Before the

Federal Trade Commission

Hearings on Competition and Consumer Protection in the 21st Century
Project Number P181201

Comments on Topic 4: The Intersection Between Privacy, Big Data, and Competition

August 17, 2018

Submitted by:
Electronic Frontier Foundation
Bennett Cyphers
Jamie L. Williams
815 Eddy St
San Francisco, CA 94109
(415) 436-9333
bennett@eff.org
jamie@eff.org

The Electronic Frontier Foundation is the leading nonprofit organization defending civil liberties in the digital world. Founded in 1990, EFF champions user privacy, free expression, and innovation through impact litigation, policy analysis, grassroots activism, and technology development. We work to ensure that rights and freedoms are enhanced and protected as our use of technology grows. EFF represents over 40,000 dues-paying members, including consumers, hobbyists, artists, computer programmers, entrepreneurs, students, teachers, and researchers.

Increasing market concentration and structural barriers to competition for Internet-related businesses threaten the values of free expression, privacy, and the innovation that has made the Internet a powerful force in daily life. It is imperative that policymakers and industry address competition issues actively and thoughtfully, avoiding approaches that will themselves harm the rights and freedoms of Internet users, or impede innovation.

A. Access to Data—and User Control Over Their Own Data—Is Critical for Competition.

In today's data-driven world, access to data is critical for competition. The web's largest platforms are well aware of this; their companies were built on consumer data. These same companies are now attempting to stop competition by cutting off competitors' access to publicly available data, blocking interoperable technologies, and failing to give users any meaningful ability to transport their data to other platforms. This is a problem. Fostering competition is an important component of the fight for online civil liberties. Consumers suffer when they have to rely on just a few platforms to communicate and learn online, and to protect their rights. Those few, dominant platforms have little incentive to protect user privacy, and sometimes even to maintain robust security practices to protect users, and they often substitute their own view of what constitutes valuable speech for that of their users or the broader public.

Facebook and its subsidiaries, for instance, are over ten times more valuable than the next two largest social media companies outside China—Twitter and Snapchat—combined. The company has cemented its dominance by buying out potential competitors before they've had a chance to grow (like Instagram) and waging wars of attrition against others (like Snapchat) when it can't. Because of its massive reach across much of the world, the platform can effectively censor public speech,[1] perform psychological experiments,[2] and potentially sway elections on the scale of a nation-state. If users don't like the way Facebook wields this power, there is nowhere else as ubiquitous or as well-populated for them to go. Facebook's trove of user data is its most valuable asset, which presents a dilemma. Thanks to network effects,[3] every user who joins a social network makes it more valuable for advertisers and more useful to everyone else. Without some access to the data Facebook has, it is virtually impossible for upstart platforms to compete with the behemoth now used by nearly a third of the world.[4]

To protect consumers and ensure competition in a data-driven world, efforts to maintain monopolistic control over Internet users and their data must be stopped, and Internet users must be given meaningful control of their own data.

B. Protecting Competition Requires New Privacy Laws and Regulations.

Protecting competition—by stopping efforts to maintain monopolistic control over user data and granting users meaningful control of their own data—requires well thought out privacy laws and regulations.

For many years, EFF has urged technology companies, legislators, and regulators to do a better job of giving technology users control over their data and protecting their users' privacy and civil liberties. EFF has, and continues to,[5] pressure the companies themselves, in hopes that mature players would see the importance of implementing real privacy protections. However, where voluntary efforts had failed to protect consumers, well thought out regulation is needed. [6]

---

[1] https://www.onlinecensorship.org/.

[2] https://www.theguardian.com/technology/2014/jul/02/facebook-apologises-psychological-experiments-on-users.

[3] https://www.vox.com/videos/2018/4/11/17226430/facebook-network-effect-video-explainer.

[4] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

[5] *See, e.g.,* https://www.eff.org/deeplinks/2018/07/facing-facebook-data-portability-and-interoperability-are-anti-monopoly-medicine.

[6] *See, e.g.,* https://www.eff.org/deeplinks/2018/08/how-improve-california-consumer-privacy-act-2018.

