L77-ZV26].

[54] *See News Feed Preview*, FACEBOOK, https://developers.facebook.com/docs/instant-articles/reference/feed-preview [https://perma.cc/7BS7-LBZ6] (last visited Apr. 25, 2018).

[55] *See* Khadeeja Safdar, *Retailers Try New Pricing Tricks to Battle Amazon on Black Friday*, WALL ST. J. (Nov. 20, 2017), https://www.wsj.com/articles/retailers-try-new-pricing-tricks-to-battle-amazon-on-black-friday-1511028271? [https://perma.cc/DN7M-4PHF]; Jeffrey Dastin, *Amazon trounces rivals in battle of the shopping 'bots'*, REUTERS (May 10, 2017), https://www.reuters.com/article/us-amazon-com-bots-insight/amazon-trounces-rivals-in-battle-of-the-shopping-bots-idUSKBN1860FK [https://perma.cc/FU92-W3TT].

competitors' applications and websites.[56] Web crawling, automated browsing of websites using a bot, powers Google search and Google displays scraped data in its search result previews.[57] Scraped data is used to measure and predict market behavior: economists use data scraping to gather research data,[58] and hedge funds use scraped data as an alternative data set to predict market trends.[59] Data scraping may also be used in audit testing to determine whether a service's behavior is discriminatory.[60]

In addition to their legitimate and beneficial uses, bots, the technology used for data scraping, also have malicious uses. Commentators often distinguish between "good bots" and "bad bots."[61] So-called "bad bots" include spam bots[62] and bots that impersonate real people, such as the Twitter bots used by Russia in

---

[56] Kate Conger, *Uber's Massive Scraping Program Collected Data About Competitors Around the World,* GIZMODO (Dec. 11, 2017, 10:03 PM), https://gizmodo.com/ubers-massive-scraping-program-collected-data-about-com-1820887947 [https://perma.cc/57VJ-823W].

[57] *See How Search organizes information,* GOOGLE, https://www.google.com/search/howsearchworks/crawling-indexing/ [https://perma.cc/Z7WC-KDDJ] (last visited Apr. 25, 2018).

[58] Alberto Cavallo & Roberto Rigobon, *The Billion Prices Project: Using Online Prices for Measurement and Research*, 30 J. ECON. PERSPECTIVES, no. 2, at 151, 154 (Spring 2016).

[59] Lindsay Fortado et al., *Hedge Funds See a Gold Rush in Data Mining*, FIN. TIMES (Aug. 28, 2017), https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787 [https://perma.cc/RAG4-EPJY].

[60] *E.g.*, Sandvig v. Sessions, 315 F. Supp. 3d 1, 9 (D.D.C. 2018) ("One way to determine whether members of protected classes are being discriminated against is to engage in 'outcomes-based audit testing.' [Plaintiffs] . . . are writing a computer program that will create bots—automated agents that will each browse the Internet and interact with websites as a human user might.").

[61] *See, e.g.*, Tom Ruff, *The Good, Bad and Ugly of 'Bots' Online*, THE HILL (Sept. 14, 2017 8:20 AM), http://thehill.com/opinion/technology/350536-the-good-bad-and-ugly-of-bots-online [https://perma.cc/6K54-7HZA]; *see Bot Traffic Report 2016*, INCAPSULA (Jan. 24, 2017), https://www.incapsula.com/blog/bot-traffic-report-2016.html [https://perma.cc/EGJ7-NLAT]; *see also* Branwell Moffat, *Good Bots, Bad Bots, And The Troublesome Ones In Between*, DIGITALIST MAG. (Jun. 21, 2017), http://www.digitalistmag.com/customer-experience/2017/06/21/good-bots-bad-bots-troublesome-ones-in-between-05163949 [https://perma.cc/JR5V-2DQ4]; Tamanna Mishra, *Good Bots Are the Internet's Worker Bees; Bad Bots Are Out to Get Us—Can You Tell Them Apart?*, YOURSTORY (Apr. 28, 2017), https://yourstory.com/2017/04/good-and-bad-bots/ [https://perma.cc/ALL6-DCE3].

[62] *See* Mishra, *supra* note 61.

the 2016 election[63] or the bots used by scalpers in the online ticket resale market.[64] Online service providers must also defend themselves against botnets, collections of malware-infested computers, which can be controlled remotely and used in coordinated ways.[65] Botnets are used in stealing data and passwords, attacking private and public networks, and carrying out Distributed Denial of Service (DDoS) attacks.[66] "Good bots" include the Googlebot; copyright bots, which look for infringing material online;[67] Reddit's moderator bots;[68] chat bots used for customer service;[69] and the bots used by Mint and Facebook.[70]

Data scraping itself is often said to be parasitic. Companies concerned about scraping of their websites argue, in essence, that scrapers are free riders that have misappropriated their content and harmed their relationships with their users.[71] Craigslist referred to a scraper of its website, 3Taps, as "unabashedly mass-harvesting and

---

[63] *See* Denise Clifton, *Twitter Bots Distorted the 2016 Election—Including Many Likely from Russia*, MOTHER JONES (Oct. 12, 2017, 6:00 AM), https://www.motherjones.com/politics/2017/10/twitter-bots-distorted-the-2016-election-including-many-controlled-by-russia/ [https://perma.cc/826X-BEUQ]; Victor Luckerson, *The Big, Bad Bot Problem*, RINGER (Mar. 8, 2018, 6:00 AM), https://www.theringer.com/tech/2018/3/8/17093982/twitter-bot-problem [https://perma.cc/ED3B-37XA].

[64] *See* Ticketmaster L.L.C. v. Prestige Entm't, Inc., 306 F. Supp. 3d 1164, 1170–71 (C.D. Cal. 2018); Jason Koebler, *The Man Who Broke Ticketmaster*, VICE (Feb. 10, 2017, 8:00 AM), https://motherboard.vice.com/en_us/article/mgxqb8/the-man-who-broke-ticketmaster [https://perma.cc/28LH-XRR2].

[65] *Policy Brief: Botnets*, INTERNET SOC'Y (Oct. 30, 2015), https://www.internetsociety.org/policybriefs/botnets/ [https://perma.cc/CA28-FAUS].

[66] *Id.*

[67] Mishra, *supra* note 61.

[68] *See AutoModerator*, REDDIT, https://www.reddit.com/wiki/automoderator [https://perma.cc/UC6Z-RDEL] (last visited Mar. 28, 2018).

[69] *See* Stuart Dredge, *Why Facebook and Microsoft say chatbots are the talk of the town*, GUARDIAN (Sept. 18, 2016, 6:26 AM), https://www.theguardian.com/technology/2016/sep/18/chatbots-talk-town-interact-humans-technology-silicon-valley [https://perma.cc/XW6R-MHS6].

[70] *See supra* notes 52–54.

[71] *E.g.*, Cara Bayles, *LinkedIn Tells 9th Circ. Startup's Bots Hurt Competition*, LAW360 (Mar. 15, 2018), https://www.law360.com/articles/1022804/linkedin-tells-9th-circ-startup-s-bots-hurt-competition [https://perma.cc/4G9B-QQGX] ("LinkedIn's attorney . . . told the panel during oral arguments in San Francisco that hiQ Labs Inc. was taking a 'free ride on the business LinkedIn built.'"); Jeffrey Kenneth Hirschey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, 29 BERKELEY TECH. L.J. 897, 920 (2014).

redistributing postings entrusted by craigslist users," and argued that this "undermine[d] the integrity of local craigslist communities, ultimately harming both craigslist and its users."[72] LinkedIn referred to a scraper of its website, hiQ, as "flagrantly violat[ing] LinkedIn's privacy commitments and member controls, and subvert[ing] the expectations of LinkedIn members."[73] In another case, Facebook said a scraper and social media aggregator, Power Ventures, interfered in its relationship with its users and induced users to provide their Facebook contacts' email addresses.[74]

As a result of these concerns, companies often attempt to limit scraping of their websites through their terms and conditions. Zillow's terms of use prohibit automated queries, specifically "screen and database scraping, spiders, robots, [and] crawlers[,]" while making an exception for search engines to the extent their scraping is fair use "allowed by applicable copyright law."[75] Etsy's terms of use specifically state, "Don't Steal Our Stuff," and assert that users "agree not to 'crawl,' 'scrape,' or 'spider' any page" of its website.[76] Facebook,[77] LinkedIn,[78] Twitter,[79] eBay,[80]

---

[72] First Amended Complaint at 2, Craigslist, Inc. v. 3Taps, Inc., 964 F. Supp. 2d 1178 (N.D. Cal. 2013) ECF no. 9.

[73] *See* LinkedIn Corporation's Supplemental Brief in Opposition to Plaintiff's Motion for a Preliminary Injunction at 2, hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017), ECF no. 50.

[74] *See* First Amended Complaint of Facebook at 2, Facebook, Inc. v. Power Ventures, Inc., 252 F. Supp. 3d 765 (N.D. Cal. 2017), ECF no. 9.

[75] *Terms of Use*, ZILLOW, https://www.zillow.com/corp/Terms.htm [https://perma.cc/7AJP-FFU9] (last visited Feb. 23, 2018). Zillow, however, offers direct downloads of certain research data.

[76] *See Terms of Use – Our House Rules,* ETSY, https://www.etsy.com/legal/terms-of-use/ [https://perma.cc/P58A-EX5B] (last visited Feb. 27, 2018). While Etsy offers an API, its terms also contain a provision prohibiting automated scraping and bots. *See API Terms of Use*, ETSY (Jan. 8, 2017), https://www.etsy.com/legal/api [https://perma.cc/58AG-98CQ].

[77] *Statement of Rights and Responsibilities*, FACEBOOK (Jan. 30, 2015), https://www.facebook.com/terms.php [https://perma.cc/2RPY-PKL8] ("[y]ou will not collect users' content or information, or otherwise access Facebook, using automated means (such as harvesting bots, robots, spiders, or scrapers) without our prior permission").

[78] *See User Agreement,* LINKEDIN, https://www.linkedin.com/legal/user-agreement [https://perma.cc/92KL-CJ5C] (last visited Apr. 8, 2018) ("You agree that you will not: . . . [d]evelop, support or use software, devices, scripts, robots, or any other

Craigslist,[81] TripAdvisor,[82] Expedia,[83] IMDB,[84] Yelp,[85] Hotels.com,[86] and Kickstarter[87] all prohibit scraping and bots in their terms and conditions, usually with an exception for bots which have been granted express permission.

---

means or processes (including crawlers, browser plugins and add-ons, or any other technology or manual work) to scrape the Services or otherwise copy profiles and other data from the Services").

[79] *Twitter Terms of Service*, TWITTER, https://twitter.com/en/tos [https://perma.cc/UH3K-MR89] (last visited Feb. 23, 2018) ("crawling the Services is permissible if done in accordance with the provisions of the robots.txt file, however, scraping the Services without the prior consent of Twitter is expressly prohibited").

[80] *eBay User Agreement*, EBAY (Nov. 1, 2017), https://pages.ebay.com/help/policies/user-agreement.html [https://perma.cc/TE8V-CXVF] ("In connection with using or accessing the Services you will not; . . . use any robot, spider, scraper, data mining tools, data gathering and extraction tools, or other automated means to access our Services for any purpose, except with the prior express permission of eBay").

[81] *Terms of Use*, CRAIGSLIST (Dec. 29, 2017), https://www.craigslist.org/about/terms.of.use.en [https://perma.cc/78RD-QEU7] ("[y]ou agree not to copy/collect CL content via robots, spiders, scripts, scrapers, crawlers, or any automated or manual equivalent").

[82] *TripAdvisor Website Terms, Conditions and Notices*, TRIPADVISOR (Feb. 15, 2018), https://tripadvisor.mediaroom.com/us-terms-of-use [https://perma.cc/CB9D-75SY] ("you agree not to; . . . access, monitor or copy any content or information of this Website using any robot, spider, scraper or other automated means or any manual process for any purpose without our express written permission").

[83] *Website Terms of Use*, EXPEDIA (Feb. 21, 2018), https://www.expedia.com/p/info-other/legal.htm [https://perma.cc/KHN9-G7XL] ("[y]ou agree not to; . . . access, monitor or copy any content or information of this Website using any robot, spider, scraper or other automated means or any manual process for any purpose without our express written permission").

[84] *IMDb Conditions of Use*, IMDB, https://www.imdb.com/conditions?ref_=ft_cou [https://perma.cc/TGR5-CA7B] (last visited Feb. 27, 2018) ("[y]ou may not use data mining, robots, screen scraping, or similar data gathering and extraction tools on this site, except with our express written consent as noted below").

[85] *Terms of Service*, YELP (Nov. 27, 2012), https://www.yelp.com/static?p=tos [https://perma.cc/2BMT-DLPC] ("[y]ou also agree not to, and will not assist, encourage, or enable others to . . . [u]se any robot, spider, site search/retrieval application, or other automated device, process or means to access, retrieve, scrape, or index any portion of the Site or any Site Content").

[86] *Terms and Conditions*, HOTELS.COM (Jan. 18, 2018), https://www.hotels.com/customer_care/terms_conditions.html [https://perma.cc/ZC9E-XMEC] (last visited Feb. 27, 2018) ("you agree not to . . . access, monitor or copy any content or information of this Website using any robot, spider, scraper or other automated means or any manual process for any purpose without our express written permission").

[87] *Terms of Use*, KICKSTARTER (Oct. 19, 2014), https://www.kickstarter.com/terms-of-use?ref=global-footer [https://perma.cc/DWD5-JLWU].

Companies also often take technical measures to prevent scraping of their websites. Measures to detect bots and scrapers include monitoring website traffic and looking for unusual traffic spikes, users completing repetitive tasks too quickly, and other behavior inconsistent with a human user.[88] Another common defensive measure is CAPTCHA, a "Completely Automated Public Turing test to tell Computers and Humans Apart," an automated method of distinguishing bots from humans by asking users to complete a task, specifically a task for which humans typically outperform computers.[89] Internet Protocol (IP) address blocking is another common means of countering bots,[90] although its effectiveness is limited.[91] IP addresses are often dynamically assigned, meaning that they change over time.[92] In addition, IP addresses can be "spoofed" to create anonymity, though this practice is more dubious.[93]

Though companies often oppose scraping of their sites, scraping differs from what is generally considered a security

---

88 *See* JonasCz, *A Guide to Preventing Webscraping*, GITHUB (July 30, 2017), https://github.com/JonasCz/How-To-Prevent-Scraping [https://perma.cc/KL37-AJ3S]; *How to Prevent Getting Blacklisted While Scraping*, SCRAPEHERO, https://www.scrapehero.com/how-to-prevent-getting-blacklisted-while-scraping/ [https://perma.cc/U2SX-56WZ] (last visited Feb. 27, 2018).

89 Deb Amlen, *What the Heck Is That?*: *CAPTCHA*, N.Y. TIMES (Feb. 26, 2018), https://www.nytimes.com/2018/02/26/crosswords/what-the-heck-is-that-captcha.html [https://perma.cc/CS6Y-ALZE].

90 *See, e.g.*, Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1067 (9th Cir. 2016)*, cert. denied*, 138 S. Ct. 313 (2017) (noting that Power Ventures circumvented Facebook's IP address block); Craigslist Inc. v. 3Taps Inc., 964 F. Supp. 2d 1178, 1180–81 (N.D. Cal. 2013) (noting that Craigslist blocked 3Taps' IP address).

91 *See* JonasCz, *supra* note 88; Bill Brenner, *Scraper and Bot Series – When Good Bots Go Bad*, AKAMAI: SIRT ALERTS BLOG (Mar. 10, 2016 9:00 AM), https://blogs.akamai.com/2016/03/scaper-and-bot-series—-when-good-bots-go-bad.html [https://perma.cc/AK42-7W4C].

92 *See Static vs. dynamic IP addresses,* GOOGLE FIBER, https://support.google.com/fiber/answer/3547208?hl=en [https://perma.cc/MXU7-ZQDX].

93 *See* Farha Ali, *IP Spoofing*, 10 INTERNET PROTOCOL J. 2, 3 (2007), https://www.cisco.com/c/en/us/about/press/internet-protocol-journal/back-issues/table-contents-38/104-ip-spoofing.html [https://perma.cc/3P9F-54K8]. IP addresses are assigned by Internet Service Providers (ISPs). As a result, with cooperation from an ISP, an IP address can potentially help identify a specific computer and its user. Individuals concerned for their anonymity or privacy will sometimes spoof their IP addresses, sometimes in order to conduct illicit activity. However, spoofing may be used for legitimate purposes as well, such as performance testing of websites.

breach or "black hat" hacking.[94] When software engineers and web developers design applications and websites, they must anticipate a number of common hacks and avoid creating vulnerabilities that allow these hacks to occur. Common hacks include cookie poisoning, hidden field manipulation, parameter tampering, cross-site scripting, exploiting backdoor and debug options, HTTP response splitting, and SQL injection.[95] These attacks often involve manipulation of HTTP header information, such as falsifying authentication information.[96] Many attacks also rely on phishing, which can be thought of as any attempt to extract information from a user using deceptive practices and social engineering, such as copycat websites or fraudulent emails and texts.[97] Scrapers, by contrast, generally act like normal users.[98]

---

[94] "White hat" hackers, by contrast, are encouraged to find and report flaws, or bugs, in web application code, and are sometimes paid a finder's fee for doing so. *See* Nick Bilton, *Hackers with Enigmatic Motives Vex Companies*, N.Y. TIMES (July 25, 2010), https://www.nytimes.com/2010/07/26/technology/26security.html [https://perma.cc/7B79-6NVJ].

