No. 19-783

IN THE
Supreme Court of the United States

NATHAN VAN BUREN,
Petitioner,

v.

UNITED STATES,
Respondent.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE ELEVENTH CIRCUIT

BRIEF OF AMICUS CURIAE THE MARKUP IN SUPPORT OF PETITIONER

Nabiha Syed
THE MARKUP
900 Broadway, Suite 202
New York, NY 10159
(347)894-3746
nabiha@themarkup.org

Katherine M. Bolger*
John M. Browning
DAVIS WRIGHT TREMAINE LLP
1251 Avenue of the Americas
New York, NY 10020
(212) 489-8230
katebolger@dwt.com

David M. Gossett
DAVIS WRIGHT TREMAINE LLP
1919 Pennsylvania Ave. NW
Washington, DC 20006

*Counsel of Record

Counsel for Amicus Curiae

LEGAL PRINTERS LLC ● Washington, DC ● 202-747-2400 ● legalprinters.com

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

Cases

Constitutional Provisions and Statutes

Other Authorities

INTEREST OF *AMICUS CURIAE*[1]

The Markup is a nonprofit news organization that conducts data-driven investigations into how powerful institutions use digital technology to reshape our society. Launched in February 2020, *The Markup* has assembled a staff of journalists who specialize in data journalism – also known as computational journalism – and publishes their work online at themarkup.org. The core mission of *The Markup* is to use the tools of data journalism to investigate how technology works, particularly when it is wielded by the online platforms, government entities, and other organizations that play an increasingly prominent role in our online (and real world) lives. *The Markup* thus has a compelling interest in ensuring that the newsgathering practices that make data journalism a valuable tool for holding the powerful accountable in our modern democracy are not chilled or made illegal by overbroad application of data security laws.

[1] All parties have provided their consent to the filing of this brief. No counsel for a party authored this brief in whole or in part, and no person or entity other than *amicus* and its counsel made a monetary contribution to the preparation or submission of this brief.

INTRODUCTION AND SUMMARY OF ARGUMENT

In 1895, pioneering journalist Ida B. Wells published *The Red Record: Tabulated Statistics and Alleged Causes of Lynching in the United States, 1892-1894*. Ida B. Wells-Barnett, THE RED RECORD (1895). In that publication and several others like it, Wells combed through public records and compiled statistical charts to demonstrate the frequency with which men, and particularly Black men, were lynched. The purpose of Wells' statistical analysis was not merely actuarial. Rather, it served as the raw material for her news reporting, which changed the way lynching was perceived and debated, both by its critics and its supporters. The power of Ms. Wells' reporting lay in her devotion to facts. As she wrote, "[n]o good result can come from any investigation which refuses to consider the facts. A conclusion that is based upon a presumption, instead of the best evidence, is unworthy of a moment's consideration." Ida B. Wells-Barnett, LYNCHING AND THE EXCUSES FOR IT (1901). By carefully gathering and marshalling indisputable facts from public records scattered across the nation, Ms. Wells revealed the frightful scale of lynching in this country; without her work, the truth may have remained hidden in plain sight.

The Markup continues the tradition of data journalism that Ms. Wells pioneered, the need for which could not be greater today. Like Ms. Wells, *The Markup* seeks to inform and influence public debate by marshaling the "best evidence" available – which is now increasingly online. Whereas Ms. Wells forced society to confront the evils of lynching by manually

searching paper records, modern data journalists analyze enormous datasets to identify abuses of power in the modern world. With increasingly powerful and sophisticated tools at their disposal, data journalists have exposed countless injustices, including: systemic racial discrimination by banks refusing to issue loans to Black customers; the re-segregation of Florida schools after integration was no longer enforced; and flaws in algorithms designed to assist with sentencing, which incorrectly rated Black defendants as posing higher risks of reoffending than White defendants.[2] Data journalists have also created public databases that track police shootings and offer sophisticated insight into our political system by aggregating individual polls.[3]

All of these projects illuminate the invisible forces at work in peoples' lives. The common thread running back to Ms. Wells is that journalists can use data to observe and critique powerful entities or entrenched systems. It is impossible to report on systematic