In particular, we believe new regulations should:

1. Address when and how online services must acquire affirmative user consent before collecting or sharing personal data, particularly where that data is not necessary for the basic operation of the service. Any request for opt-in consent should be easy to understand and should clearly advise the user what data the operator seeks to gather, how the operator will use it, how long the operator will keep it, and with whom the operator will share it. The request should be renewed any time the operator wishes to use or share data in a new way or gather a new kind of data. And the user should be able to withdraw consent, including for particular purposes.

2. Create an affirmative "right to know" so that users can learn what personal data companies have gathered about them, where they got it, and with whom these companies have shared it (including the government).

3. Create an affirmative right to data access and "data portability," so users can get a complete copy of their data from a service provider and move it to a different platform. The data should be easy to understand, machine-readable, and available in widely-adopted standard formats when applicable.

4. Create new mechanisms for users to hold companies accountable for data breaches and other privacy failures. Companies that suffer data breaches should have to report the breaches to the public in a timely manner. In cases where a breach was caused by inadequate security practices, it should be easier for the people harmed—including those suffering non-financial harms—to take negligent companies to court.

Any new regulations must be judicious and narrowly tailored. Overly burdensome regulation and technology mandates can stifle innovation, especially by small companies. If regulations are too byzantine, only the largest corporations will be able to comply, and the regulations themselves will act as a barrier to entry for smaller competitors. To that end, policymakers should consider tailoring new obligations based on the size of the service in question, taking into account revenue, number of employees, number of users, and other factors.

There are many other considerations to take into account when drafting new privacy rules for the digital economy. For example, policymakers should consider whether and how the rights and obligations they create can be waived, especially where users and companies have unequal bargaining power, such as when a user is prompted to agree to dozens of pages of terms before using a service. Privacy regulations must be balanced against First Amendment interests. For example, the "right to know" should extend to data a newspaper's website has collected about a user's browsing habits, but must not cover a reporter's investigative file. And at every step, policymakers should consult with data experts so they understand what data can be collected and used, under what circumstances.

There are few easy answers in privacy regulation, and no new regulation will ever be a panacea. Still, we believe the stakes are too high for inaction. We hope the Commission will use the powers at its disposal, including its Section 5 authority and sound recommendations to Congress, to advance sensible, user-friendly privacy laws and regulations.

C. Protecting Competition Also Requires Stopping Efforts by Major Internet Platforms to Use Computer Crime Statutes to Maintain Monopolistic Control Over Data.

New legislation is not enough. To ensure competition in a digital age, we must also put an end to efforts to abuse existing laws to maintain monopolistic control over user data.

Specifically, major Internet companies are currently attempting to co-opt a notoriously imprecise, per-Internet criminal anti-"hacking" statute intended to target computer break-ins, the Computer Fraud and Abuse Act ("CFAA"), and their state law equivalents, and transform these laws into tools for conducting anti-competitive behavior under the color of the law.[7] To protect competition, abuse of the CFAA must stop.

Congress passed the CFAA—which has been dubbed the "worst law in technology"[8]—in 1986, in response to a series of malicious computer break-ins. The law makes it a crime to access a computer "without authorization" but fails to tell us what that means. This vague language has enabled the law to metastasize in some jurisdictions from a law meant to target malicious "hacking" of private computer systems, into a tool for companies and websites to selectively enforce their computer use preferences and policies—such as terms of service prohibitions on using automated web browsing tools to access information—against competitors.

Platforms have taken advantage of this in a number of ways. In recent years, large companies—including Microsoft-owned LinkedIn[9]—have amped up efforts to use the CFAA's civil enforcement provision to punish competitors for using commonplace automated web browsing tools to access information they've published publicly online for the rest of the world to see. As USC Gould Law Professor Orin Kerr has explained, however, posting information publicly on the web and then telling someone they are not authorized to access it is "like publishing a newspaper but then forbidding someone to read it."[10] This is a clear abuse of a law meant to target criminals.

Automated web browsing—also referred to as "web scraping"[11]—is the process of using a computer script to send tailored queries to websites to retrieve specific pieces of content. The technique is used across the web for countless applications, such as aggregating information from multiple sources and identifying and extracting data for analysis.