[95] *See The Dirty Dozen: Preventing Common Application-Level Hack Attacks*, IBM (Dec. 2007), ftp://ftp.software.ibm.com/software/rational/web/whitepapers/r_wp_dirtydozen.pdf [https://perma.cc/E3LT-V6L7]; Sumit Siddharth & Pratiksha Doshi, *Five Common Web Application Vulnerabilities*, SYMANTEC CONNECT (Apr. 27, 2006), https://www.symantec.com/connect/articles/five-common-web-application-vulnerabilities [https://perma.cc/YMC2-2584].

[96] *See* Akash Mahajan, *Introduction to HTTP Response Headers for Security*, INFOSEC INST. (Aug. 13, 2012), http://resources.infosecinstitute.com/http-response-headers/#gref [https://perma.cc/7KX8-AA8Y].

[97] *See Phishing*, FTC (July 2017), https://www.consumer.ftc.gov/articles/0003-phishing [https://perma.cc/EMY3-D6RT].

[98] An ordinary user visits a website by clicking a link or typing a Uniform Resource Locater (URL), or web address, into a browser. *See What is a URL?*, ORACLE, https://docs.oracle.com/javase/tutorial/networking/urls/definition.html [https://perma.cc/N5NK-9SVG] (last visited Mar. 26, 2018). The browser is a client making a one-way HTTP request to a web server over TCP/IP and the server responds with a one-way response. *See an Overview of HTTP*, MOZILLA, https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview [https://perma.cc/4VVK-KVL7] (last updated June 24, 2018); *A Typical HTTP session*, MOZILLA (Mar. 26, 2018), https://developer.mozilla.org/en-US/docs/Web/HTTP/Session [https://perma.cc/R5J6-VPT6]. The first HTTP request a human user makes to a website, via a browser, is typically a "GET" request, which only retrieves data. *See HTTP Request Methods*, MOZILLA, https://developer.mozilla.org/en-US/docs/Web/HTTP/Methods [https://perma.cc/25UX-S7KP] (last visited May 13, 2018). To scrape a website, a bot will make numerous HTTP "GET" requests, parse the website's code, and store the information its programmer has built it to retrieve. *See* Hartley Brody, *Web Scraping References: A*

B. Traditional Treatment of Data by Law

Historically, U.S. intellectual property law has not protected pure information or facts. The Copyright Act of 1976 protects "original works of authorship fixed in any tangible medium of expression," but this protection does not extend to ideas, concepts, discoveries, or facts.[99] In *Feist Publications v. Rural Telephone Service Co.*, Feist copied telephone directory data from Rural for its own, more expansive telephone directory, and Rural sued Feist claiming copyright infringement.[100] Rural argued that its telephone directory was a "compilation,"[101] which the Copyright Act defines as "a work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship."[102] Because copyright is about expression and originality, the Supreme Court stated that "sweat of the brow" does not entitle a work to copyright protection.[103] The copyright to a compilation of facts, or data, is "thin," because it involves minimal creativity and originality; the specific selection and arrangement of those facts may be copyrightable, but the raw facts and data themselves are not.[104]

In addition, in some instances, works created through copying have been found to be fair use. Fair use, originally a judicial

---

Simple Cheat Sheet for Web Scraping with Python, HARTLEY BRODY (Feb. 18, 2017), https://blog.hartleybrody.com/web-scraping-cheat-sheet/ [https://perma.cc/E2U9-CTQU]. In essence, a scraping bot is simply a web client, similar to a web browser.

99 17 U.S.C. § 102 (2012); Feist Publ'n, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 344–45 (1991) ("The most fundamental axiom of copyright law is that '[n]o author may copyright his ideas or the facts he narrates.'").

100 *Feist*, 499 U.S. at 342–44.

101 *Id.* at 341.

102 17 U.S.C. § 101 (2012). A compilation includes collective works, "such as a periodical issue, anthology, or encyclopedia, in which a number of contributions, constituting separate and independent works in themselves, are assembled into a collective whole." *Id.*

103 *Feist*, 499 U.S. at 352–54. The EU has made the opposite policy choice through the sui generis right, which rewards substantial investment in data collection. *See* Council Directive 96/9, 1996 O.J. (EC).

104 *See Feist*, 499 U.S. at 348–49; David E. Shipley, *Thin but Not Anorexic: Copyright Protection for Compilations and Other Fact Works*, 15 J. INTELL. PROP. L. 91, 95 (2007). In contrast, the work on an expressive work, such as a novel, is said to be "thick." *See* Fleener v. Trinity Broad. Network, 203 F. Supp. 2d 1142, 1149 (C.D. Cal. 2001).

doctrine, is enshrined in Section 107 of the Copyright Act.[105] Use of a copyrighted work may be transformative if it "adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message."[106] Because copying is so fundamental to the functioning of the Internet[107] and is often necessary for interoperability of software applications,[108] courts have sometimes found that online services that involve extensive copying—such as search engines—are fair use.[109] In *Authors Guild v. Google, Inc.*, the Second Circuit held that making digital copies of books to enable search, and providing short "snippets" of those books, was such a transformative use.[110] However, in the same court's decision in *Fox News Network, LLC v. TVEyes, Inc.*, it noted that such transformative use does not extend to redistribution of content that denies the copyright holder revenue.[111]

Despite copyright's limited protection of data, certain kinds of data are specifically protected by law and certain means of obtaining data are limited by law. Various statutes protect medical and financial data, but other types of data receive more limited protection or none at all.[112] For user data, the U.S. offers what is termed a "sectoral model," meaning that "Congress passes

---

105 *See* 17 U.S.C. § 107 (2012) (factors to be considered in determining fair use are: "(1) purpose and character of the use . . . ; (2) nature of the copyrighted work; (3) amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for the value of the copyrighted work)."

106 *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

107 *See supra* Section I.A.

108 *See supra* Section I.A; *see also* Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510, 1522–23 (9th Cir. 1992), *as amended* (Jan. 6, 1993).

109 *See, e.g.*, Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1166 (9th Cir. 2007).

110 *See* Authors Guild v. Google, Inc., 804 F.3d 202, 216–17 (2d Cir. 2015).

111 *See* Fox News Network, LLC v. TvEyes, Inc., 883 F.3d 169, 174 (2d Cir. 2018) ("TVEyes's re-distribution of Fox's audiovisual content serves a transformative purpose in that it enables TVEyes's clients to isolate from the vast corpus of Fox's content the material that is responsive to their interests, and to access that material in a convenient manner. But because that re-distribution makes available virtually all of Fox's copyrighted audiovisual content—including all of the Fox content that TVEyes's clients wish to see and hear—and because it deprives Fox of revenue that properly belongs to the copyright holder, TVEyes has failed to show that the product it offers to its clients can be justified as a fair use.").

112 *See* Joel R. Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?*, 44 FED. COMM. L.J. 195, 198, 210 (1992).

narrowly tailored laws that barely infringe on the marketplace's role of self-regulation, and the Federal Trade Commission (FTC) and the Department of Commerce monitor businesses relying primarily on industry standards[.]"[113] Many patented inventions include a database as a claim element,[114] and a new type of database, if found to be novel and non-obvious, could theoretically be patentable.[115] Data that is kept secret, with measures taken to protect the secrecy of the data, may be offered protection under trade secret law.[116]

Negotiated data license agreements are an increasingly prevalent means of protecting data through contract law,[117] but contracts of adhesion—non-negotiated, form contracts, used in reoccurring transactions where the parties have unequal bargaining power—are also often enforced online.[118] In *ProCD, Inc. v.*

---

[113] *See* Bradyn Fairclough, *Privacy Piracy: The Shortcomings of the United States' Data Privacy Regime and How to Fix It*, 42 J. CORP. L. 461, 463 (2016).

[114] *See, e.g.*, U.S. Patent No. 9,900,355 (filed Oct. 5, 2016); U.S. Patent No. 9,900,353 (filed Oct. 5, 2016); U.S. Patent No. 9,900,339 (filed Feb. 27, 2017); U.S. Patent No. 9,900,162 (filed Nov. 11, 2015). A search of U.S. patents run on February 23, 2018 for patents containing the word "database" in at least one claim came up with over 100,000 search results.

[115] *See* 35 U.S.C. §§ 102–103 (2012).

[116] *See* 18 U.S.C. § 1839(3) (2012); Molly Hubbard Cash, *Keep It Secret, Keep It Safe: Protecting Trade Secrets by Revisiting the Reasonable Efforts Requirement in Federal Law*, 23 J. INTELL. PROP. L. 263, 285–86 (2016); *see also* Heather Roark Parker, *Trade Secrets and Patent Protection: The Unlikely Power Couple Under the AIA*, 32 SYRACUSE J. SCI. & TECH. L. 1, 19–20 (2016).

[117] *See* Daniel Glazer et. al., *Data as IP and Data License Agreements*, THOMSON REUTERS: PRACTICAL LAW, https://1.next.westlaw.com/Document /I5f5951a21c8a11e38578f7ccc38dcbee/View/FullText.html? [https://perma.cc/7ZQ3-PLC6] (last visited Feb. 23, 2018); *see Community Data License Agreement*, LINUX FOUNDATION, https://cdla.io/ [https://perma.cc/Z367-CA45] (last visited Feb. 23, 2018); *see Bloomberg and Twitter Sign Data Licensing Agreement*, BLOOMBERG (Sept. 16, 2015), https://www.bloomberg.com/company/announcements/bloomberg-and-twitter-sign-data-licensing-agreement/ [https://perma.cc/8CSX-LWDF].

[118] *E.g.*, Fagerstrom v. Amazon.com, Inc., 141 F. Supp. 3d 1051, 1056 (S.D. Cal. 2015), *aff'd,* Wiseley v. Amazon.com, Inc., 709 F. App'x 862 (9th Cir. 2017). *Compare* William J. Condon, Jr., *Electronic Assent to Online Contracts: Do Courts Consistently Enforce Clickwrap Agreements?*, 16 REGENT U. L. REV. 433 (2004) (discussing the enforceability of clickwrap and browserwrap contracts and arguing that both types of contracts should be enforceable), *with* Juliet M. Moringiello & 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, 470–71 (2013) (arguing that courts are increasingly applying traditional notions of notice and assent to online contracts).

Zeidenberg—which, like *Feist*, involved telephone directory data—the Seventh Circuit found that a "shrinkwrap" license agreement on software protected a compilation of data where copyright could not.[119] In the Internet era, shrinkwrap licenses, which referred to licenses that go into effect when the plastic wrapping is taken off a CD-ROM case,[120] have been replaced by clickwrap and browserwrap licenses.[121] A clickwrap license is an agreement that goes into effect when a website user is offered terms and conditions and clicks "I agree," while browserwrap licenses are terms and conditions that a user is said to have agreed to by virtue of using an application ("app") or website.[122] Terms of use that appear on a website, such the Zillow and Etsy terms of use,[123] are an example of a browserwrap license. Clickwrap and browserwrap licenses share a defining characteristic: consumers almost never read them.[124]

Companies rely on other legal means for protecting their data as well. Copyright claims have not had particularly success in data scraping cases, and as a result proprietors of social media and other user-based websites have attempted to prohibit third parties from copying their data under the Computer Fraud and Abuse Act (CFAA), state hacking statutes, and the related tort of trespass to chattels.[125]

---

119 *See* ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1455 (7th Cir. 1996).

120 *Id.* at 1449.

121 *See* Moringiello & Reynolds, *supra* note 118, at 461–62.

122 *Id.* at 465–67.

123 *See supra* Section I.A.

124 *See generally* Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services* (Information, Communication & Society, Working Paper, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465 [https://perma.cc/H2RB-LLYG] (last revised Aug. 18, 2018); hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1107 (N.D. Cal. 2017) ("It is unlikely, however, that most users' *actual* privacy expectations are shaped by the fine print of a privacy policy buried in the User Agreement that likely few, if any, users have actually read."); *see* Aaron Smith, *Half of Online Americans Don't Know what a Privacy Policy Is*, FACT TANK BLOG (Dec. 4, 2014), http://www.pewresearch.org/fact-tank/2014/12/04/half-of-americans-dont-know-what-a-privacy-policy-is/ [https://perma.cc/85Q3-DRJT].

125 *See infra* Section I.C.

C. *Data Scraping as a Cause of Action*

When companies find the data on their websites and applications scraped, they may turn to a number of legal causes of action in search of a remedy. These causes of action commonly include: (1) trespass to chattels; (2) violations of the Computer Fraud and Abuse Act (CFAA) and state computer crime statutes; (3) breach of contract; (4) copyright infringement and violations of the Digital Millennium Copyright Act (DMCA).[126] Recently, antitrust claims have also come into play in data scraping cases.

1. Trespass to Chattels

"A trespass to a chattel may be committed by intentionally . . . using or intermeddling with a chattel in the possession of another [,]" when "the chattel is impaired as to its condition, quality, or value, or . . . the possessor is deprived of the use of the chattel for a substantial time[.]"[127] States may have their own formulations of trespass specific to intangible property. For example, California common law specifically acknowledges trespass to chattels as "encompass[ing] unauthorized access to a computer system where (1) defendant intentionally and without authorization interfered with plaintiff's possessory interest in the computer system; and (2) defendant's unauthorized use proximately resulted in damage to plaintiff."[128]

Trespass to chattels is commonly argued in data scraping cases, under the theory that a defendant's scraping interfered with a plaintiff's use of its website and servers by consuming intangible resources such as network and server capacity.[129] These harms are often acknowledged to be minimal. In *eBay v. Bidder's Edge*, a

---

[126] Other causes of action that are less common include trademark infringement and other Latham Act related civil actions, misappropriation, unfair competition, intentional interference with contractual relationship, interference with prospective business advantage, fraud, Sherman Act claims, and trade secret-related claims.

[127] *See* RESTATEMENT (SECOND) OF TORTS §§ 217(b), 218(b)–(c) (AM. LAW. INST. 1965); Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 438 n.58 (2d Cir. 2004).

[128] *See* Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 980 (N.D. Cal. 2013) (internal quotation marks omitted).

[129] *See id.*; eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1064–66 (N.D. Cal. 2000); *see also* Snap-on Bus. Sols. Inc. v. O'Neil & Assocs., 708 F. Supp. 2d 669, 679 (N.D. Ohio 2010); *Register.com*, 356 F.3d at 404.

cause about aggregation of auction data, eBay stated that "the load on its servers resulting from [Bidder's Edge's] web crawlers represents between 1.11% and 1.53% of the total load on eBay's listing servers."[130] eBay's argument was partially metaphorical: eBay argued that Bidder's Edge's activities "should be thought of as equivalent to sending in an army of 100,000 robots a day to check the prices in a competitor's store."[131] The court disagreed with the metaphor, but decided that allowing the scraping to continue "unchecked . . . would *encourage* other auction aggregators" to crawl eBay's website, which had the potential reduce its performance.[132] In addition, despite noting that courts rarely grant preliminary injunctions based on ongoing trespasses to chattels, the court decided to rely on cases related to real property as instructive, again comparing eBay's website to a physical auction house.[133] As a result, the court granted eBay a preliminary injunction which prohibited Bidder's Edge from using any robot or crawler on eBay's website without written authorization.[134]

2. Computer Fraud and Abuse Act

The CFAA is the primary legal means by which companies offering web-based services attempt to block scraping of their applications. In 1984, Congress passed its first computer-crime statute, the Counterfeit Access Device and Computer Fraud and Abuse Act ("CADCFAA"),[135] which was soon amended to create the CFAA,[136] and later expanded to create a civil cause of action

---

130 *See eBay*, 100 F. Supp. 2d at 1064. Contrast this with *Snap-on*, 708 F. Supp. 2d at 679–80, where the scraper program apparently caused enough of a traffic spike that Snap-On's website crashed. *Snap-On* has a very different fact pattern than the other data scraping cases. In *Snap-On*, Snap-On had a negotiated agreement with a third-party, Mitsubishi. After a contract dispute with Snap-On regarding data ownership and portability, Mitsubishi turned to the defendant O'Neil to scrape Snap-On's database, which was password-protected. Snap-On won at trial.

131 *See eBay*, 100 F. Supp. 2d at 1065.

132 *Id.* at 1066 (emphasis added).

133 *Id.* at 1067.

134 *Id.* at 1073.

135 Counterfeit Access Device and Computer Fraud and Abuse Act (CADCFAA) of 1984, Pub. L. No. 98-473, § 2102(a), 98 Stat. 2190 (codified as amended at 18 U.S.C. § 1030 (2012)).

136 Computer Fraud and Abuse Act of 1986, Pub. L. No. 99-474, 100 Stat. 1213 (codified as amended at 18 U.S.C. § 1030 (2012)).

within the CFAA.[137] While the CFAA was originally envisioned as an anti-hacking or computer trespass statute,[138] the language of the CFAA is much broader. In data scraping cases, an individual typically runs afoul of the CFAA's civil provisions when he "intentionally accesses a computer without authorization or exceeds authorized access" and obtains information from that computer.[139] The computer must be a "protected computer," a computer involved in interstate commerce or communication,[140] or any computer connected to the Internet.[141] The term "exceeds authorized access" is defined to mean "to access a computer with authorization and to use such access to obtain or alter information in the computer that the accessor is not entitled so to obtain or alter [.]"[142] In addition, to be eligible for a civil remedy, the violation must have resulted in certain harms, the most expansive being a "loss" to one or more persons of at least $5,000, occurring during any one-year period.[143] The term loss is defined to include reasonable costs to a victim, such as the cost of "responding to an offense, conducting a damage assessment, and restoring the data, program, system, or information to its condition prior to the offense and any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service[.]"[144]

The key phrases of the CFAA are "without authorization" and "exceeds authorized access," which have been interpreted in numerous ways by federal courts and legal scholars. These

---

137 *See* S. REP. NO. 104-357, at 11–12 (1996). The civil provision was added in the 1994 amendment.