---

[2] Bill Dedman, *The Color of Money*, ATLANTA JOURNAL-CONSTITUTION (1988), http://powerreporting.com/color/color_of_money.pdf; Nathaniel Lash, *Why Pinellas County is the worst place in Florida to be black and go to public school*, TAMPA BAY TIMES (Aug. 12, 2015), https://www.tampabay.com/news/education/k12/why-pinellas-county-is-the-worst-place-in-florida-to-be-black-and-go-to/2241065/; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[3] Wesley Lowery et al., *994 people shot dead by police in 2015*, WASH. POST (2015), https://www.washingtonpost.com/graphics/national/police-shootings/; Nate Silver et al., *Latest Polls*, FIVETHIRTYEIGHT.COM (June 7, 2020), https://projects.fivethirtyeight.com/polls/president-general/national/.

trends in our society caused by powerful governmental or private actors – such as racial discrimination, police misconduct, or anti-competitive behavior – without collecting enough information to understand the big picture. Indeed, many of these abuses would remain invisible if journalists lacked the power to collect raw data from online sources. Data journalism thus enables citizens to comprehend our increasingly complex world and, in doing so, gives us all the power to demand change.

Yet the newsgathering techniques that make data journalism possible could be effectively outlawed by the broad reading of the CFAA that the government urges and that the Eleventh Circuit endorsed. As the Eleventh Circuit acknowledges, its interpretation of the CFAA empowers the operator of any website "to legislate what counts as criminal behavior through their internal policies or terms of use." Pet. App. 27a. In practice, this means that journalists may risk criminal liability – and prison time – for violating even the most trivial terms of service for any website they investigate. Worse still, any website owner seeking to avoid legitimate oversight could instantly, or perhaps even retroactively, make it a crime for journalists to investigate their services by deliberately tweaking the terms of service to prohibit newsgathering activities. Many websites have already taken such steps. Jacquellena Carrero, Note, *Access Granted: A First Amendment Theory of Reform of the CFAA Access Provision*, 120 COLUM. L. REV. 131, 134 (2020); Komal S. Patel, Note, *Testing the Limits of the First Amendment: How Online Civil Rights Testing Is Protected Speech Activity*, 118 COLUM. L. REV. 1473, 1494 (2018).

The chilling effect on data journalism is obvious. *The Markup*'s reporters and other data journalists routinely harvest data from government databases and digital platforms. This is not "hacking." Rather, *The Markup*'s reporters gain authorized access to the data – often by setting up legitimate accounts – and have no more (or less) access than any other authorized user. They do not intentionally circumvent technology designed to limit access to non-public information and they endeavor – as far as possible – to abide by the terms of service for every website they investigate. And, because its reporters have nothing to hide, *The Markup* publishes "Show Your Work" articles describing how data was collected for any given news article. But even with these efforts, *The Markup*'s journalists could violate the CFAA because it is impossible to guarantee perfect compliance with every website's "protean" terms of use, which are invariably "long, dense, and subject to change." *Sandvig v. Barr*, --- F. Supp. 3d ---, 2020 WL 1494065, at *10 (D.D.C. Mar. 27, 2020), *appeal filed*, No. 20-5153 (D.C. Cir. May 28, 2020).

The First Amendment cannot tolerate such arbitrary restrictions on newsgathering. In an unbroken line of cases establishing the right to publish lawfully obtained information concerning a matter of public interest, this Court has made it clear that the government cannot criminalize "routine newspaper reporting techniques." *Smith v. Daily Mail Publ'g Co.*, 443 U.S. 97, 103 (1979). This Court has also held that journalists cannot be punished for obtaining and publishing newsworthy information after it has been "publicly revealed," even if that information was legally required to remain secret.