---

[7] *See generally* Jamie L. Williams, *Automation is Not "Hacking": Why Courts Must Reject Attempts to Use the CFAA as an Anti-Competitive Sword*, 24 B.U. J. Sci. & Tech. L. X (forthcoming 2018).

[8] https://www.newyorker.com/news/news-desk/fixing-the-worst-law-in-technology.

[9] https://www.eff.org/deeplinks/2017/08/judge-cracks-down-linkedins-shameful-abuse-computer-break-law.

[10] *See* Orin S. Kerr, *Norms of Computer Trespass*, 116 Colum. L. Rev. 1143, 1169 (2016).

[11] https://www.eff.org/deeplinks/2018/04/scraping-just-automated-access-and-everyone-does-it.

The web is the largest, ever-growing data source on the planet. It's a critical resource for journalists, academics, businesses, and everyday people alike. Meaningful access sometimes requires the assistance of technology, to automate and expedite an otherwise tedious process of accessing, collecting and analyzing public information. As a technical matter, web scraping is simply machine automated web browsing. There is nothing that can be done with a web scraper that cannot be done by a human with a web browser. As one district court judge recently recognized, Web scraping "is merely a technological advance that makes information collection easier; it is not meaningfully different from using a tape recorder instead of taking written notes, or using the panorama function on a smartphone instead of taking a series of photos from different positions."[12]

Use of automated web browsing can help competition by lowering startup information barriers[13] and enable consumers to find deals and discounts online.[14] It can also help uncover unfair or deceptive business practices. ProPublica, for example, used automated web browsing to uncover that Amazon's pricing algorithm was hiding the best deals from its customers.[15] And because broader access to datasets can help correct bias in how algorithms are currently trained, it can also help identify and correct issues of algorithmic bias.[16]

It is important to understand that web scraping is a *widely used* method of interacting with the content on the web: everyone does it—even (and especially) the companies trying to convince courts to punish others for the same behavior. Companies use automated web browsing products to gather web data for a wide variety of uses. [17] Some examples from industry include manufacturers tracking the performance ranking of products in the search results of retailer websites, companies monitoring information posted publicly on social media to keep tabs on issues that require customer support, and businesses staying up to date on news stories relevant to their industry across multiple sources. E-commerce businesses use automated web browsing to monitor competitors' pricing and inventory, and to aggregate information to help manage supply chains. Businesses also use automated web browsers to monitor websites for fraud, perform due diligence checks on their customers and suppliers, and to collect market data to help plan for the future. Gartner has even recommended that all businesses treat the web as their largest data source and predicts that the ability to compete in the digital economy will depend on the ability

---

[12] *Sandvig v. Sessions*, No. CV 16-1368 (JDB), 2018 WL 1568881, at *7 (D.D.C. Mar. 30, 2018).

[13] *See* Rory Van Loo, *Rise of the Digital Regulator*, 66 Duke L.J. 1267, 1285–89 (2017).

[14] *See* Complaint, Sw. Airlines Co. v. Roundpipe LLC, No. 3:18-CV-33 (N.D. Tex. filed Jan. 5, 2018) (lawsuit by Southwest Airlines against a company that used automated web browsing software to enable customers to check flight prices and take advantage of the airline's own rebooking deals).

[15] https://www.propublica.org/article/how-we-analyzed-amazons-shopping-algorithm.

[16] *See* Amanda Levendowski, *How Copyright Law Can Fix AI' s Implicit Bias Problem*, 93 Wash. L. Rev. (forthcoming 2018).

[17] https://www.import.io/post/13-ways-use-web-scraping-tools/.

to curate and leverage web data: "Your company's biggest database isn't your . . . internal database. Rather it's the Web itself."[18]

Even the very companies trying to misuse the CFAA to punish competitors for using automated web browsing tools have used—and continue to use—these same techniques to build their businesses.[19]

Boston University Law Professor Andrew Sellars recently analyzed the sixty-one opinions generated via web scraping cases in the last twenty years. He reported that the "vast majority of these opinions concern claims brought by direct commercial competitors or companies in closely adjacent markets to each other."[20] The CFAA is first and foremost a criminal statute. The fact that these unauthorized Web scraping cases are *consistently* about blocking competition—and not about punishing criminals for breaking into private computer systems—demonstrates that the law is clearly being abused.