138 *See* H.R. REP. NO. 98-894, at 8–11 (1984) ("Compounding this is the advent of the activities of so-called 'hackers' who have been able to access (trespass into) both private and public computer systems, sometimes with potentially serious results . . . . For example, the motion picture 'War Games' showed a realistic representation of the automatic dialing and access capabilities of the personal computer."); S. REP. NO. 99-432, at 7 (1986) ("Second, section 2(b) will clarify the present 18 U.S.C. 1030(a)(3), making clear that it applies to acts of simple trespass against computers belonging to, or being used by or for, the Federal Government.").

139 18 U.S.C. § 1030(a)(2)(C) (2012).

140 18 U.S.C. § 1030(e)(2)(B) (2012).

141 *E.g.*, United States v. Nosal, 676 F.3d 854, 859 (9th Cir. 2012) (noting that "protected computer" refers to "all computers with Internet access").

142 18 U.S.C. § 1030(e)(6) (2012).

143 18 U.S.C. §§ 1030(c)(4)(A)(i)(I) & (g) (2012).

144 18 U.S.C. § 1030(e)(11) (2012).

approaches to interpretation of the CFAA include: (1) the agency approach; (2) the contract approach; (3) the plain meaning approach; (4) the trespass approach; and (5) the code-based approach. The agency approach is sometimes applied in employment contexts, and looks to whether a user violated the duty of loyalty he owes to his employer under agency law.[145] It rarely applies in the data-scraping context, which often involves parties with no legal relationship.[146] The contract approach looks to whether a user of a website or application violated its terms and conditions,[147] which often results in enforcement of browserwrap and clickwrap contracts.[148] The First, Fifth, and Eleventh Circuits have adopted this approach, along with several district courts.[149] The plain meaning approach looks to the plain meaning of "exceeds authorized access," and has been adopted by the Fourth and Ninth Circuits.[150] Accordingly, in *United States v. Nosal*, a criminal CFAA case, the Ninth Circuit considered the meaning of the phrase, "not entitled so to obtain or alter" and applying the rule of lenity, found that "the CFAA does not extend to violations of use restrictions" but instead targets "unauthorized procurement or alteration of information[.]"[151] The computer trespass approach acknowledges that the CFAA was intended as a computer trespass statute,[152] and suggests imposing elements similar to those of trespass to determine whether an individual has exceeded

---

145 *See* Lee Goldman, *Interpreting the Computer Fraud and Abuse Act*, 13 U. PITT. J. TECH. L. POL'Y 1, 15 (2012).

146 *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058 (9th Cir. 2016)*, cert. denied*, 138 S. Ct. 313 (2017); Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962 (N.D. Cal. 2013); CouponCabin LLC v. Savings.com, Inc., No. 2:14-CV-39-TLS, 2017 WL 83337 (N.D. Ind. Jan. 10, 2017). Contrast these with a case like *United States v. Nosal*, 676 F.3d 854 (9th Cir. 2012), which involved employment, or agency, relationship.

147 *See* Goldman*, supra* note 145, at 6–7; Patricia L. Bellia, *A Code-Based Approach to Unauthorized Access Under the Computer Fraud and Abuse Act*, 84 GEO. WASH. L. REV. 1442, 1455–56 (2016).

148 *See, e.g.*, Sw. Airlines Co. v. Farechase, Inc., 318 F. Supp. 2d 435, 439–40 (N.D. Tex. 2004) (even if plaintiff's use agreement was not an enforceable contract, defendant knew that its terms prohibited scraping and bots); *see* Craigslist, Inc. v. Naturemarket, Inc., 694 F. Supp. 2d 1039, 1056–57 (N.D. Cal. 2010) (in part because defendants violated website terms of use, plaintiff alleged a claim for CFAA violation).

149 *See* Goldman, *supra* note 145, at 7–8.

150 *See id.* at 13.

151 *See* United States v. Nosal, 676 F.3d 854, 863–64 (9th Cir. 2012).

152 *See* H.R. REP. NO. 98-894, at 10–11 (1984).

authorized access.[153] Under the code-based approach, originally proposed by Orin S. Kerr,[154] a user acts without authorization only when he circumvents code that regulates access to the protected computer.[155] However, Kerr has suggested that it is valid to view the CFAA as a computer trespass statute.[156]

The application of the CFAA in data scraping cases highlights the inconsistencies in interpretation of the statute. *CollegeSource, Inc. v. AcademyOne, Inc.*, decided in the Eastern District of Pennsylvania, offers a narrow interpretation of the CFAA applied to data scraping.[157] CollegeSource accused AcademyOne of scraping and republishing college course catalogs and course information, which CollegeSource had collected and archived.[158] In evaluating CollegeSource's CFAA cause of action, the court noted that CollegeSource's materials were available to the public and that AcademyOne had not engaged in hacking.[159] Although CollegeSource had sent AcademyOne a cease and desist letter,[160] the court rejected CollegeSource's argument that by violating its terms of use, AcademyOne had "exceeded authorization," noting that it had previously found those same terms of use unenforceable under contract law.[161]

Notice is increasingly important in CFAA cases, as illustrated by *Craigslist Inc. v. 3Taps Inc.*,[162] which was decided in the

---

153 *See* Josh Goldfoot & Aditya Bamzai, *A Trespass Framework for the Crime of Hacking*, 84 GEO. WASH. L. REV. 1477, 1483 (2016) (advocates a three-element test for unauthorized access: "(1) the entry (or access) violates an express or implied prohibition; (2) the violator knew, or should have known, of the prohibition's existence; and (3) the prohibition is material or related to the underlying policy of trespass").

154 *See* Orin S. Kerr, *Cybercrime's Scope: Interpreting "Access" and "Authorization" in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596 (2003).

155 *See* David J. Rosen, *Limiting Employee Liability Under the CFAA: A Code-Based Approach to "Exceeds Authorized Access,"* 27 BERKELEY TECH. L.J. 737, 747 (2012); Bellia, *supra* note 147, at 1457.

156 Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1153–54 (2016) [hereinafter *Computer Trespass*].

157 *See generally* CollegeSource, Inc. v. AcademyOne, Inc., No. CIV.A. 10-3542, 2012 WL 5269213 (E.D. Pa. Oct. 25, 2012), *aff'd*, 597 F. App'x 116 (3d Cir. 2015).

158 *Id.* at *1.

159 *Id.* at *4, 14.

160 *Id.* at *5.

161 *Id.* at *15.

162 964 F. Supp. 2d 1178 (N.D. Cal. 2013).

Northern District of California after *Nosal*. In *3Taps*, the court defined the issue as "whether Craigslist had the power to revoke, on a case-by-case basis, the general permission it granted to the public to access the information on its website."[163] The court noted that Craigslist "affirmatively communicated its decision to revoke 3Taps' access through its cease-and-desist letter and IP blocking efforts."[164] In *Facebook, Inc. v. PowerVentures, Inc.*, decided in 2016, the Ninth Circuit took up this reasoning, holding that the permission of Facebook's users to access their accounts on Facebook's website was "not sufficient to constitute authorization after Facebook issued the cease and desist letter."[165]

CouponCabin LLC v. Savings.com, Inc., a case involving coupons and coupon codes, advances an even broader view of CFAA liability.[166] In *CouponCabin*, the court noted that CFAA liability "may exist in certain situations where a party's authorization to access electronic data—including publicly accessible electronic data—has been affirmatively rescinded or revoked."[167] However, even though defendant Linfield Media was not given actual notice that its access was unauthorized, the court found that CouponCabin's technological blocking measures acted as constructive notice.[168]

The CFAA is often used as the basis for injunctions of scraping activity.[169] Under a traditional preliminary injunction analysis, a court may consider the public interest, but this does not always occur. *In Citizens Information Associates, LLC v. Justmugshots.com*, a case about scraping of mugshots and arrest

---

163 *Id.* at 1182.

164 *Id.* at 1184.

165 *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1068 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017).

166 *See* CouponCabin LLC v. Savings.com, Inc., No. 2:14-CV-39-TLS, 2017 WL 83337 (N.D. Ind. Jan. 10, 2017).

167 *Id.* at *3.

168 *Id.*

169 *See* Citizens Info. Assocs., LLC v. Justmugshots.com, No. 1-12-CV-573-LY, 2012 WL 12874898, at *2 (W.D. Tex. Dec. 18, 2012) ("To obtain a preliminary injunction, Citizens must demonstrate: (1) a substantial likelihood of success on the merits of its claim; (2) a substantial threat of irreparable injury or harm if the injunction is not granted; (3) that the threatened injury to Citizens outweighs any harm the injunction might cause to D'Antonio; and (4) that granting the injunction will not disserve the public interest.").

records, the court stated that because increased public access to this information is arguably in the public interest, it could not grant an injunction.[170] In *EF Cultural Travel BV v. Explorica, Inc.*, a case where a scraper bot was used to gather price information, which was then used to undercut a competitor's prices, the First Circuit reviewed a district court's grant of a preliminary injunction.[171] The case was complicated by the fact that Explorica, the scraper, was founded by EF's former employees, who then used proprietary tour codes to assist in the scraping of EF's website.[172] The court noted the problem of assessing a CFAA "loss," and determined that the effort and time spent *assessing* the potential damage to EF's computer systems constituted a loss, suggesting that a more narrow definition of loss would "reward sophisticated intruders."[173] However, one of the former EF employees had arguably breached a confidentiality agreement—it is unclear why the court ruled on the CFAA and not this likely breach of contract.[174]

Critics describe the CFAA as flawed, overbroad, and criminalizing ordinary behavior.[175] Commentators note that the CFAA does not define several of its key terms, including "access" and "authorization,"[176] and that that the use of the term "computer" makes little sense in the Internet context, when content is stored on

---

170 *Id.* at *2–3.

171 EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 580 (1st Cir. 2001).

172 *Id.* at 579–80.

173 *Id.* at 585.

174 *Id.* at 583–84.

175 *E.g.*, Tim Wu, *Fixing the Worst Law in Technology*, NEW YORKER (Mar. 18, 2013), https://www.newyorker.com/news/news-desk/fixing-the-worst-law-in-technology [https://perma.cc/AJ7M-J6JT]; Tiffany Curtis, Note and Comment, *Computer Fraud and Abuse Act Enforcement: Cruel, Unusual, and Due for Reform*, 91 WASH. L. REV. 1813 (2016); Samantha Jensen, *Abusing the Computer Fraud and Abuse Act: Why Broad Interpretations of the CFAA Fail*, 36 HAMLINE L. REV. 81, 83–84 (2013); Jonathan Keim, *Updating the Computer Fraud and Abuse Act*, 16 ENGAGE, no. 3, at 31, 32–33 (Oct. 2015), https://fedsoc.org/commentary/publications/updating-the-computer-fraud-and-abuse-act-1 [https://perma.cc/VR6G-58DK].

176 *E.g.*, Andrew T. Hernacki, *A Vague Law in A Smartphone World: Limiting the Scope of Unauthorized Access Under the Computer Fraud and Abuse Act*, 61 AM. U. L. REV. 1543, 1554 (2012).

many computers and servers via the cloud.[177] The CFAA has been commonly applied in instances where no actual hacking has occurred. In one case, the CFAA was found to apply to an employee who deleted all data on his employer-provided computer before returning it,[178] and in another case, to a local official who forwarded another's emails without permission.[179] In another instance, prosecutors attempted to use the CFAA to criminalize the creation of a fake social media profile in violation of terms of service, which was found to violate the void-for-vagueness doctrine.[180] Password sharing by a former employee was also found to be a CFAA violation.[181] Noting the Supreme Court's decision in *Packingham v. North Carolina*, which holds that "to foreclose access to social media altogether is to prevent the user from engaging in the legitimate exercise of First Amendment rights,"[182] the American Civil Liberties Union (ACLU) has even suggested that the CFAA is unconstitutional because it chills exercise of free speech by making it illegal to conduct certain kinds of online research.[183]

3. Breach of Contract

A contract is fundamentally a promise recognized by law as enforceable if broken, or breached.[184] As courts have noted, the Internet "has not fundamentally changed the principles of contract."[185] In order to be binding, contracts online still require "a 'meeting of the minds' and a manifestation of 'mutual assent.'"[186]

---

177 Amanda B. Gottlieb, Note, *Reevaluating the Computer Fraud and Abuse Act: Amending the Statute to Explicitly Address the Cloud*, 86 FORDHAM L. REV. 767, 778–79 (2017).

178 *See* Int'l Airport Ctrs., L.L.C. v. Citrin, 440 F.3d 418, 420–21 (7th Cir. 2006).

179 *See* Steinbach v. Vill. of Forest Park, No. 06-C-4215, 2009 WL 2605283, at *1, 5–6 (N.D. Ill. Aug. 25, 2009).

180 *See* United States v. Drew, 259 F.R.D. 449, 457–467 (C.D. Cal. 2009).

181 *See* United States v. Keys, 703 F. App'x 472, 474 (9th Cir. 2017); *accord* United States v. Nosal, 844 F.3d 1024, 1040–41 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 314 (2017).

182 137 S. Ct. 1730, 1737 (2017).

183 *See* Sandvig v. Sessions, 315 F.Supp.3d 1, 32 (D.D.C. 2018).

184 *See* RESTATEMENT (SECOND) OF CONTRACTS § 1 (AM. LAW INST. 1981).

185 *See* Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 403 (2d Cir. 2004).

186 *See* Van Tassell v. United Mktg. Grp., LLC, 795 F. Supp. 2d 770, 789 (N.D. Ill. 2011).

However, many courts are increasingly willing to enforce contracts of adhesion that appear online, such as clickwrap and browserwrap agreements.[187] In evaluating these types of contracts, courts typically evaluate whether the "structure of the contract or website gives users reasonable notice of the terms or requires express assent."[188] Online users, however, rarely read terms of service and website privacy policies,[189] in part because few websites either situate their contracts in a manner that encourages users to read them or offer terms that can be easily read.[190]

Breach of contract arguments are not uncommon in data scraping cases.[191] An early example of this trend can be found in *Register.com, Inc. v. Verio, Inc.*, decided in 2000.[192] In *Register.com*, the Southern District of New York granted an injunction against Verio, a scraper of WHOIS data, under Register.com's breach of contract claim.[193] The terms of use in the case were a browserwrap agreement, published on the "home page of [Register.com's] Internet website."[194] Though the terms of use stated that "[b]y submitting this query, you agree to abide by these

---

187 *See* Erin Canino, *The Electronic "Sign-in-Wrap" Contract: Issues of Notice and Assent, the Average Internet User Standard, and Unconscionability*, 50 U.C. DAVIS L. REV. 535, 541 (2016); *accord* AT&T Mobility LLC v. Concepcion, 563 U.S. 333, 346–47 (2011).

188 *See* Canino, *supra* note 187, at 541.

189 *See* Obar & Oeldorf-Hirsch, *supra* note 124; Debra Cassens Weiss, *Chief Justice Roberts Admits He Doesn't Read the Computer Fine Print*, A.B.A. J. (Oct. 20, 2010, 12:17 PM), http://www.abajournal.com/news/article/chief_justice_roberts_admits_he_doesnt_read_the_computer_fine_print [https://perma.cc/7JBJ-KSYP].

190 *See, e.g.*, *Terms of Use*, TICKETMASTER, http://www.ticketmaster.com/h/terms.html [https://perma.cc/JS9D-92YZ] (last visited Apr. 11, 2018) (The main page of Ticketmaster's website states, at the very bottom of the page, "By continuing past this page, you agree to our Terms of Use." The terms of use are in small font, 12 pixels, which is roughly equivalent to 9 point font, and are around 5000 words long, which would take an average reader 25 minutes to read at a speed of 200 words per minute.); *see* Canino, *supra* note 187, at 554–55; Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 HARV. L. REV. 1173, 1178–80 (1983).

191 *See, e.g.*, Craigslist, Inc. v. Kerbel, No. C-11-3309 EMC, 2012 WL 3166798, at *14 (N.D. Cal. Aug. 2, 2012) (holding that plaintiff stated a claim for breach of contract). *See generally* Register.com, Inc. v. Verio, Inc., 126 F. Supp. 2d 238 (S.D.N.Y. 2000), *aff'd as modified*, 356 F.3d 393 (2d Cir. 2004).

192 126 F. Supp. 2d at 238.

193 *Id.* at 243, 248 (WHOIS data is information about domain names, and falls under the purview of ICANN, which assigns and regulates domain names).