Oklahoma Publ'g Co. v. District Ct., 430 U.S. 308, 311 (1977). It makes no difference to the First Amendment whether the information was gathered online or in an analog format. As this Court has recognized, the Internet is a "vast democratic forum[]," *Reno v. ACLU,* 521 U.S. 844, 868-69 (1997) – the "modern public square." *Packingham v. North Carolina*, 137 S. Ct. 1730, 1737 (2017). Data journalism serves an important role in helping Americans observe and understand the forces at play in that public square that might otherwise remain hidden. A statute that allows powerful forces like the government or wealthy corporate actors to unilaterally criminalize newsgathering activities by blocking these efforts through the terms of service for their websites would violate the First Amendment.

But this Court can avoid a collision course with the First Amendment by giving the CFAA the rational interpretation urged by petitioner. As Judge Bates recently explained in rejecting the government's broad reading of the CFAA, "[i]f this Court determines that the [CFAA] does *not* actually criminalize" journalists for "violating consumer websites' terms of services," then "the Court need not dive … into the First Amendment policy." *Sandvig*, 2020 WL 1494065, at *7. Accordingly, this Court should reverse the Eleventh Circuit and hold that, if a person has authority to access material on a computer system, that access does not violate the CFAA – even if it is for a reason beyond those for which access was granted.

ARGUMENT

The Markup is at the vanguard of data journalism, which has a venerable tradition but has rapidly

evolved in response to the explosion of data now available to reporters. Essentially, data journalists use digital tools to do what investigative reporters have always done: gather the information necessary to reveal matters of public concern that would otherwise remain hidden, and then translate it into news stories for the public's benefit. The First Amendment protects basic newsgathering techniques that make journalism possible. This Court should reject a broad reading of the CFAA that would thwart those First Amendment principles by potentially outlawing the tools that *The Markup* and other news organizations use to gather data online for the purpose of investigative reporting.

I. Data Journalism Is a Vital Form of Journalism.

The roots of data journalism can be found in the public-service tradition of investigative journalism. The tradition stretches back at least as far as Ms. Wells and other reporters who manually gathered, analyzed, and published data. This form of journalism is necessary in order to identify structural problems or patterns in our society that cannot be discerned by looking at incidents in isolation. For instance, Ms. Wells' masterful use of data disproved the fiction that lynching was sporadic and showed it for what it was – a systemic campaign of terror.

Another classic example of data journalism is *The Color of Money*, a Pulitzer-Prize winning series of reports about racially discriminatory mortgage lending practices that appeared in *The Atlanta Journal-Constitution* in 1988. Dedman, *supra* note 2. For his classic report, Bill Dedman used the Freedom

of Information Act to obtain "computer tape" containing lenders' reports relating to a total of 109,000 loan decisions made in metro Atlanta between 1981 and 1986. *Id*. Using statistical analysis and publicly available demographic information, Mr. Dedman was able to determine that "[w]hites receive five times as many home loans from Atlanta's banks and savings and loans as blacks of the same income – and that gap has been widening each year." *Id*. at 3. The power of Mr. Dedman's reporting lay once again in his use of data, which proved that the discrimination was not intermittent but rather an institutional practice across a whole city. Once these systemic injustices were revealed, it was possible to effect meaningful change.

In the years since, the growth of personal computing and the Internet has caused an explosion in the amount of data that our society generates – which, in turn, has expanded the horizon of what is possible through data journalism. Since the beginning, the unique strength of data journalism has been its ability to discern and articulate structural problems with our democracy by identifying trends that are simply not apparent from focusing on one-off anecdotes. A prime example of the insights made possible by data journalism is *The Washington Post*'s initiative led by Wesley Lowery to create a national database tracking police shootings, which won the 2015 Pulitzer Prize. The information in the database was used as the basis for a series of reports about the frequency of police shootings and who the victims were likely to be; it exposed the reality that "unarmed black men are seven times more likely than whites to die by police gunfire." Sandhya Somashekhar et al.,

Black and Unarmed, WASH. POST (Aug. 8, 2015), https://www.washingtonpost.com/sf/national/2015/08/08/black-and-unarmed/?itid=sf_. There is a clear link between this reporting and Wells' earlier work on lynching. In both cases – and Dedman's *Color of Money* series – the data allowed journalists to observe and expose injustices that would have otherwise remained hidden.