The companies seeking to abuse the CFAA in this way are subverting the web's open access norms.[21] These short-sighted and opportunistic efforts threaten open access to information across the Internet, including by investigative journalists, researchers, academics, and individual consumers. And in an era of algorithms and artificial intelligence, lack of access to data is a barrier to product innovation and competition.

LinkedIn, one company involved in recent web scraping litigation, characterizes its reliance on the CFAA as about protecting user privacy, not about stifling competition.[22] But the company's proposed rule—imposing criminal CFAA liability for automated access of publicly available user data by competitors that LinkedIn has told to "go away"—will not truly protect the privacy interests of LinkedIn users who decide to publish their information publicly online. The data will still be freely available on the Web for anyone else to access and use, without consequence. LinkedIn's privacy policy acknowledges the inherent lack of privacy in data users post publicly on its site and makes no promises to users about LinkedIn's ability to protect it:

---

[18] https://www.forbes.com/sites/gartnergroup/2015/02/12/gartner-predicts-three-big-data-trends-for-business-intelligence/.

[19] Microsoft-owned LinkedIn, for example, one company seeking to use the CFAA to block automated Web scraping by a competing service, acknowledges in its privacy policy that it uses automated tools, *i.e.*, Web scraping, to "collect public information about you, such as professional-related news and accomplishments" and makes that information available on its own website—unless a user opts out via adjusting their default privacy settings. *See* LinkedIn, Privacy Policy, §§ 1.1-1.2 (effective May 8, 2018), https://www.linkedin.com/legal/privacy-policy.

[20] *See* Andrew Sellars, *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act*, 24 B.U. J. Sci. & Tech. L. 424, X (forthcoming 2018) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3221625).

[21] *See* Orin S. Kerr, *Norms of Computer Trespass*, 116 Colum. L. Rev. 1143, 1162–64 (2016).

[22] See *hiQ Labs, Inc. v. LinkedIn Corp.*, 273 F. Supp. 3d 1099, 1106 (N.D. Cal. 2017).

"Please do not post or add personal data to your profile that you would not want to be publicly available."[23] What is needed to protect privacy is comprehensive, thoughtful privacy regulation that LinkedIn, its parent company Microsoft, and all other websites and Internet service providers would be subject to.[24]

Platforms have also used the CFAA to go after companies for creating interoperable software and shut down follow-on innovation. Social media giant Facebook, for example, for a decade has pursued litigation against a company that tried back in 2008 to provide a social media aggregation service for users of Facebook and other social media platforms.[25] This service, had it not been stifled in the cradle, could have been a great boon to those who often switch between services like Facebook, LinkedIn, and Twitter, or who struggle to remember who's a friend, who's a contact, and who's a follower, or where they received any given message. Facebook sent the company, Power Ventures, a cease and desist letter and set up an ineffective IP address block. When Power continued to provide its social media aggregation services to Facebook users, Facebook turned to the CFAA. In order to provide its aggregation services, Power Ventures had used—with permission—the valid Facebook login credentials of its users. Facebook claimed that Power Ventures had violated the CFAA by continuing to use these valid credentials after receipt of the cease and desist letter. And in 2016, it convinced the Ninth Circuit to go along with this theory of liability. At Facebook's urging, the court contorted previously clear CFAA precedent and opened the door for even more abuse of the CFAA,[26] including many of the pending automated web browsing cases that are threatening competition and open access across the web today (which consistently rely on the Ninth Circuit's decision in this case).

D. Mergers Involving Data from Third-Party Trackers—Including User Location Data—Must Receive Special Scrutiny.

Finally, to protect both competition and consumers, merging of rich first-party datasets with *third-party trackers*—systems that use ads and other third-party plugins to track user habits around the web and on mobile devices—must receive special scrutiny. Such mergers present privacy risks to users and exacerbate existing network effects and make it difficult for companies without comparable datasets to compete.