194 *Id.* at 245.

terms[,]" the defendant argued that "it was not asked to click on an icon indicating that it accepted the terms."[195] The court asserted that by submitting a WHOIS query, Verio agreed to be bound by the terms of use, forming a clickwrap agreement that Verio then breached.[196] In a similar but more recent case, *Craigslist, Inc. v. Kerbel,* the Northern District of California held that Craigslist had alleged a valid contract, stating uncritically that Kerbel assented to the terms of use "each time he access[ed] the website."[197]

In other data scraping cases, courts analyzed whether a user had actual or constructive notice of a website's terms of use in order to determine whether a contract was formed. In *DHI Group, Inc. v. Kent*, a case about online job boards, the court noted that while browserwrap agreements rarely give consumers actual or constructive notice, it was plausible that the defendant Oilpro had constructive notice because its own website contained the same provisions prohibiting use of automated means to download data.[198] The court also confined its conclusion to cases where both parties were sophisticated businesses using browserwrap agreements on their websites.[199] In *College Source, Inc. v. AcademyOne, Inc.*, the court considered a "Copyright and Disclaimer" notice located on CollegeSource's PDF catalogs and website, and noting the lack of "essential elements of contract formation," granted summary judgment in favor of the defendant on the plaintiff's breach of contract claim.[200] In other cases, the issue of whether terms of use were a browserwrap contract was raised, but left undecided.[201]

Snap-On Business Solutions Inc. v. O'Neil & Associates, Inc. is an unusual data scraping case where a party arguably breached a

---

195 *Id.* at 248 (internal quotation marks omitted).

196 *See id.*

197 Craigslist, Inc. v. Kerbel, No. C–11–3309 EMC, 2012 WL 3166798, at *14 (N.D. Cal. Aug. 2, 2012) (argued over the auto-posting and reposting of classified ads).

198 *See* DHI Grp., Inc. v. Kent, No. H–16–1670, 2017 WL 4837730 (S.D. Tex. Oct. 26, 2017).

199 *Id.*

200 *See* CollegeSource, Inc. v. AcademyOne, Inc., No. 10–3542, 2012 WL 5269213, at *10 (E.D. Pa. Oct. 25, 2012).

201 *E.g.*, Sw. Airlines Co. v. Farechase, Inc., 318 F. Supp. 2d 435, 441 (N.D. Tex. 2004).

negotiated agreement rather than contract of adhesion, and where the plaintiff suffered real harm to its servers and temporary loss of service.[202] Snap-On and Mitsubishi negotiated a series of agreements in which Mitsubishi contributed data such as parts catalogs to Snap-On, which provided electronic part catalogs for clients in the automotive and heavy equipment industries.[203] In particular, Mitsubishi agreed in the Web Hosting Agreement to not use its access to Snap-On's website for any purpose other than administering user names and passwords to authorized users.[204] However, after Mitsubishi asked Snap-on for a copy of its data with Snap-on's enhancements, which included hot spots, links, and photographs, Snap-on refused, and Mitsubishi began to discuss creating a new database with O'Neil & Associates.[205] O'Neil then offered to create a scraping tool to retrieve the data from Snap-On's system, and received thirty login credentials from Mitsubishi.[206] However, O'Neil's scraping—which was performed without rate limiting, or slowing down of requests so as not to overwhelm a server—created enormous spikes in Snap-On's website traffic that caused the website to crash.[207] It is unclear why Snap-On sued O'Neil for breach of contract instead of Mitsubishi, but the court found that there was a sufficient dispute of material fact to preclude summary judgment on Snap-On's CFAA, breach of contract, and copyright claims.[208]

In analyzing CFAA claims, courts often consider whether a user violated a website's terms of use as part of the "exceeds authorized access" analysis. In *Southwest Airlines Co. v. Farechase, Inc.*, the court stated that regardless of whether Southwest's use agreement was an enforceable contract, the defendant had constructive knowledge of the terms, and thus was aware its access was unauthorized.[209] In *Craigslist, Inc. v.*

---

202 708 F. Supp. 2d 669 (N.D. Ohio 2010).
203 *Id.* at 672.
204 *Id.* at 673.
205 *Id.*
206 *Id.* at 674.
207 *Id.* at 675.
208 *Id.* at 678, 683, 686.
209 *See* Sw. Airlines Co. v. Farechase, Inc., 318 F. Supp. 2d 435, 439–40 (N.D. Tex. 2004).

Naturemarket, Inc., the court stated that, "Plaintiff alleged that Defendants accessed its computers in violation of the TOUs, and therefore without authorization" and thus granted the plaintiff default judgment on its CFAA claim.[210] In contrast, in *Cvent, Inc. v. Eventbrite, Inc.*, the court noted that Cvent's terms of use were "not displayed on the website in any way in which a reasonable user could be expected to notice them."[211] The court noted that the link to access the terms was "buried at the bottom of the first page, in extremely fine print" and that the terms themselves were "several pages long."[212]

4. Copyright

In data scraping cases, copyright infringement is often alleged and dismissed.[213] In *Ticketmaster Corp. v. Tickets.com, Inc.*, a case about scraping of ticket and event information, the court denied the plaintiff's request for a preliminary injunction on its copyright claim, noting that "[t]he major difficulty with many of plaintiff's theories and concepts is that it is attempting to find a way to protect its expensively developed basic information from what it considers a competitor and it cannot do so."[214] However, parties who have registered a copyright on the entirety of their website are sometimes allowed to proceed with such claims, based on the notion that the organization and arrangement of the information on a website is copyrightable.[215] In the same *Ticketmaster* case, in a

---

210 694 F. Supp. 2d 1039, 1057 (N.D. Cal. 2010); *accord* CouponCabin LLC v. Savings.com, Inc., No. 2:14-CV-39-TLS, 2017 WL 83337, at *5 (N.D. Ind. Jan. 10, 2017) (holding that because both parties were sophisticated businesses, the browserwrap agreement was enforceable).

211 739 F. Supp. 2d 927, 932 (E.D. Va. Sept. 14, 2010).

212 *Id.* at 933.

213 *See, e.g.*, eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1072 (N.D. Cal. 2000) ("BE argues that the trespass claim . . . 'is similar to eBay's originally filed but now dismissed copyright infringement claim'"); *see Naturemarket*, 694 F. Supp. 2d at1056; *see also* Ticketmaster Corp. v. Tickets.com, Inc., No. 99CV7654, 2000 WL 1887522, at *3 (C.D. Cal. Aug. 10, 2000), *aff'd*, 2 F. App'x 741 (9th Cir. 2001) [hereinafter *Tickets.com I*]; *see also* Allure Jewelers, Inc. v. Ulu, No. 1:12CV91, 2012 WL 4322519 (S.D. Ohio Sept. 20, 2012) (dismissing Allure's copyright claim based on late registration).

214 *Tickets.com I*, 2000 WL 1887522, at *3.

215 *See, e.g.*, DHI Group, Inc. v. Kent, No. CV H-16-1670, 2017 WL 4837730, at *4 (S.D. Tex. Oct. 26, 2017) ("Since Oilpro alleges the entire website, including the page

later decision, the court—having accepted that Ticketmaster's website was copyrightable—evaluated Tickets.com's copying and determined that its spidering activity was fair use.[216] In *Craigslist Inc. v. 3Taps Inc.*, the court noted that Craigslist's allegation of a compilation copyright hinged on its exclusive licenses of its users' posts and held that its terms of use did not involve the writing necessary to grant an exclusive license.[217] In addition, scrapers, by virtue of circumventing an IP block or traffic monitoring software, are sometimes found to have potentially violated the DMCA.[218]

Copyright is said to be in tension with contract law, and courts have sometimes applied the doctrine of preemption to resolve the conflict.[219] Section 301 of the Copyright Act states that "all legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright . . . are governed exclusively by this title . . . .[N]o person is entitled to any such right or equivalent right in any such work under the common law or statutes of any State."[220] In data scraping cases, claims of copyright

---

layout and organization of the member profile pages, is part of its registered copyright and that DHI published this information on its own website, Oilpro has stated a plausible claim for copyright infringement."); *see Naturemarket*, 694 F. Supp. 2d at 1056 (granting default judgment on Craigslist's copyright claim); *see also* Craigslist, Inc. v. Kerbel, No. C-11-3309 EMC, 2012 WL 3166798, at *9 (N.D. Cal. Aug. 2, 2012); Facebook, Inc. v. Power Ventures, Inc., No. C 08-5780 JF (RS), 2009 WL 1299698, at *4 (N.D. Cal. May 11, 2009). However, the "look and feel" of a website itself is generally not copyrightable, nor is the underlying CSS; the HTML and CSS together may be copyrightable if sufficiently expressive. *See* Media.net Advert. FZ-LLC v. NetSeer, Inc., 156 F. Supp. 3d 1052, 1062, 1065–67 (N.D. Cal. 2016).

216 *See* Ticketmaster Corp. v. Tickets.com, Inc., No. CV997654HLHVBKX, 2003 WL 21406289, at *5–6 (C.D. Cal. Mar. 7, 2003) (favorably comparing the copying to reverse engineering, noting that it was temporary and intended to extract public facts, and observed the lack of infringing material on Tickets.com's website.) [hereinafter *Tickets.com II*].

217 Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 973–74 (N.D. Cal. 2013). Facebook's copyright claim was also voluntarily dismissed in its case via Fed. R. Civ. P 41(A)(1), though the exact reasons why are unclear. *See* Facebook, Inc. v. Power Ventures, Inc., No. 08-CV-05780-LHK, 2017 WL 3394754, at *2 (N.D. Cal. Aug. 8, 2017) ("On February 18, 2011, Judge Ware granted the parties' stipulation to dismiss Facebook's DMCA claim, copyright and trademark infringement claims, and claims for violations of California Business and Professions Code Section 17200.").

218 *E.g.*, *DHI Grp.*, 2017 WL 4837730, at *5.

219 *See* Guy A. Rub, *Copyright Survives: Rethinking the Copyright-Contract Conflict*, 103 VA. L. REV. 1141, 1159 (2017).

220 *See* 17 U.S.C. § 301(a) (2012).

preemption are rarely raised and, where they are raised, often denied.[221]

One data scraping case where preemption was found to apply is *Cvent*, where the district court found that copyright preempted the plaintiff's Virginia Computer Crimes Act (VCCA) claim.[222] The VCCA is similar to the CFAA, but specifically states as an element that violator must obtain "property of services by false pretenses," or embezzle or commit larceny, or convert "the property of another."[223] Because the plaintiff's allegation of the VCCA violation was based on copying, the court found that copyright preempted the claim.[224] In *Southwest Airlines*, the court found that Southwest's misappropriation claim for "fare, route, and scheduling information" was similarly preempted by copyright law.[225]

5. Antitrust

Following the *Facebook v. Power Ventures* decision, a scraper of LinkedIn's website, hiQ, sought a declaratory judgment that it was not violating the CFAA or other laws by scraping the site.[226] hiQ's business model depends on collecting data from LinkedIn and analyzing it to provide services to employers, including a service called "Keeper" aimed at alerting employers of employees who are at risk of being recruited away.[227] In May 2017, LinkedIn sent hiQ a cease and desist letter telling hiQ to stop scraping its website and noting the terms of its user agreement, which prohibit

---

221 *See, e.g.*, Snap-on Bus. Sols. Inc. v. O'Neil & Assocs., Inc., 708 F. Supp. 2d 669, 680 (N.D. Ohio 2010) (holding that copyright did not preempt plaintiff's trespass to chattels claim); *see3Taps, Inc.*, 942 F. Supp. 2d at 977 (holding that copyright did not preempt plaintiff's breach of contract claim).

222 *See* Cvent, Inc. v. Eventbrite, Inc., 739 F. Supp. 2d 927, 934–35 (E.D. Va. 2010).

223 *Id.* at 934 ("The elements of a violation of the VCCA are that the defendant (1) uses a computer or computer network; (2) without authority; and (3) either obtains property or services by false pretenses, embezzles or commits larceny, or converts the property of another.").

224 *Id.* at 935.

225 *See* Sw. Airlines Co. v. Farechase, Inc., 318 F. Supp. 2d 435, 440–41 (N.D. Tex. 2004).

226 *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1103–04 (N.D. Cal. 2017), *appeal filed*, No. 17-16783 (9th Cir. Sept. 6, 2017).

227 *Id.* at 1104.

scraping.[228] In addition to alleging violations of the CFAA and Digital Millennium Copyright Act (DMCA), LinkedIn also claimed that hiQ had committed trespass to chattels by scraping its website,[229] and expressed concern about users' privacy.[230] hiQ, in turn, argued that LinkedIn's decision to block its access to data "was made for an impermissible anticompetitive purpose—namely that it want[ed] to monetize this data itself with a competing product."[231] The court stated that, "the Sherman Act prohibits companies from leveraging monopoly power to 'foreclose competition or gain a competitive advantage, or to destroy a competitor.'"[232] Noting LinkedIn's market dominance and previous contradictory positions regarding user privacy taken in other litigation, the court found that the issues raised by hiQ supported granting a preliminary injunction.[233] The injunction prohibited LinkedIn from blocking hiQ's access to its website while the litigation proceeded.[234] LinkedIn has since appealed the decision to the Ninth Circuit, the same circuit that decided *Facebook v. Power Ventures*.

II. DATA AND PUBLIC POLICY

> "Technology is neither good nor bad; nor is it neutral[.]"[235]

This Part discusses public policy justifications for both intellectual property and traditional property law and how these justifications should be applied to data. Section II.A suggests a framework with which to evaluate the success of a public policy around data and argues that copyright offers the correct balance of incentives. Section II.B discusses the use of trespass and trespass

---

228 *Id.*
229 *Id.* at 1104–05.
230 *Id.* at 1118.
231 *Id.* at 1117.
232 *Id.* at 1118 (citing Otter Tail Power Co. v. United States, 410 U.S. 366, 377 (1973)).
233 *Id.*
234 *Id.* at 1120.
235 *See* Melvin Kranzberg, *Technology and History: "Kranzberg's Laws,"* 27 TECH. & CULTURE, no. 3, at 547, 554 (July 1986).

metaphors in application CFAA, and argues that these claims rely on misleading analogies that treat the Internet as a physical place.

A. Balancing Exclusive Rights in Data

Intellectual property law involves many trade-offs, generating incentives to create and invent and resulting in occasional tragedies in the failure to reward "sweat of the brow." The U.S. Constitution explicitly endorses a utilitarian approach to intellectual property, giving Congress the power to "promote the Progress of Science and useful Arts" by offering time-limited exclusivity to authors and inventors.[236] Under this justification, patents incentivize invention; copyrights incentivize creative expression; trademarks incentivize investment and quality;[237] and trade secrets disincentivizes certain types of unfair competition.[238] The same justifications for areas of intellectual property limit their reach: utility patents must be useful;[239] copyrighted materials must be original works of authorship and fixed in tangible medium of expression;[240] federally-registered trademarks must be distinctive and used in interstate commerce;[241] and trade secrets must be secret.[242] A utilitarian analysis considers whether the benefits of a policy outweigh its costs, and whether a policy successfully achieves its stated objectives. When considering possible protection of data, this entails examining the incentives created by a policy and its social and economic consequences.

Lack of protection or exclusivity in certain areas of intellectual property law can create negative spaces where innovation and

---

[236] *See e.g.,* U.S. CONST. art. I, § 8, cl. 8; Adam D. Moore, *A Lockean Theory of Intellectual Property*, 21 HAMLINE L. REV. 65 (1997); Eric E. Johnson, *Intellectual Property and the Incentive Fallacy*, 39 FLA. ST. U. L. REV. 623, 624 (2012). Lockean labor theory and personhood theory offer competing justifications for intellectual property, but neither justification is acknowledged as valid by U.S. law. EU law, by contrast, recognizes both theories as valid.

[237] *See* 1 MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 3:1 (5th ed. 2018).

[238] *See* Michael Risch, *Why Do We Have Trade Secrets?*, 11 MARQ. INTELL. PROP. L. REV. 1, 14–15 (2007).

[239] 35 U.S.C. § 101 (2012).

[240] 17 U.S.C. § 102 (2012).

[241] 15 U.S.C. § 1051 (2012).

[242] 18 U.S.C § 1839(3)(A) (2012).

competition thrive.[243] The fashion and restaurant industries both lack comprehensive intellectual property protection for their participants' creations,[244] but the industries continue to prosper.[245] The open source movement,[246] and lack of intellectual property protection for programming languages[247] are both essential to software development, a thriving industry.[248] Lack of protection of information and data can also create tragedies where the law fails to reward an individual's investment in research and data collection while allowing others exploit the fruits of their labor.

Miller v. Universal City Studios, Inc. offers a compelling example of a tragedy caused by copyright's failure to protect facts.[249] The plaintiff, Gene Miller, wrote a nonfiction book about a kidnapping in which the victim was buried alive, later adapted by Universal into a screenplay and TV movie, without crediting or compensating Miller.[250] Because the book was based in fact, despite Miller's year-and-a-half of original research, the information conveyed by the book was not copyrightable.[251] The same "negative spaces" which enable competition and innovation are also home to such tragedies. No matter how much time and effort is expended to perfect a recipe, the lists of ingredients and

---

243 *See* Elizabeth L. Rosenblatt, *A Theory of IP's Negative Space*, 34 COLUM. J. L. & ARTS 317, 349 (2011).

244 *See id.* at 325–28.

245 *See e.g.* Imran Amed, et al., *The State of Fashion 2018,* MCKINSEY & CO. (Nov. 2017), https://www.mckinsey.com/~/media/mckinsey/industries/retail/our%20insights/renewed%20optimism%20for%20the%20fashion%20industry/the-state-of-fashion-2018-final.ashx [https://perma.cc/JH92-S45E]; Hudson Riehle, *Restaurant Industry 2017 and Beyond*, NAT'L REST. ASS'N (May 20, 2017), https://www.restaurant.org/Downloads/PDFs/Events-Groups/Fast-Casual-Show-State-of-Industry-Presentation-Ma.pdf [https://perma.cc/KSS7-37FP].