Indeed, within the last decade, news organizations like *The Markup*, *ProPublica*, *The Intercept*, *Five Thirty Eight*, and others – as well as journalists at traditional news organizations – have pioneered new analytical reporting techniques that have made it possible to use data to reveal hitherto unseen truths about the way our world works. The discipline has developed rapidly to adapt to a society that is increasingly awash with data, and data journalists have won at least 3 Pulitzer Prizes as well as countless other awards. The vibrant field of data journalism also covers a wide range of subject matter and generates a diverse range of output, from hard-hitting investigative reports about matters of intense public concern to innovative graphical reporting that uses pictures or video to bring data to life. A few examples include:

- *ProPublica* reporters demonstrated how certain online hate groups were able to monetize their content despite websites' claims to the contrary. Julia Angwin et al., *Despite Disavowals, Leading Tech Companies Help Extremist Sites Monetize Hate*, PROPUBLICA (Aug. 19, 2017), https://www.propublica.org/

article/leading-tech-companies-help-extremist-sites-monetize-hate.

- *Taken* is a collaborative reporting effort to collect, analyze and publish data-driven investigative reports on the prevalence of civil asset forfeiture across the nation, which is a process by which police departments seize property from civilians and sell or keep it for their own benefit. *Taken*, PULITZER CENTER, https://taken.pulitzercenter.org/.

- *USA Today* exposed the prevalence of "fill-in-the-blank" legislation written by corporations, interest groups or lobbyists, revealing that legislatures introduced more than 10,000 model bills within the last eight years. *Rob O'Dell* et al., *Copy, paste, legislate*, USA TODAY (Apr. 3, 2019), https://www.usatoday.com/pages/interactives/asbestos-sharia-law-model-bills-lobbyists-special-interests-influence-state-laws/

- *Five Thirty Eight* and others aggregate data from different pollsters to create "polls of polls" that more accurately reflect public opinion. *See* Nate Silver et al., *Latest Polls*, FIVETHIRTYEIGHT.COM (June 7, 2020), https://projects.fivethirtyeight.com/polls/president-general/national.

Rather than speaking solely about data journalism in the abstract, however, we think it will help the Court to understand the implications of this case to explain how *The Markup* operates – and thus how the government's interpretation of the CFAA could

effectively prohibit *The Markup* from continuing to function.

II. The Markup Uses Data Journalism Techniques to Report on the Technology Shaping Our Society.

The Markup takes a scientific approach to news reporting that was created by the editor-in-chief and founder of *The Markup*, Julia Angwin. Ms. Angwin previously led investigative teams at *ProPublica* and *The Wall Street Journal*. She is a winner and two-time finalist for the Pulitzer Prize in journalism. *The Markup* employs a roster of data journalists, who have, over the course of their careers, used data-driven journalism to reveal racial biases in criminal sentencing software that discriminates against black defendants; document housing discrimination on Facebook; and break other news stories that would not have been possible without access to data and sound analytical reporting techniques. *See, e.g.*, Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016) (Pulitzer Prize Finalist for Explanatory Reporting); Julia Angwin, et al., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016). These journalists follow the "The Markup Method" of reporting, which is a three step process:

- **Build.** We ask questions and collect or build the datasets we need to test our hypotheses.

- **Bulletproof.** We bulletproof our stories through a rigorous review process, inviting external experts and even the subjects of investigations to challenge our findings.

- **Show our work.** We share our research methods by publishing our datasets and our code. And we explain our approach in detailed methodological write-ups.[4]

In short, *The Markup's* journalists are "dedicated to deciphering the invisible infrastructure of our world" because its readers "deserve to understand the forces that shape … reality." Nabiha Syed, *A Letter from the President*, THE MARKUP (Feb. 25, 2020), https://themarkup.org/2020/02/25/president-letter-nabiha-syed.

Much like the scientific method, The Markup Method is about developing conclusions from observable facts. In addition to being rigorous and transparent, *The Markup* – like many, but not all, other organizations devoted to data journalism – makes reasonable efforts not to intentionally violate the terms and conditions of the digital platforms it investigates. *The Markup* also aims to be transparent when it seeks to access data (to the extent possible) and does not deliberately circumvent technological measures website operators have imposed to restrict access to information. In other words, *The Markup* does not intentionally "hack" into sources of information that it is not authorized to see by breaking security measures and, to the extent possible, it takes precautions to avoid exceeding the terms of its authorized access.