In 2007, Google purchased Doubleclick, a third-party advertising and tracking company. The merger was reviewed by the Commission at the time, and the majority determined that the competition and privacy concerns were not sufficient to challenge the acquisition. In 2013, Facebook acquired a similar product, Atlas, from Microsoft, which they have since folded into their own brands.

---

[23] See LinkedIn, Privacy Policy, § 1.1 (effective May 8, 2018), https://www.linkedin.com/legal/privacy-policy.

[24] *See* Section B, *supra*.

[25] https://www.eff.org/cases/facebook-v-power-ventures.

[26] https://www.eff.org/deeplinks/2016/12/take-two-ninth-circuit-revises-two-password-sharing-decisions-fails-fix-cfaa-mess.

Today, Facebook's and Google's tracking networks are the two largest on the English-speaking Internet by far. Facebook tracking code, including social plugins and its invisible "pixel," is present on nearly 25% of the top one million sites on the Internet. The company's ad network also covers 40% of the top 500 most popular mobile apps. By some metrics, Google's reach is even broader. Rich tracking code for Doubleclick is present on over 20% of the top million sites; including Google Analytics and other services, code from Google is present on approximately three quarters of sites on the web.

In addition to their third-party tracking capabilities, both of these companies have massive first-party data stores. That gives them the ability to link data from their third party trackers with the data that users have provided them voluntarily, including real names, demographic data, contacts, communication, and interests.

We believe these kinds of mergers and acquisitions raise both privacy and competition concerns.

From a privacy perspective, mergers between tracking companies and first-party data stores create risks to users that are not present in their component parts. Normally, third-party tracking companies creates anonymous, ad-hoc profiles for users as they browse the web. They have difficulty linking one user's activity across different devices, and when a user clears cookies or switches to a new browser, the tracking company may have to start building a new profile from scratch. However, when a Facebook user browses the web, their activity can be immediately and permanently linked to their Facebook identity via Facebook's cookies. When a user uploads a photo or comments on a friend's post, they implicitly consent to giving the company their data. But when they leave facebook.com to browse the web, they may not realize that Facebook is *still tracking them*. Even if they do, the company offers no way to opt out of that collection or to delete the data after the fact. The result is a potent, permanent profile of that user's digital life, combining data they have chosen to share with data collected surreptitiously while they might have felt anonymous.

From a competition perspective, the mergers exacerbate existing network effects and make it difficult for companies without comparable datasets to compete. They give the companies competitive advantages for both their first-party platforms and third-party advertising products. Facebook touts their ability to advertise to "real people"—that is, to use information from Facebook profiles to target individuals outside of Facebook products. Third-party ad platforms that do not possess a similar first-party dataset cannot hope to do the same. Furthermore, these companies have a privileged view of the landscape of the Internet, and therefore of their competition. This gives some companies "a relative advantage in accessing and analyzing data to discern threats well before others, including the government."[27]

There are some behavioral remedies that we believe could mitigate the harms of these mergers. After acquiring Doubleclick, Google volunteered to keep the data it collected through

[27] *See* Maurice E. Stucke, *Should We Be Concerned About Data-opolies?*, 2 Geo. L. Tech. Rev. 275, 305 (2018) (available at https://ssrn.com/abstract=3144045 or http://dx.doi.org/10.2139/ssrn.3144045).

Doubleclick separate from the rest of its user data. Commissioner Harbour, in her dissenting statement for the investigation, predicted that the company would eventually reverse this policy, and in 2016, it did. Today, it might make sense to enforce a similar policy: require that data from third-party tracking networks must be "siloed" away from first-party data so that anonymous web activity cannot be linked to rich digital identities.

Finally, we believe traditional metrics for assessing these mergers are insufficient, and new means of evaluation are needed in the future. In her dissent, Commissioner Harbour wrote, "Traditional competition analysis of Google's acquisition of DoubleClick fails to capture the interests of all the relevant parties." We agree, and we believe that mergers between data collectors should be scrutinized more strictly than they have in the past, and on more comprehensive grounds. We hope to engage in an ongoing conversation about how to assess competitive harms caused by consolidation in the age of big data.