246 *See* Marcus Maher, *Open Source Software: The Success of an Alternative Intellectual Property Incentive Paradigm*, 10 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 619, 695 (2000).

247 *See* Elizabeth G. Lowry, *Copyright Protection for Computer Languages: Creative Incentive or Technological Threat?*, 39 EMORY L.J. 1293, 1306 (1990).

248 *See The $1 Trillion Economic Impact of Software*, BSA (June 2016), http://softwareimpact.bsa.org/pdf/Economic_Impact_of_Software_Report.pdf [https://perma.cc/3LEU-AGEZ].

249 *See* Miller v. Universal City Studios, Inc., 650 F.2d 1365 (5th Cir. 1981).

250 *Id.* at 1367–68.

251 *Id.* at 1372.

procedures contained in recipes are ultimately not copyrightable.[252] Copying and knock-off brands are rampant in the fashion industry, free-riding on the hard work of the original designers,[253] enabled by the failure of copyright to protect clothing.[254]

On the opposite side of the spectrum, too many exclusive rights can create a different sort of tragedy. When multiple stakeholders are able to exclude others from use of a resource, a tragedy of the anticommons emerges.[255] In a tragedy of the anticommons, property becomes locked into inefficient uses because exclusive rights holders create barriers that prevent optimal use.[256] In his 2003 article, *Cyberspace as a Place and the Tragedy of the Anticommons*, Dan Hunter suggests that the network resources that constitute the Internet are a form of commons, and in the early days of the Internet, the public had free and open access to websites; however, as time passed, websites increasingly became enclosed.[257]

Another concern that arises from allocating too many exclusive rights is one of competition and barriers to entry. It is often said that Internet's openness and decentralization was essential to its early development.[258] But the Internet's value as a communication mechanism, as well the value of widely-adopted user-based platforms like Facebook, LinkedIn, and Reddit, is derived partially

---

[252] *See, e.g.*, Publications Int'l, Ltd. v. Meredith Corp., 88 F.3d 473, 480 (7th Cir. 1996) ("The identification of ingredients necessary for the preparation of each dish is a statement of facts. There is no expressive element in each listing; in other words, the author who wrote down the ingredients for 'Curried Turkey and Peanut Salad' was not giving literary expression to his individual creative labors.").

[253] *See* Katy Steinmetz, *The Knockoff Economy: How Copying Hurts—and Helps—Fashion*, TIME (Sept. 10, 2012), http://style.time.com/2012/09/10/the-knockoff-economy-how-copying-hurts-and-helps-fashion/ [https://perma.cc/W5AE-T6KZ].

[254] *See, e.g.*, Whimsicality, Inc. v. Rubie's Costume Co., 891 F.2d 452, 455 (2d Cir. 1989) ("We have long held that clothes, as useful articles, are not copyrightable.").

[255] *See* Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621, 623–24 (1998).

[256] *See* Dan Hunter, *Cyberspace as Place and the Tragedy of the Digital Anticommons*, 91 CAL. L. REV. 439, 444 (2003) [hereinafter *Cyberspace as Place*].

[257] *Id.* at 511.

[258] *See* Lawrence Lessig, *Cyberspace's Architectural Constitution*, Lecture given at www9 in Amsterdam, Netherlands (June 12, 2000), https://cyber.harvard.edu/works/lessig/www9.pdf [https://perma.cc/2J5K-KTJB].

from network effects.[259] Network effects are phenomena that occur when the value of a good or service increases as the number of people who use it increases,[260] and have been said to create barriers to entry[261] and encourage monopoly power in technology spaces.[262] Because users contribute content to user-based services, over time they are said to develop a type of "collective inertia" tying them to the platform.[263] Antitrust law exists in tension with intellectual property law, as by its nature, intellectual property law offers limited monopolies and antitrust law prohibits monopolization.[264]

As a result, any policy creating property rights around data must balance incentives to create and innovate against potential creation of too many property rights, which can stifle innovation and competition. To the extent they are expressive and original, websites and web applications are works of authorship fixed in a tangible medium of expression.[265] *The New York Times*' copyright on its newspaper is no less strong because it is simultaneously published in print and online.[266] But, copyright does not protect facts.[267] While a story is created by its author, a fact exists in the world, and like a scientific principle, is only discovered.[268]

---

[259] *See* Mark A. Lemley, *Antitrust and the Internet Standardization Problem*, 28 CONN. L. REV. 1041, 1045–47 (1996) [hereinafter *Antitrust & Internet Standardization*].

[260] *See* Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479, 481 (1998). Network effects are also sometimes called positive network externalities.

[261] *See* Gregory J. Werden, *Network Effects and Conditions of Entry: Lessons from the Microsoft Case*, 69 ANTITRUST L.J. 87, 108–09 (2001).

[262] *See* John T. Soma & Kevin B. Davis, *Network Effects in Technology Markets: Applying the Lessons of Intel and Microsoft to Future Clashes Between Antitrust and Intellectual Property*, 8 J. INTELL. PROP. L. 1, 3–4 (2000).

[263] *See Antitrust & Internet Standardization*, *supra* note 259, at 1050–51.

[264] *See* Maureen A. O'Rourke, *Striking A Delicate Balance: Intellectual Property, Antitrust, Contract, and Standardization in the Computer Industry*, 12 HARV. J.L. & TECH. 1, 2–3 (1998).

[265] 17 U.S.C. § 102(a) (2012).

[266] *See* A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1026 (9th Cir. 2001), *as amended* (Apr. 3, 2001*), aff'd sub nom.* 284 F.3d 1091 (9th Cir. 2002) (denying Napster an implied license based on its argument that the record companies had encouraged digital redistribution of their copyrighted works).

[267] *See supra* Section I.B.

[268] *See* Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 347 (1991) ("[F]acts do not owe their origin to an act of authorship. The distinction is one between

Because data scraping is essentially a form of copying using bots, it falls firmly within the subject matter of copyright law. Data scraping and web crawling are fundamentally tools for copying information, facts, and data online.[269] A scraping bot accesses websites and makes copies of those websites, parses the websites' code, and stores information in a database.[270]

In data scraping cases, through the language of "authorization," the CFAA is used to assert a right to exclude, one of the bundle of rights in property.[271] Providers of websites and applications own their computers and typically own or have a leasehold estate on their servers.[272] They also own their intellectual property rights and have a license to user content. But unless the content being copied is original and expressive, these companies do not own the data itself or the underlying information it contains.[273]

However, in data scraping cases, companies use the CFAA to assert something akin to an exclusive right to data. In one data scraping case, the court, discussing the CFAA and citing *Feist*, asked "[w]hy should the copyright symbol, which arguably does not protect the substantive information anyway . . . or the provision of page-by-page access for that matter, be taken to suggest that

---

creation and discovery: The first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence."); Melville Nimmer,1 NIMMER ON COPYRIGHT § 2.11 (2018).

269 *See supra* Section I.A.

270 *See supra* Section I.A.

271 *See* Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 435 (1982) ("The power to exclude has traditionally been considered one of the most treasured strands in an owner's bundle of property rights.").

272 Most major websites and web applications are hosted by Amazon Web Services or another hosting provider, meaning that these services *lease* their server space and capacity. *See generally All Customer Success Stories*, AMAZON, https://aws.amazon.com /solutions/case-studies/all/ [https://perma.cc/7GAA-W3Q8] (last visited Apr. 10, 2018); Klint Finley, *The Amazon S3 Outage Is What Happens When One Site Hosts Too Much of the Internet*, WIRED (Feb. 28, 2017, 4:20 PM), https://www.wired.com/2017/02/happens-one-site-hosts-entire-internet/ [https://perma.cc/B5ES-CU6U]; Mike Williams, *Best Cloud Hosting Providers in 2018*, TECHRADAR (Feb. 1, 2018), https:// www.techradar.com/news/best-cloud-hosting-providers [https://perma.cc/L8ZL-QSES].

273 *Contra* EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 583 (1st Cir. 2001) ("Explorica's wholesale use of EF's travel codes to facilitate gathering EF's prices from its website reeks of use—and, indeed, abuse—of proprietary information that goes beyond any unauthorized use of EF's website").

downloading information at higher speed is forbidden[?]"[274] Plaintiffs in data scraping cases typically allege a "sweat of the brow" argument that is rejected by copyright,[275] often terming scrapers "free riders."[276] Even when CFAA and trespass claims rely on arguments about hypothetical damage to servers, the fundamental disputes are about copying of data.[277]

B. Real Property Metaphors and Trespass Online

In addition to suggesting that there is an exclusive property right in data itself, parties making CFAA and trespass claims also argue that the Internet itself is analogous to a physical place. Both trespass and the CFAA are concerned with the idea of authorization, and litigants opposed to data scraping often suggest that there can be something analogous to an unauthorized entry on a public website.[278]

---

274 *See* EF Cultural Travel BV v. Zefer Corp., 318 F.3d 58, 63 (1st Cir. 2003).

275 *See supra* Sections I.B and I.C.

276 *See, e.g.*, Appellant's Opening Brief at 1, hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783, 2017 WL 4518160 (9th Cir. Sept. 6, 2017), ECF no. 6 ("This case poses the question whether LinkedIn has the right to protect itself from anonymous data-scraping "bots" deployed by hiQ—a company that seeks to free ride on the fruits of LinkedIn's labor and investment by scraping massive volumes of data from LinkedIn's computer servers and then repackaging and selling that data to others.").

277 *See, e.g.*, Fidlar Techs. v. LPS Real Estate Data Sols., Inc., 810 F.3d 1075, 1078, 1080 (7th Cir. 2016) (upholding the district court's granting of summary judgment to defendant LPS on plaintiff Fidlar's CFAA claim, noting that LPS's web harvester was primarily used for copying data, and did not alter the data or disrupt Fidlar's services); *see* Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 404 (2d Cir. 2004) ("While Verio's robots alone would not incapacitate Register's systems, the court found that if Verio were permitted to continue to access Register's computers through such robots, it was 'highly probable' that other Internet service providers would devise similar programs to access Register's data, and that the system would be overtaxed and would crash. We cannot say these findings were unreasonable.").

278 *See, e.g.*, Plaintiff Craigslist, Inc.'s Opposition to Renewed Motion to Dismiss; Response to Brief by Amici Curie, Craigslist, Inc. v. 3Taps, Inc., 964 F.Supp.2d 1178 (N.D. Cal. 2013) (No. CV 12–03816 CRB), 2013 WL 12308283; Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1068 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017) (comparing Power Ventures to a person who wants to borrow a friend's jewelry that is held in a safe deposit box at a bank); Craigslist Inc. v. 3Taps Inc., 964 F. Supp. 2d 1178, 1187 (N.D. Cal. 2013) ("The law of trespass on private property provides a useful, if imperfect, analogy. Store owners open their doors to the public, but occasionally find it necessary to ban disruptive individuals from the premises.").

Since the 1990s, the tort of trespass has been used in ways that rely on fundamental misunderstandings of the subject matter of online property rights.[279] A typical company providing an online service owns its computers as chattel and typically has either a property interest or leasehold estate in its servers.[280] Its interest in its website and code is based on copyright law, and its interest in its brand and domain name is based on trademark law. In most cases, no other property rights exist. The property rights that do exist offer clear claims and remedies. If a person steals a computer or a physical hard drive, the claim to be made is the tort of conversion.[281] If a person copies a work of authorship without permission or fair use, the claim is copyright infringement.[282] If a person harms a computer or denies its possessor of its use, the claim is trespass to chattels.[283]

Trespass to chattels, however, has been routinely applied in cases involving the Internet in ways that imply that cyberspace is a place where real property exists.[284] As a result, judges have applied rules about trespass to land to chattels without the constraints of real property law. In a classic trespass to real property case, "although a visitor may be an invitee when first entering a home, he may be demoted to a licensee or trespasser under certain

---

[279] Many academic articles have been written suggesting that cyber trespass attempts to create new property rights, including those explicitly discussed below, and that these new property rights are unconstrained and do not belong in means of communication. *See e.g.*, Mary Anne Bendotoff & Elizabeth R. Gosse, *"Stay Off My Cyberproperty!": Trespass to Chattels on the Internet*, 6 INTELL. PROP. L. BULL. 12, 17 (2001); Laura Quilter, *The Continuing Expansion of Cyberspace Trespass to Chattels*, 17 BERKELEY TECH. L.J. 421, 437–42 (2002); Eric J. Feigin, *Architecture of Consent: Internet Protocols and Their Legal Implications*, 56 STAN. L. REV. 901, 931–32 (2004).

[280] Today, few companies maintain their own web servers; instead, most online companies lease servers from large cloud hosting providers, such as Amazon Web Services (AWS). *See supra* note 272.

[281] *See* RESTATEMENT (SECOND) OF TORTS § 222A (AM. LAW INST. 1965). Under New York common law, computer files themselves and other intangible property may be subject to the tort of conversion, *e.g.*, Thyroff v. Nationwide Mut. Ins. Co., 8 N.Y.3d 283, 292-93 (2007). This rule does not create new rights but merely creates continuity for old property rights, and is constrained by the "merger doctrine," meaning the property must be theoretically representable in paper form, such as a stock certificate, a promissory note, or a physical client list. *Id.*at 291–92.

[282] 17 U.S.C. § 501(a) (2012).

[283] *See* RESTATEMENT (SECOND) OF TORTS § 217 (AM. LAW INST. 1965).

[284] *See* Hunter, *supra* note 256, at 483–88.

circumstances—such as when an invitation is unequivocally revoked."[285] When the metaphor of a cyberspace as a place is applied to a website, a user who enters with permission is a common law invitee, and when that permission is withdrawn or the authorization to enter is exceeded, the user becomes a trespasser.[286] But this analogy is deeply flawed: while computers and servers are chattels, they are not real property.

If courts applied trespass to chattels to computers following the traditional constraints of tort law, its use would be significantly more limited. A trespass to chattels claim requires intent and use or intermeddling with a chattel in possession of another in a manner which impairs the chattel or deprives the possessor of its use.[287] Because dispossession includes barring a possessor's access to chattel or destroying a chattel while it is in another's possession, installing ransomware or malware can be reasonably viewed as trespass to chattels, because both effectively deny the possessor the use of the chattel.[288] As this example suggests, trespass to chattels requires "substantial" actual harm that is more than theoretical or de minimis, a requirement that has not been applied in cyber-trespass cases.[289] The metaphor in *eBay* of an auction house filled with robots ignores the fact that websites are not physical places, and additionally overlooks how little of eBay's traffic came from bots.[290] Because eBay's server, like other servers of public websites, intentionally communicates with other computers and servers, the use or intermeddling element of trespass to chattels is difficult to apply.[291] How does a bot meet the requirement of use or intermeddling when a server is intended for communication with the public? It is additionally difficult to see how a bot constituting

---

285 *See* Rogers v. Martin, 63 N.E.3d 316, 326 n.4 (Ind. 2016); *see* Estate of Joshua S. Cilley v. Lane, 985 A.2d 481, 486 (Me. 2009) ("A licensee who is asked to leave and refuses becomes a trespasser").

286 Hunter, *supra* note 256, at 482.

287 RESTATEMENT (SECOND) OF TORTS §§ 217(b), 218(b)–(c) (AM. LAW. INST. 1965).

288 RESTATEMENT (SECOND) OF TORTS § 221(c)–(d) (AM. LAW. INST. 1965).

289 Steven Kam, *Intel Corp. v. Hamidi: Trespass to Chattels and A Doctrine of Cyber-Nuisance*, 19 BERKELEY TECH. L.J. 427, 433–35 (2004).

290 *See* eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1067 (N.D. Cal. 2000).

291 RESTATEMENT (SECOND) OF TORTS §§ 217(b), 218(b)–(c) (AM. LAW. INST. 1965).

around 1% of traffic would impair the server or deprive eBay of its use.[292]

However, while trespass claims are still made in data scraping cases,[293] in recent years CFAA claims—which may be understood as computer trespass claims[294]—have prevailed.[295] In *Facebook*, while the court rejected the contract-based approach to the CFAA, it embraced a trespass to real property approach by treating the cease and desist letter as notice that Power Ventures' implied permission, or authorization, to access Facebook's website had been revoked.[296] The same reasoning can be seen in *3Taps*, where the court recognized Craigslist as having a right to exclude 3Taps from its website, one of the essential rights in property.[297]

The use of trespass metaphors in CFAA cases has been widely criticized in legal scholarship. In *Cyberspace as a Place and the Tragedy of the Anticommons*, Dan Hunter suggests that because language shapes perceptions of reality, the cyberspace-as-a-place metaphor leads to the application of spatial assumptions online.[298] In the 1990s, we surfed the web, hung out in chatrooms, used email addresses, and worried about application backdoors, all language reflecting a view of cyberspace as land.[299] This use of metaphor is particularly damaging when adopted by courts in the context of computer and network "trespass."[300]

---

[292] *See eBay*, 100 F. Supp. 2d at 1063.

[293] *See, e.g.*, Couponcabin LLC v. Savings.com, Inc., No. 2:14-CV-39-TLS, 2016 WL 3181826, at *2 (N.D. Ind. June 8, 2016). Facebook, notably, did not allege a claim of trespass in its complaint against Power Ventures. First Amended Complaint at 1, Facebook, Inc. v. Power Ventures, Inc., No. 08-CV-5780-LHK, 2013 WL 5372341 (N.D. Cal. Sept. 25, 2013), 2009 WL 3561632.

[294] Goldfoot & Bamzai, *supra* note 153, at 1482–83 ("the CFAA established that 'trespassing' violated computer owners' rights"); *See e.g.*, QVC, Inc. v. Resultly, LLC, 159 F. Supp. 3d 576, 590 (E.D. Pa. 2016) ("Moreover, '[t]he general purpose of the CFAA was to create a cause of action against computer hackers (e.g., electronic trespassers).'").

[295] *See, e.g.*, Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058 (9th Cir. 2016), *cert. denied*, 138
S. Ct. 313 (2017).

[296] *Id.* at 1067–68.

[297] *See* Craigslist Inc. v. 3Taps Inc., 964 F. Supp. 2d 1178, 1184 (N.D. Cal. 2013).

[298] Hunter, *supra* note 256, at 477–78.

[299] *Id.* at 454–55.

[300] *Id.* at 482.

Others have taken a more optimistic view of trespass metaphors online. In his 2016 article *Norms of Computer Trespass*, Orin Kerr works within the trespass metaphor, noting the importance of social norms to physical trespass and suggesting that courts look to Internet norms to rule in online trespass cases.[301] He argues that courts should apply a presumption of openness to the web and view efforts to regulate access such as "terms of use, hidden addresses, cookies, and IP blocks . . . as merely [sic] speed bumps rather than virtual barriers."[302] His test for trespass is a bright-line test, drawn when a user (or bot) bypasses an authentication requirement.[303] In the petition for certiorari of *Facebook v. Power Ventures*, the Cato Institute, writing as amicus curiae, suggested applying a landlord-tenant metaphor to the facts of the case.[304] Cato noted the prevalence of password sharing, and suggested that the average Facebook user views himself as a tenant, able to invite guests onto the website, without his landlord's, or Facebook's, permission.[305] In *hiQ*, the court analogized LinkedIn's ban on hiQ accessing its website to a store owner banning members of the public from viewing a sign from a public sidewalk.[306]

When using a metaphor to describe the Internet, it is essential to consider the limitations of the analogy when making inferences. Public websites, by their nature, require their servers to communicate with the computers of their visitors. Any person with an Internet connection can make a request from a server and receive a response, making the Internet seemingly like a public place. However, a client, like a web browser or bot, which makes an HTTP get request to a website is more analogous to a customer calling a 1-800 number than to a customer visiting a mall. The nature of the interaction is communication, not a physical entry. As

---

301 *See generally Computer Trespass*, *supra* note 156, at 1143.

302 *Id.* at 1161.

303 *Id*.

304 *See* Brief of the Cato Institute as Amicus Curiae Supporting the Petition for Certiorari, Power Ventures, Inc. v. Facebook, Inc., No. 16–1105, 138 S.Ct. 313 (2017), 2017 WL 2391509, *11–12.

305 *Id*.

306 *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1112–13 (N.D. Cal. 2017).

a result, metaphors that treat the Internet as a place can only be applied to the extent the relevant analogy conforms to the technology under consideration.

In contrast to the balance offered by copyright law, the CFAA and trespass have been applied in data scraping cases to insinuate exclusive rights that are unlimited and nearly absolute. Even traditional real property rights are limited by doctrines like nuisance and easement, but no comparable limitations exist for the CFAA and trespass claims in data scraping cases.

III. SOLUTIONS

The problems found in data scraping cases result, in part, from a lack of claims tailored to activity online. The CFAA was first enacted 1984 prior to widespread use of computers,[307] and its amendments have primarily served the purpose of expanding its use by prosecutors,[308] with little focus on its civil causes of action. While Internet norms exist in a positive sense,[309] they may also be normatively created with the help of legislators and judges. As a result, this Note proposes both legislative and interpretative solutions whenever possible, discussing (1) the CFAA; (2) breach of contract; and (3) copyright.

Specifically, Section III.A argues that the CFAA should be interpreted under a plain meaning analysis to reflect a clearer understanding about the extent to which any ordinary user is "authorized" to access a particular website, and "entitled" to obtain information from that website in the context of computers and servers. Like other statutes in the information space, the CFAA should be amended to contain exceptions; it should also be amended with clearer language. Section III.B argues that while interpretations of the CFAA often rely on terms of use, we should evaluate those terms using contract law, and improve terms of use and websites so that users are aware of what they have agreed to. Section III.C suggests that data scraping cases are often copyright

---

307 *See generally* S. REP. NO. 104-357, at 3 (1996).

308 *See* Tiffany Curtiss, *Computer Fraud and Abuse Act Enforcement: Cruel, Unusual, and Due for Reform*, 91 WASH. L. REV. 1813, 1814 (2016).

309 *See supra* Section I.A.

compilation cases where the copying is enabled by technological means, and that copyright could handle these cases using a *Feist v. Rural* analysis and fair use.

A. The CFAA Should Not Be Used to Penalize Data Scraping

1. Data Scraping Is Not Encompassed by the Contextual Meaning of "Exceeds Authorized Access"

"The starting point for interpreting a statute is the language of the statute itself."[310] In the CFAA, the phrase "exceeds authorized access" is defined to mean "to access a computer with authorization and to use such access to obtain or alter information in the computer that the accessor is not entitled so to obtain or alter[.]"[311] Though courts commonly apply a more common-sense analysis to understanding what it means to exceed authorized access,[312] under the rule against surplus age, the definition must be given meaning or the words lose their effect.[313] In addition, by the same principle, the idea of exceeding authorized access must differ from accessing without authorization. The words of the definition of "exceeds authorized access" themselves offer a two-part test: (1) first, we examine whether the user accessed a computer *with authorization*; (2) second, we evaluate whether the user used this access to obtain or alter information that he or she was *not entitled to so obtain or alter*. However, the meanings of the terms, "authorization" and "entitled," remain unclear.

To determine plain meanings of words, courts often look to dictionary definitions as a starting point, though the results may be indeterminate.[314] The Oxford English Dictionary defines "authorize" as to "[g]ive official permission for or approval to (an

---

310 *See* Consumer Prod. Safety Comm'n v. GTE Sylvania, Inc., 447 U.S. 102, 108 (1980).

311 18 U.S.C. § 1030(e)(6) (2012).

312 *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1067 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017) (holding that the lack of permission from Facebook to access its website, as indicated by the cease and desist letter, was determinative).

313 *See, e.g.*, Duncan v. Walker, 533 U.S. 167, 174 (2001).

314 *See Looking It Up: Dictionaries and Statutory Interpretation*, 107 HARV. L. REV. 1437, 1445–46 (1994).

undertaking or agent)."[315] Merriam Webster defines "authorize" as first, "to endorse, empower, justify, or permit by or as if by some recognized or proper authority (such as custom, evidence, personal right, or regulating power)" and, second, "to invest especially with legal authority[.]"[316] Authorization is then defined as "the act of authorizing[.]"[317] Black's Law Dictionary defines authorization as "[o]fficial permission to do something; sanction or warrant" or as the "official document granting such permission."[318] In the context of the CFAA applied online, few of these definitions appear particularly relevant. Ordinary human users do not have legal authority to visit websites, nor do they have official permission. The Merriam-Webster definition at least indicates that custom may play a role in determining whether a user has authorization, and that authorization could be similar to the concept of permission.

However, the ambiguity of a statutory term does not depend solely on dictionary definitions, nor can the words of such a term be viewed in isolation from one another.[319] To determine whether a statute is ambiguous, courts look to the language of the statute itself, "the specific context in which that language is used, and the broader context of the statute as a whole."[320] "[T]he meaning of a word . . . must be drawn from the context in which it is used."[321] In short, in statutory interpretation, context matters.

Authorization has a specific meaning in the context of computers and servers. A 1996 Internet Engineering Task Force (IETF) publication discusses this meaning, stating that: "Authorization refers to the process of granting privileges to processes and, ultimately, users. This differs from authentication in that authentication is the process used to identify a user. Once

---

315 *Authorize*, OXFORD DICTIONARIES, https://en.oxforddictionaries.com/definition/us/authorize [https://perma.cc/66BD-39Q3] (last visited Mar. 30, 2018).

316 *Authorize*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/authorize? [https://perma.cc/DK5K-TGZK] (last visited Mar. 30, 2018).

317 *Id.*

318 *Authorization*, Black's Law Dictionary (10th ed. 2014).

319 *See* Yates v. United States, 135 S. Ct. 1074, 1081 (2015).

320 *See* Robinson v. Shell Oil Co., 519 U.S. 337, 341 (1997).

321 *See* Deal v. United States, 508 U.S. 129, 132 (1993).

identified (reliably), the privileges, rights, property, and permissible actions of the user are determined by authorization."[322]

On a website, an authorization policy defines what a user can see and do.[323] Thus, on a public website, all users have "authorization" to view public resources, or URLs. Authorization may be very granular, granting specific individuals and groups abilities to read, write, modify, and delete resources.[324] This version of authorization offers a bright line rule: if a user *can* view a resource without hacking, a user is *authorized* to view that resource. In contrast, a pseudo-public website, one which, like Facebook or LinkedIn, uses a login, is said to have an "authentication" requirement.[325] A user is also authorized if she has credentials—typically, a username and password—that grant such authenticated access.

Colloquially, in computer security, when access is described as "unauthorized," it typically means that black-hat hacking has occurred, or that a user does not have "credentials" to access an online resource such as a website. In another white paper about authentication, after noting that one method of authentication can be possession of an item, such as a credit card or proximity badge, the author states, "[p]ossession based authentication is clearly subject to theft or use by an *unauthorized* individual if lost or stolen."[326] In an article describing Cambridge Analytica's access to Facebook, the author states that "[t]here was no unauthorized external hacking involved[.]"[327]

---

322 *Site Security Handbook,* IETF NETWORK WORKING GROUP (B. Fraser ed., Sept. 1997), https://tools.ietf.org/html/rfc2196 [https://perma.cc/8AGA-CFV7].

323 Dave Piscitello, *Access Controls, User Permissions and Privileges*, ICANN BLOG (Jan. 19, 2016), https://www.icann.org/news/blog/access-controls-user-permissions-and-privileges [https://perma.cc/F62Y-2ACP].

324 *Id.*

325 I refer to Facebook and LinkedIn as "pseudo-public" because nearly anyone with an email address may join these websites, but both have an authentication requirement to view specific content.

326 *See* Doug Graham, *It's All About Authentication*, SANS INST. (Mar. 15, 2003), https://www.sans.org/reading-room/whitepapers/authentication/its-about-authentication-1070 [https://perma.cc/2GD2-K8S2].

327 *See* Ido Kilovaty, *The Cambridge Analytica Debacle Is Not a Facebook "Data Breach." Maybe It Should Be.*, TECHCRUNCH (Mar. 17, 2018),

Even if we accept a looser definition of authorization, the dictionary definitions of "authorization" and "authorize" suggest that the concept of authorization may be based in norms. Password sharing is such a norm. Many households have a shared Netflix or Hulu account,[328] and spouses often have access to each other's online bank accounts for the purpose of paying bills.[329] Companies often have an official Twitter account, and until 2015 a shared username and password was required for multiple employees to have access.[330] In addition to the commonality of password sharing, the express prohibition of "trafficking" in passwords with "intent to defraud" elsewhere in the CFAA suggests that mere sharing of passwords without the requisite *mens rea* is outside the CFAA's scope.[331] As a result, users of websites with an authentication barrier, such as a login requirement, should be considered to have authorization if they access the website with the permission of the account holder.

Courts should adopt this interpretation of "authorization" by creating a judicial presumption of authorization in CFAA cases involving public websites or valid login information. The

---

https://techcrunch.com/2018/03/17/the-cambridge-analytica-debacle-is-not-a-facebook-data-breach-maybe-it-should-be/ [https://perma.cc/U2SJ-TYTG].

328 Reuters, *People Sharing Passwords are a Growing Problem for Netflix*, FORTUNE (Jul. 11, 2017), http://fortune.com/2017/07/11/netflix-hulu-password-sharing/ [https://perma.cc/P5WR-2XM4]; *see* David Nield, *How to Safely Share Your HBO, Netflix, and Other Streaming Logins With Friends*, GIZMODO (Aug. 4, 2017, 11:18 AM), https://fieldguide.gizmodo.com/how-to-safely-share-your-hbo-netflix-and-other-stream-1797530211 [https://perma.cc/A6VM-Z7PA].

329 *See* Ruchika Tulshyan, *Is Your Spouse Your Biggest Online Security Risk?*, FORBES (Aug. 23, 2013, 11:32 AM), https://www.forbes.com/sites/ruchikatulshyan/2013/08/23/is-your-spouse-your-biggest-online-security-risk/#57d292436de6 [https://perma.cc/3VT3-DSV8]; Eliana Dockterman, *Your Password or Your Privacy: Why Partners Share—And Why They Shouldn't*, TIME (Feb. 24, 2014), http://healthland.time.com/2014/02/24/the-complicated-politics-of-sharing-passwords-with-a-partner/ [https://perma.cc/Y49H-H4SD].

330 *See* Greg Kumparak, *Twitter Finally Lets You Share Team Accounts Without Sharing Passwords*, TECHCRUNCH (Feb. 17, 2015), https://techcrunch.com/2015/02/17/share-twitter-account/ [https://perma.cc/EUR8-8SR4].

331 18 U.S.C. § 1030(a)(6) (2012) ("Whoever . . . (6) knowingly and with intent to defraud traffics (as defined in section 1029) in any password or similar information through which a computer may be accessed without authorization, if— (A) such trafficking affects interstate or foreign commerce; or (B) such computer is used by or for the Government of the United States.").

presumption of authorization to access a public website can then only be overcome by a showing that a user did not have permission—implied or express—to use login credentials or that a user "hacked" the website. This presumption of authorization would apply no matter whether a defendant is accused of accessing a protected computer "without authorization" or in a manner that "exceeds authorized access."

As a result, the second part of the test for interpreting the phrase "exceeds authorized access," which looks to the meaning of "used this access to obtain or alter information he or she was not entitled to so obtain or alter" needs to refer to something that differs from the concept of authorization based on the canon of meaningful variation.[332] The word entitled is thus a key word of the statutory definition, because the phrase rests on whether an individual is "not entitled." Merriam-Webster defines entitle as "to give a title to" or "to furnish with proper grounds for seeking or claiming something[.]"[333] Black's Law Dictionary defines "entitle" as "[t]o grant a legal right to or qualify for."[334]

Users do not precisely have a "title," property right, or legal right to access websites. However, users do have something equivalent to a "lawful entry" onto a public website. If an ordinary user were not "entitled" to "obtain" public files on a public website, then the CFAA would be so overbroad as to be meaningless. In addition, because the CFAA is also a criminal statute, an overbroad violation violates the rule of lenity.[335] Because the word "entitled" is not commonly used in the context of computers and servers, it is necessary to determine the specific meaning of the word in context of the CFAA.

---

[332] The canons of presumption of consistent usage and meaningful variation require interpretation of the same or similar terms in a statute in the same way. *See* Jacob Scott, *Codified Canons and the Common Law of Interpretation*, 98 GEO. L.J. 341, 368–69 (2010).

[333] *Entitled*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/entitled? [https://perma.cc/9ZF8-JG38] (last visited Mar. 18, 2018).

[334] *Entitle*, Black's Law Dictionary (10th ed. 2014).

[335] United States v. Santos, 553 U.S. 507, 514 (2008) ("The rule of lenity requires ambiguous criminal laws to be interpreted in favor of the defendants subjected to them."); LVRC Holdings LLC v. Brekka, 581 F.3d 1127, 1134–35 (9th Cir. 2009) (applying the rule of lenity to the CFAA).

One word that has a similar meaning to "entitled" is the word "privileged." Both words are concerned with the idea of rights, and like an entitlement, a privilege is something that may be granted.[336] In the context of computers and servers, measures used "to implement authorization policies are called user access controls, user permissions[,] or user privileges."[337] In websites and applications with a login requirement, users have privileges to specific files, or resources, with authentication protocols that restrict access.[338] This concept of access controls, permissions, and privileges could be used to give specific meaning to the word entitled. A user would then violate the second prong of the proposed test by using authorized access to obtain or alter information he or she did not have permissions or privileges to so obtain or alter, based on the technical meanings of the terms "permissions" and "privileges."

Under this proposed test, a user would then violate the CFAA's "exceeds authorized access" provision by breaching an authentication barrier, or by accessing resources she lacked privileges or permissions to access. Because this interpretation looks to technological access barriers, circumvention measures that do not define privileges or permissions or effectively restrict access, such as an IP address block or CAPTCHA, would not be considered a CFAA violation. This interpretation also rejects the idea of applying spatial norms to the Internet. Instead, plaintiffs alleging CFAA violations would need to describe the authorization, privileges, and permissions granted to users of their websites and applications generally, and then contrast the behavior of the defendant.

This proposed test also properly limits the scope of the CFAA to hacking. A bot that accesses a website with a user's permission is simply another web client which, like a web browser, is copying

---

336 *Privilege*, Black's Law Dictionary (10th ed. 2014) (defining privilege as a "special legal right, exemption, or immunity granted to a person or class of persons; an exception to a duty.").

337 Piscitello, *supra* note 323.

338 A. Arthur Fisher, *Authentication and Authorization: The Big Picture with IEEE 802.1X*, SANS INSTITUTE (Dec. 21, 2001), https://www.sans.org/reading-room/whitepapers/authentication/authentication-authorization-big-picture-ieee-8021x-123 [https://perma.cc/UG9R-TRB6].

a website on behalf of the user. Because web crawlers and data scraping bots access resources with authorization, privileges, and permissions, their activity would not be covered by a narrow interpretation of the CFAA. Moreover, scraping is not hacking. To the extent hacking "get[s] inside a computer," or trespasses, then hacking refers to, for example, "gaining [unauthorized] access to the stored contents of a computer system, gaining access to the processing capabilities of a system, or intercepting information being communicated between systems."[339] When a bot or crawler interacts with a website in a way that is hard to distinguish from a human user, courts should find that no CFAA violation has occurred.

The plain-meaning interpretation of the CFAA proposed here is very much like a code-based approach, but the focus is on using specialized understandings of terms to give clear meaning to the words of the CFAA.[340] However, the approach proposed here allows courts to consider evidence about how ordinary users access applications and websites with authorization, and then contrast the behavior of an accused CFAA violator.

2. Data Scraping Rarely Results In A "Loss"

To be eligible for a civil remedy, a violation of the CFAA must have resulted in, at a minimum, a "loss" to one or more persons of at least $5,000, occurring during any one-year period.[341] "Loss" is defined as, "any reasonable cost to any victim, including the cost of responding to an offense, conducting a damage assessment, and restoring the data, program, system, or information to its condition prior to the offense, and any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service[.]"[342]

---

339 Julie J.C.H. Ryan, *How Do Computer Hackers "Get Inside" a Computer?*, SCI. AM. (Aug. 16, 2004), https://www.scientificamerican.com/article/how-do-computer-hackers-g/?print=true [https://perma.cc/7P5V-LTYR].

340 *See* Rosen, *supra* note 155, at 760 ("Under the proposed code-based approach, an employee exceeds authorized access when she (1) encounters a code-based barrier on her employer's computer and then (2) proceeds to use her authorized access to obtain or alter information that exists behind the barrier.").

341 18 U.S.C. §§ 1030(c)(4)(A)(i)(I) & (g) (2012).

342 18 U.S.C. § 1030(e)(11) (2012).

In data scraping cases, parties often allege that the "loss" occurred from responding to the scraping and determining the identity of the scraper.[343] The canon of noscitur a sociis[344] allows us to look at the entire definition to understand the meanings of "cost of responding to an offense" and "conducting a damage assessment" from these phrases' associates.[345] The rest of the definition discusses restoring data and interruption of service, implying that "damage" must be more than a little extra traffic on a website, and "responding" may require more than setting up an IP address block. In data scraping cases, there is rarely an interruption of service,[346] as creators of data scraping bots often take measures to ensure that they are "polite" and behave like ordinary users.[347]

In recent years, courts have sometimes recognized that "actual disruptions in service, not mere access" is required for CFAA "damage."[348] Similar reasoning could be applied to the concept of "loss." Courts should carefully scrutinize the basis of any CFAA "loss" in data scraping cases to determine what, if any, harm actually occurred. If harm to computers or servers occurred, then

---

343 *E.g.*, Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1066 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313, 199 L. Ed. 2d 206 (2017) ("It is undisputed that Facebook employees spent many hours, totaling more than $5,000 in costs, analyzing, investigating, and responding to Power's actions"); EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 585 (1st Cir. 2001) ("Appellees unquestionably suffered a detriment and a disadvantage by having to expend substantial sums to assess the extent, if any, of the physical damage to their website caused by appellants' intrusion. That the physical components were not damaged is fortunate, but it does not lessen the loss represented by consultant fees.").

344 *Noscitur a sociis* means "it is known from its associates," and is a canon of statutory interpretation that looks to the meaning of a statutory term based on the words and phrases surrounding it. *See* Yates v. United States, 135 S. Ct. 1074 (2015).

345 *See* Graham Cty. Soil & Water Conservation Dist. v. U.S. ex rel. Wilson, 559 U.S. 280, 287 (2010).

346 *E.g.*, *Power Ventures*, 844 F.3d at 1066; CollegeSource, Inc. v. AcademyOne, Inc., No. CIV.A. 10-3542, 2012 WL 5269213, at *15 (E.D. Pa. Oct. 25, 2012), *aff'd*, 597 F. App'x 116 (3d Cir. 2015) (holding that while CollegeSource did not assert damage or interruption of a computer, it could claim a loss based on "internal investigation of AcademyOne's websites, its hiring of a computer expert, and its subsequent security measures."); *EF Cultural Travel*, 274 F.3d at 585.

347 *See supra* Part I; *cf.* Snap-on Bus. Sols. Inc. v. O'Neil & Assocs., Inc., 708 F. Supp. 2d 669, 675–76 (N.D. Ohio 2010) (noting that O'Neil's software stopped crashing Snap-On's website once he limited the rate of requests).

348 Fidlar Techs. v. LPS Real Estate Data Sols., Inc., 810 F.3d 1075, 1085 (7th Cir. 2016).

the time and resources spent investigating and responding to that harm may be properly encompassed by a CFAA loss.

3. Ultimately, the CFAA Should Be Amended To Clarify its Meaning And Add Exceptions And Preemption Provisions

Even with these limitations, the words of the CFAA are generally thought to be overbroad,[349] and this notion is supported when the CFAA is compared with similar statutes. The DMCA, for instance, states that, "[n]o person shall circumvent a technological measure that effectively controls access to a work protected under this title."[350] The DMCA is a much broader statute than the CFAA, but is specifically tailored to protect its underlying property right, that of copyright. However, the DMCA also contains numerous exceptions, including for reverse engineering,[351] and encryption research,[352] and protection of personal identifiable information ("PII").[353] The Stored Communication Act ("SCA") is also concerned with the concept of "authorization" and includes a provision for whoever "intentionally exceeds an authorization . . . and thereby obtains, alters, or prevents authorized access" to electronic communications.[354] Like the DMCA, the SCA also contains numerous exceptions, including for providers of electronic communications services[355] and for required disclosures based on court orders.[356] Even the Espionage Act, a very broadly

---

[349] *See, e.g.*, Sandvig v. Sessions, 315 F.Supp.3d 1, 25 (D.D.C. 2018) ("By providing for both civil and criminal enforcement of websites' limitless ToS—including enforcement by the same entities that write the ToS—a broader reading of the CFAA 'would appear to criminalize a broad range of day-to-day activity' and 'subject individuals to the risk of arbitrary or discriminatory prosecution and conviction,' raising Fifth Amendment concerns."); *see* Orin S. Kerr, *Vagueness Challenges to the Computer Fraud and Abuse Act*, 94 MINN. L. REV. 1561, 1561–62 (2010) ("The CFAA has become so broad, and computers so common, that expansive or uncertain interpretations of unauthorized access will render it unconstitutional."); *see also* Wu, *supra* note 175.

[350] 17 U.S.C. § 1201(a) (2012).

[351] 17 U.S.C. § 1201(f) (2012).

[352] 17 U.S.C. § 1201(g) (2012).

[353] 17 U.S.C. § 1201(i) (2012). PII is any data that can be used to identify an individual, such as name, social security number, address, phone number, etc. *See Guidance on the Protection of Personal Identifiable Information*, U.S. DEPARTMENT OF LABOR, https://dol.gov/general/ppii [https://perma.cc/K2CL-WW8B] (last visited Sept. 30, 2018).

[354] 18 U.S.C. § 2701(a) (2012).

[355] 18 U.S.C. § 2701(c)(1) (2012).

[356] 18 U.S.C. § 2703 (2012).

written law—concerned with "unauthorized possession" and whether a person is "entitled to receive" materials related to national security[357]—has been limited over time by courts. Though courts have not ruled the Espionage Act to be unconstitutionally vague, modern courts often limit the statute's terms through inference.[358] In addition, the Espionage Act has been limited in the past through First Amendment jurisprudence.[359]

The CFAA is not written with any such exceptions, and courts have been reluctant to limit the CFAA's scope thus far. Several amendments have been proposed to narrow the scope of the CFAA,[360] but none has ever made it out of committee.[361] At a minimum, the CFAA should be amended to contain similar exceptions to the DMCA.[362] Specifically, the CFAA should contain explicit exceptions for copying of data, reverse engineering, and security research. The CFAA's main provisions

---

357 18 U.S.C. § 793(e) (2012).

358 *See* Laura Barandes, *A Helping Hand: Addressing New Implications of the Espionage Act on Freedom of the Press*, 29 CARDOZO L. REV. 371, 374 (2007); United States v. Morison, 844 F.2d 1057, 1074 (4th Cir. 1988) ("The defendant would also indict the phrase 'entitled to receive' as vague. The defendant finds this phrase vague because it does not spell out exactly who may 'receive' such material. However, any omission in the statute is clarified and supplied by the government's classification system provided under 18 U.S.C. App. 1 for the protection of the national security and the district judge so ruled.").

359 *See* New York Times Co. v. United States, 403 U.S. 713, 714–15 (1971).

360 *See, e.g.*, Active Cyber Defense Certainty Act, H.R. 4036, 115th Cong. (2017) (adding provisions permitting "active cyber defense measures"); *see* Personal Data Privacy and Security Act of 2011, S. 1151, 112th Cong. (2011) (adding additional penalties related to fraud offenses, defining some computers as "critical infrastructure computers," and limiting section 1030(g)'s applicability to terms of use violations); *see also* Aaron's Law Act of 2013, H.R. 2454, 113th Cong. (2013).

361 18 U.S.C. § 1030 was most recently amended in 2008 to add provisions intended to combat cyber-extortion. *See* Identity Theft Enforcement and Restitution Act of 2008, Pub. L. 110-326, Title II, §§ 203, 204(a), 205—208, 122 Stat. 3561, 3563 (2008). In general, most proposed bills to amend section 1030 made since 2000 seek to broaden the CFAA or create additional penalties. *See, e.g.*, Botnet Protection Act of 2016, S.2931, 114th Congress (2016). Versions of Aaron's Law were proposed in 2013 and 2015 in both the House and Senate and were referred to committee, but no hearings or markup sessions were held. *See* Aaron's Law Act of 2013, S.1196, 113th Congress (2013); Aaron's Law Act of 2013, H.R.2454, 113th Congress (2013); Aaron's Law Act of 2015, H.R.1918, 114th Congress (2015); Aaron's Law Act of 2015, S.1030, 114th Congress (2015).

362 17 U.S.C. § 1201(f)–(i) (2012).

should also be clarified. Narrow and specific definitions of terms such as "authorization" and "access" would help courts limit the CFAA's scope.

The Electronic Frontier Foundation's ("EFF") proposal,[363] a clarification of Aaron's Law, is an appropriate amendment of the CFAA that would limit its use in data scraping cases. Aaron's Law proposed, along with modifications to the CFAA's criminal penalties, striking the phrase "exceeds authorized access" from the CFAA and replacing "without authorization" to "access without authorization."[364] The EFF's modification of Aaron's Law suggests defining "access without authorization" as

> to circumvent technological access barriers to a computer, file, or data without the express or implied permission of the owner or operator of the computer to access the computer, file, or data, but does not include circumventing a technological measure that does not effectively control access to a computer, file, or data.[365]

The term "without the express or implied permission" is specifically noted to "not include access in violation of a duty, agreement, or contractual obligation, such as an acceptable use policy or terms of service agreement, with an Internet service provider, Internet website, or employer."[366]

However, the EFF's proposal uses the similar language to the DMCA,[367] making circumvention of a technological access barrier

---

[363] *See CFAA Revisions – Penalties and Access*, EFF, https://www.eff.org/document/eff-cfaa-revisions-penalties-and-access [http://perma.cc/6EGX-VG4N] (last visited Apr. 1, 2018).

[364] Aaron's Law Act of 2013, H.R. 2454, 113th Cong. (2013); Aaron's Law Act of 2013, S.1196, 113th Cong. (2013). Aaron's Law, as proposed in 2013, defines access without authorization as "(A) to obtain information on a protected computer; (B) that the accessor lacks authorization to obtain; and (C) by knowingly circumventing one or more technological or physical measures that are designed to exclude or prevent unauthorized individuals from obtaining that information."

[365] *CFAA Revisions*, *supra* note 363.

[366] *Id*.

[367] The DMCA states, "[n]o person shall circumvent a technological measure that effectively controls access to a work protected under this title." *See* 17 U.S.C. § 1201(a) (2012). EFF's proposal defines "access without authorization" as "to circumvent technological access barriers to a computer, file, or data without the express or implied

the basis of a CFAA violation.[368] This language is clarified by the statement that circumvention of a technological measure that "does not effectively control access to a computer, file, or data" is not a CFAA violation.[369] The EFF's proposal would effectively make the agency and contract approaches[370] to the CFAA obsolete, preventing the CFAA from being used to enforce unenforceable contracts of adhesion. However, while the phrase "does not effectively control access" should require courts to obtain expert witness testimony from software engineers and digital security professionals, it is possible that a court could hold that an IP address block or CAPTCHA effectively controls access to a file or data. One possibility to prevent this would be to list examples of measures that effectively control access in the statute, such as an authentication barrier, and measures that do not, such as an IP address block.

In addition, if the CFAA is reformed, its amendments could potentially include a preemption clause that is similar to Section 301 of the Copyright Act.[371] This would help reduce the use of trespass as a cause of action in computer misuse cases, as well as data scraping cases, and create predictability and stability for online service providers.

B. Online Contracting Can Be Improved

In data scraping cases, judges often dismiss a clickwrap or browserwrap contract as unenforceable, but turn around and decide that the same unenforceable terms of service make scraping a CFAA violation.[372] It is even more common for websites to prohibit bots, spiders, and scrapers from their websites in their terms of service,[373] and users not to have read those terms.[374] This

---

permission of the owner or operator of the computer to access the computer, file, or data, but does not include circumventing a technological measure that does not effectively control access to a computer, file, or data[,]" *see CFAA Revisions*, *supra* note 363.

368 *See CFAA Revisions*, *supra* note 363; 17 U.S.C. § 1201(a) (2012).

369 *CFAA Revisions*, *supra* note 363.

370 *See supra* Section I.C.2.

371 *See supra* Section I.C.4.

372 *See supra* Section I.C.

373 *See supra* Section I.A.

374 *See supra* Section I.B.

presents a dilemma: contract law requires that users have actual or constructive notice of the terms of an agreement,[375] but websites owners are loath to clearly present readable terms to their users in a way that encourages users to read them.[376]

Courts have sometimes suggested that, were websites to make their terms more accessible, users would more effectively be bound by terms of service.[377] Online contracts that give users actual or constructive notice of terms are, in fact, achievable, and possibly through the same technology that enables data scraping, bots. Reddit is an example of a platform that effectively gives its users constructive notice of terms using moderators and moderator bots.[378] Reddit's AutoModerator, a bot, allows human moderators to enforce the rules of Reddit and its subreddits, by programming the bot to remove inappropriate links as well comments containing certain words and phrases and to leave comments on threads noting a subreddit's rules.[379] In addition to or in lieu of offering a CAPTCHA when a service provider notices bot-like activity, websites could offer a short-form agreement that discusses prohibited activities. This could apply to other types of prohibited activity as well. Facebook, for example, which prohibits hate speech and threats in its terms of use, could create a bot that looks

---

375 *See supra* Section I.C.

376 *See* Alex Hern, *I Read All the Small Print on the Internet and It Made Me Want to Die*, GUARDIAN (Jun. 15, 2015 6:56 AM), https://www.theguardian.com/technology/2015/jun/15/i-read-all-the-small-print-on-the-internet [https://perma.cc/LK38-XLRC] ("Perhaps the best marker of how little Apple cares about the terms of service it requires its users to read can be found several paragraphs down the iCloud terms and conditions.").

377 *See* Nguyen v. Barnes & Noble Inc., 763 F.3d 1171, 1179 (9th Cir. 2014) (citations omitted) ("While failure to read a contract before agreeing to its terms does not relieve a party of its obligations under the contract, the onus must be on website owners to put users on notice of the terms to which they wish to bind consumers. Given the breadth of the range of technological savvy of online purchasers, consumers cannot be expected to ferret out hyperlinks to terms and conditions to which they have no reason to suspect they will be bound.").

378 *See generally* REDDIT, https://www.reddit.com/ [https://perma.cc/NV84-CR36] (last visited Apr. 2, 2018).

379 *See Moderator,* REDDIT, https://www.reddit.com/wiki/automoderator [https://perma.cc/SBW7-E97L] (last visited Apr. 2, 2018).

for those kinds of speech and leaves comments on posts that may violate its policy, with a link to its terms of use.[380]

In addition, terms of use, which are often unwieldy for laypersons and written with legal jargon,[381] could be rewritten to be shorter and contain links explaining specific policies in detail. One service that already approaches an appropriate level of readability in its terms of use is Etsy.[382] Etsy refers to its terms of use as "house rules," and each paragraph of the terms of use contains a short phrase up front, in bold, which summarizes what the paragraph is about.[383] One paragraph on Etsy's website states, "*Don't Try to Harm Our Systems.* You agree not to interfere with or try to disrupt our Services, for example by distributing a virus or other harmful computer code."[384] Similar to the use of brand awareness surveys to prove acquired distinctiveness in trademark law,[385] surveys could help establish actual or constructive notice in online contracting. Such surveys would determine how well ordinary users can locate, read, and understand a website's terms of use, and could either be used as evidence during litigation, or to help companies better construct their terms of use.

---

[380] *See Terms of Service,* FACEBOOK, https://www.facebook.com/terms.php [https://perma.cc/AR8D-FRL3] (last visited Apr. 2, 2018).

[381] *See, e.g.*, *User Agreement*, LINKEDIN, *supra* note 78 ("TO THE EXTENT PERMITTED UNDER LAW (AND UNLESS LINKEDIN HAS ENTERED INTO A SEPARATE WRITTEN AGREEMENT THAT OVERRIDES THIS CONTRACT), LINKEDIN AND ITS AFFILIATES (AND THOSE THAT LINKEDIN WORKS WITH TO PROVIDE THE SERVICES) SHALL NOT BE LIABLE TO YOU OR OTHERS FOR ANY INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR PUNITIVE DAMAGES, OR ANY LOSS OF DATA, OPPORTUNITIES, REPUTATION, PROFITS OR REVENUES, RELATED TO THE SERVICES (E.G. OFFENSIVE OR DEFAMATORY STATEMENTS, DOWN TIME OR LOSS, USE OF, OR CHANGES TO, YOUR INFORMATION OR CONTENT").

[382] *See generally* ETSY, https://www.etsy.com/ [https://perma.cc/MNP9-YP47] (last visited Apr. 2, 2018).

[383] *See Terms of Use,* ETSY, https://www.etsy.com/legal/terms-of-use/ [https://perma.cc/6TVM-QQG6] (last visited Apr. 2, 2018).

[384] *Id.* (emphasis in original).

[385] *See, e.g.*, Nola Spice Designs, L.L.C. v. Haydel Enterprises, Inc., 783 F.3d 527, 546 (5th Cir. 2015) ("While survey evidence is not required to establish secondary meaning, it is 'the most direct and persuasive way of establishing secondary meaning.'").

C. *Copyright Offers the Correct Balance of Incentives and Remedies and Should Preempt Equivalent State Law Claims*

1. Many Online Data Scraping Cases Are Simply Post-Feist Cases Where Copying Is Enabled by Technological Means

Copyright law protects fixed works of authorship based on originality and expression, and but does not protect factual information or data. In data scraping cases, the data involved is typically data that would be analyzed as a compilation under *Feist*.[386] In the cases decided after *Feist*, books estimating the fair market values of rare coins[387] and projecting the values of used cars[388] were found to be copyrightable, while blank forms, part numbers,[389] settlement prices of futures contracts,[390] a collection of recipes,[391] and charts of winning numbers in illegal gambling operations[392] were all denied copyright protection. To be copyrightable, a compilation of facts must exhibit subjectivity in which facts are included or how they are arranged.[393] Online data providers may exhibit this kind of subjectivity: for example, Zillow offers a feature called the "Zestimate," Zillow's "estimated market value for an individual home[.]"[394] Like books estimating fair market values of coins or used cars, this data is likely copyrightable, because a subjective judgment call has been made in order to create the estimate.

---

386 *See generally*, Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

387 CDN Inc. v. Kapes, 197 F.3d 1256, 1260 (9th Cir. 1999).

388 CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc., 44 F.3d 61, 67 (2d Cir. 1994).

389 ATC Distribution Grp., Inc. v. Whatever It Takes Transmissions & Parts, Inc., 402 F.3d 700, 705 (6th Cir. 2005).

390 *See* New York Mercantile Exch., Inc. v. IntercontinentalExchange, Inc., 497 F.3d 109, 114–15 (2d Cir. 2007).

391 *See* Tomaydo-Tomahhdo, LLC v. Vozary, 629 F. App'x 658, 661–62 (6th Cir. 2015).

392 *See* Victor Lalli Enterprises, Inc. v. Big Red Apple, Inc., 936 F.2d 671, 673 (2d Cir. 1991).

393 *See* Miriam Bitton, *Protection for Informational Works After Feist Publications, Inc. v. Rural Telephone Service Co.*, 21 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 611, 631 (2011).

394 *See Zestimate*, https://www.zillow.com/zestimate/ [https://perma.cc/T6B6-7UUH] (last visited Apr. 8, 2018).

Post-*Feist* cases often involve collections of information that are similar to the information and data scraped in contemporary CFAA cases. For example, *Middle America Title Co. v. Kirk*[395] and *Fidlar Technologies v. LPS Real Estate Data Solutions, Inc.*[396] are both cases involving real estate title information. *Middle America* was decided using a post-*Feist* copyright analysis, while *Fidlar* was a CFAA case.[397]

Middle America sought copyright protection for a compilation of land title data.[398] The court stated that Middle America's failed to show that its selection of facts "involved some kind of creative spark," noting that the alleged work "simply contains a list of all the facts" with "no creativity . . . shown in the selection."[399] As a result, the court denied Middle America copyright protection.[400]

By contrast, in *Fidlar*, LPS copied Fidlar's real estate title data using a "web harvester," but Fidlar sued for violations of the CFAA and trespass to chattels.[401] The Seventh Circuit held that the downloading of data was not a CFAA violation due to the lack of damage, specifically the absence of disruptions in service, but noted that Fidlar's claim and LPS's intrusion was "trespassory in nature."[402] In effect, the court's ruling suggests that it viewed LPS as having accessed Fidlar's software without authorization, but that the lack of resulting damage to the server prevented the CFAA's application.[403]

Many other data scraping cases involve similar allegations of copying what is essentially factual data, such as auction price data,[404] domain registration data,[405] ticket and event information,[406]

---

395 59 F.3d 719 (7th Cir 1995).
396 810 F.3d 1075 (7th Cir. 2016).
397 *Mid Am.*, 59 F.3d at 721; *Fidlar*, 810 F.3d at 1076.
398 *Mid Am.*, 59 F.3d at 721.
399 *Id.* at 723.
400 *Id.*
401 *Fidlar*, 810 F.3d at 1075–77.
402 *Id.* at 1084–85.
403 *See id.*
404 *See* eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1061–62, 1073 (N.D. Cal. 2000).
405 *See* Register.com, Inc. v. Verio, Inc., 126 F. Supp. 2d 238 (S.D.N.Y. 2000), *aff'd as modified*, 356 F.3d 393 (2d Cir. 2004).

and arrest records.[407] Instead of being litigated under the CFAA, these cases should be decided under a *Feist* analysis, resulting in outcomes which would reflect copyright's careful balance of incentives.[408]

2. In Cases Where the Content Being Scraped Is Expressive, We Should Allow User-Based Services to Sue On Behalf of Their Users

Some compilations involved in online data scraping cases, such as LinkedIn, Craigslist, and Facebook, involve what are arguably creative works authored by users. These services' cases have not thus far been decided on copyright issues because these services have non-exclusive licenses to user content.[409] It is well-established law that "[a] non-exclusive license conveys no ownership interest, and the holder of a nonexclusive license may not sue others for infringement."[410] An exclusive license requires a writing, and in *Craigslist Inc. v. 3Taps Inc.*, the court held that Craigslist's terms of use did not constitute the required writing.[411]

---

406 Ticketmaster Corp. v. Tickets.com, Inc., No. 99CV7654, 2000 WL 1887522 (C.D. Cal. Aug. 10, 2000), *aff'd*, 2 F. App'x 741 (9th Cir. 2001).

407 Citizens Info. Assocs., LLC v. Justmugshots.com, No. 1-12-CV-573-LY, 2012 WL 12874898 (W.D. Tex. Dec. 18, 2012).

408 *See supra* Section I.B.

409 *See Linkedin User Agreement*, *supra* note 78 ("... you own the content and information that you submit or post to the Services and you are only granting LinkedIn and our affiliates the following non-exclusive license: A worldwide, transferable and sublicensable right to use, copy, modify, distribute, publish, and process, information and content that you provide through our Services, without any further consent, notice and/or compensation to you or others."); *Facebook Terms of Service*, https://www.facebook.com/terms.php [https://perma.cc/W74Z-RL7A] (last visited Apr. 8, 2018) ("you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook"); *Craigslist Term of Use*, https://www.craigslist.org/about/terms.of.use [https://perma.cc/7A6E-5T7H] (last visited Apr. 8, 2018) ("You grant us a perpetual, irrevocable, unlimited, worldwide, fully paid/sublicensable license to use, copy, display, distribute, and make derivative works from content you post").

410 *See* Melville Nimmer, 1 NIMMER ON COPYRIGHT § 3.05 (2018) ("if the copyright owner ... is merely a nonexclusive licensee ... then if an infringer copies ... the copyright owner ... will not have standing to sue for the infringement"); Davis v. Blige, 505 F.3d 90, 101 (2d Cir. 2007).

411 Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 973–74 (N.D. Cal. 2013). Facebook's copyright claim was also voluntarily dismissed in its case via Fed. R. Civ. P 41(A)(1), though the exact reasons why are unclear. *See* Facebook, Inc. v. Power

The lack of standing in copyright infringement lawsuits for user-based services explains why services like Facebook and LinkedIn have resorted to the CFAA as a potential remedy for copying of their websites.[412] While Power Ventures, for example, had either an express or implied license to its users content, it did not have such a license to their friends' content and arguably infringed some Facebook users' copyrights.[413] Facebook likely did not have standing to sue for copyright infringement, even though its arrangement and selection of user posts in a user's newsfeed is arguably expressive. hiQ's service is potentially harmful to LinkedIn users' privacy, but LinkedIn also lacks standing to sue based on copyright infringement. It seems likely that social media users would balk at giving services the exclusive licenses that would permit user-based services to sue for copyright infringement.

One solution to this problem would be to allow user-based services to sue on behalf of their users in derivative form,[414] or for social media companies faced with data scraping to hire attorneys to file class actions on behalf of their users. Obtaining class certification in copyright cases is often very difficult,[415] but users of a single social media service whose content has been copied may be more likely to meet the standards of class action certification under the Federal Rules of Civil Procedure.[416]

Ventures, Inc., No. 08-CV-05780-LHK, 2017 WL 3394754, at *2 (N.D. Cal. Aug. 8, 2017) ("On February 18, 2011, Judge Ware granted the parties' stipulation to dismiss Facebook's DMCA claim, copyright and trademark infringement claims, and claims for violations of California Business and Professions Code Section 17200.").

412 *See supra* Introduction.

413 *See, generally*, Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017).

414 This could be potentially modeled after shareholder derivative lawsuits.

415 *See, e.g.*, Authors Guild, Inc. v. Google Inc., 721 F.3d 132, 134–35 (2d Cir. 2013) (vacating the class certification of people holding a copyright interest in books copied by Google until fair use was considered); *see also* Football Ass'n Premier League Ltd. v. YouTube, Inc., 297 F.R.D. 64, 65–66 (S.D.N.Y. 2013) ("Generally speaking, copyright claims are poor candidates for class-action treatment").

416 *See generally* Fed. R. Civ. P. 23.

3. Fair Use Should Be Applied Broadly to Intermediary Copying Online

Because copying is so essential to the functioning of the Internet,[417] and many uses of data scraping are arguably transformative,[418] copyright, with its fair use analysis, could result in more equitable outcomes in data scraping cases. The Second Circuit's decisions in *Authors Guild* and *TVEyes* recognize that intermediary copying is often transformative.[419] As a result, the quantity of redistributed material and the effects on the potential market for the copyrighted work are the key elements of an online fair use analysis. In *Authors Guild*, the court held that despite its wholesale digital copying of the plaintiff's books, Google's copying was performed with a "highly transformative purpose" despite its commercial motivation.[420] Because Google made the plaintiffs' works more accessible, and constructed its snippet feature "in a manner that substantially protects against its serving as an effectively competing substitute for Plaintiffs' books[,]" the court found that Google's service was non-infringing fair use.[421]

TVEyes' service, like Google Books, involved recording "essentially all television broadcasts as they happen" and using closed captioning to create a text-searchable transcript, thus allowing its clients "to efficiently sort through vast quantities of television content in order to find clips that discuss items of interest to them."[422] This search feature, like Google Books, was held to be transformative.[423] However, TVEyes also offered a "Watch" feature, which allowed its customers to "*view* up to ten-minute, unaltered video clips of copyrighted content."[424] The court found that because "TVEyes ma[de] available virtually the entirety

---

417 *See supra* Section I A.

418 *See, e.g.*, Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015); Fox News Network, LLC v. TVeyes, Inc., 883 F.3d 169 (2d Cir. 2018); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007).

419 *Authors Guild*, 804 F.3d at 229; *TVeyes*, 883 F.3d at 169.

420 *See Authors Guild*, 804 F.3d at 218.

421 *See id.* at 222.

422 *TVeyes*, 883 F.3d at 174–75.

423 *Id.* at 177.

424 *Id.*

of the Fox programming that TVEyes users want to see and hear[,]" TVEyes' Watch feature was infringing and not fair use.[425]

Both of the Second Circuit's decisions in these cases reflect the careful balance of copyright law, and illustrate the importance of copying to creation of new and innovative services online.

4. Copyright Preempts Equivalent State Claims Involving Copying of Data

Section 301 of the Copyright Act preempts state laws which involve "legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright," such that "no person is entitled to any such right or equivalent right in any such work under the common law or statutes of any State."[426] Trespass, misappropriation, state computer crime statutes, and certain actions for breach of contract are preempted by the copyright act to the extent they attempt to create rights that are equivalent to those of copyright. Data scraping cases argued under trespass in particular often attempt to create property rights online in websites and web content, and should be analyzed carefully to determine the nature of the underlying claims. If a trespass or misappropriation claim argues, in essence, that a scraper copied data without permission, this claim should either be dismissed as duplicative of a CFAA claim, if one has been alleged, or as preempted by copyright law.

CONCLUSION

PowerVentures and hiQ both used data scraping of social media to power their services, but their uses of scraping had very different consequences for user privacy. hiQ's service is potentially harmful to LinkedIn's users, as it alerts their employers of the possibility that they may be seeking other job opportunities.[427] PowerVentures, in contrast, expressed commitment to the privacy of its users, but copied users' Facebook

---

425 *Id.* at 179.

426 17 U.S.C. § 301(a) (2012).

427 hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1104 (N.D. Cal. 2017).

friends' data as well as their own.[428] Even if the CFAA were to be amended, and copyright applied in data scraping cases, these privacy harms would still not be redressed.

The emergence of user-based services since the 1990s and the constant threat of large data breaches have led to increasing concerns about user privacy and security. The bargain offered by user-based services such as Google and Facebook is that, in exchange for free online services, users give these companies—whose core business is usually advertising—the ability to use their data for ad targeting.[429] Companies like Google, Facebook, and Microsoft have access to enormous amounts of personally identifying PII, as well as non-identifying but sensitive information like search history, browsing history, and private communications.[430] User-based platforms collect large amounts of personal data, including "'volunteered data' shared intentionally by consumers, 'observed data' obtained by recording consumer actions online, and 'inferred data' derived from analyzing volunteered and observed data."[431]

Many of the wrongs that occur on the Internet are fundamentally privacy wrongs, but the U.S. sectoral model only protects specific kinds of data and combats specific kinds of harms,[432] such as those that are considered "unfair and deceptive acts or practices," which fall under FTC jurisdiction.[433]

---

428 *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1062 (9th Cir. 2016), *cert. denied*, 138 S. Ct. 313 (2017).

429 Mark Hachman, *The Price of Free: How Apple, Facebook, Microsoft and Google Sell You to Advertisers*, PCWORLD (Oct. 1, 2015 3:00 AM), https://www.pcworld.com/article/2986988/privacy/the-price-of-free-how-apple-facebook-microsoft-and-google-sell-you-to-advertisers.html [https://perma.cc/9UQ2-ZCGY]; Jathan Sadowski, *Companies Are Making Money from Our Personal Data – but at What Cost?,* GUARDIAN (Aug. 31, 2016 9:00 AM), https://www.theguardian.com/technology/2016/aug/31/personal-data-corporate-use-google-amazon [https://perma.cc/MGH5-V5FZ].

430 *See e.g.*, *In re* Google, Inc. Privacy Pol'y Litig., No. C-12-01382-PSG, 2013 WL 6248499, at *2 (N.D. Cal. Dec. 3, 2013); *In re* Facebook Internet Tracking Litig., 140 F. Supp. 3d 922, 927–28 (N.D. Cal. 2015).

431 Maureen K. Ohlhausen & Alexander P. Okuliar, *Competition, Consumer Protection, and the Right (Approach) to Privacy*, 80 ANTITRUST L.J. 121, 131 (2015).

432 *See supra* Part II.

433 15 U.S.C. § 45(a)(1). For example, the FTC has proposed that services based on user data implement a "privacy by design" model, which includes "providing reasonable security for consumer data, collecting only the data needed for a specific business

Commentators have suggested that the U.S. adopt a European-style privacy model, creating individual rights around data privacy.[434] While this type of approach would ensure privacy, any policy adopted by the U.S. should be tailored to overall U.S. policy goals and a U.S. style of governance. One way to accomplish such a style of governance would be to enact a federal statute that encourages privacy class actions. The statute would create statutory damages, and could potentially be based around creating a duty owed to users of these popular user-based services.

---

purpose, retaining data only as long as necessary to fulfill that purpose, safely disposing of data no longer being used, and implementing reasonable procedures to promote data accuracy." *Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers*, FTC (Dec. 2010), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-bureau-consumer-protection-preliminary-ftc-staff-report-protecting-consumer/101201privacyreport.pdf [https://perma.cc/Q9LK-RYWQ].

[434] *E.g.*, *America Should Borrow from Europe's Data-Privacy Law*, ECONOMIST (Apr. 5, 2018), https://www.economist.com/news/leaders/21739961-gdprs-premise-consumers-should-be-charge-their-own-personal-data-right [https://perma.cc/78AU-AMK9]. Individuals in the EU have a right to data portability, allowing users to move their data from one service to another. *See* Paul De Hert, et al., *The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services*, 34 COMPUT. L. & SEC. REV., no. 2, 193, 195 (Apr. 2018).