---

[4] *See* Julia Anguin, *A Letter from the Editor*, THE MARKUP (Feb. 25, 2020), https://themarkup.org/2020/02/25/editor-letter-julia-angwin.

But even taking these precautions, data journalists could still face criminal liability for even minor or inadvertent infractions of a website's terms and conditions if this Court adopts the Eleventh Circuit's broad reading of the CFAA. That interpretation potentially criminalizes two of the most fundamental methods that data journalists like *The Markup* use to collect data: accessing publicly available digital platforms to investigate how they work, and using automated tools to extract content from web sites (also known as data scraping).

A. The Markup Gathers News by Observing What Can Be Done on Online Platforms.

One of the most elementary ways that *The Markup* gathers information is for one of its reporters to simply sign up for an account on an online service – whether private-sector or governmental – and use that account to observe how the website works.

For instance, *The Markup* recently published an article that investigated whether Facebook allowed advertisers on its website to target people that Facebook identified as being interested in "pseudoscience," and thus possibly susceptible to dangerous conspiracy theories. *See* Aaron Sankin & Maddy Varner, *Want to Find a Misinformed Public? Facebook's Already Done It*, THE MARKUP (Feb. 25, 2020) https://themarkup.org/coronavirus/2020/04/23/want-to-find-a-misinformed-public-facebooks-already-done-it. In order to gather the facts needed to report this story, Maddy Varner, an investigative data journalist from *The Markup*, logged onto Facebook using her real name and *The Markup*'s official account. *Id.* Once she obtained authorized

access to the platform, she purchased advertisements and was able to target more than 50 million people on Facebook and Instagram who were deemed to be interested in "pseudoscience." *Id*. *The Markup* also found evidence that users interested in "pseudoscience" were being targeted with advertisements for products that were connected to prominent conspiracy theories – such as hats purporting to protect the user from electromagnetic radiation from 5G cellular equipment, which conspiracy theories have touted as a possible cause of COVID-19. Facebook ultimately removed its option to target users interested in "pseudoscience" in response to the story. *Id*.

The reporting techniques that *The Markup* used to test and report about Facebook's safeguards against the sale of pseudoscience items are hardly novel or controversial. It is a modern twist on a venerable reporting practice that stretches back at least as far as Nellie Bly, who posed as a psychiatric patient to gather information for a series of classic undercover reports about brutal conditions at the Women's Lunatic Asylum on Roosevelt Island. *See* Nellie Bly, *TEN DAYS IN A MAD HOUSE* (1887). *See also J.H. Desnick v. American Broadcasting Cos.*, 44 F.3d 1345, 52-56 (7th Cir. 1995) (holding that undercover recordings made by journalists posing as patients of an ophthalmologist for news report were lawful). And there is no question that the reporting served a public goal – once Facebook became aware of the pseudoscience category, it took steps to eliminate it.

Even so, these routine newsgathering efforts – which would be entirely unremarkable in the analog

world – could potentially expose *The Markup* and its reporters to severe criminal liability under the Eleventh Circuit's reading of the CFAA. Instead of acting responsibly, as they did in this example, all Facebook needed to do was change their terms of service to ban newsgathering and *The Markup* reporters would face criminal liability. Similarly, the Eleventh Circuit's rule would also have put *The Markup's* reporters at the mercy of Google when these reporters signed up for a Gmail account in order to test how Google handled emails from political campaigns. *See* Leon Yin & Adrianne Jeffries, *Show Your Work: How We Examined Gmail's Treatment of Political Emails*, THE MARKUP (Feb. 26, 2020). Moreover, website operators seeking to avoid scrutiny could change their terms of service to prohibit newsgathering and unilaterally – or perhaps even retroactively – make it a crime for any journalist to access a website in order to gather information for a news story. Allowing the investigated entity to so unilaterally control the fate of the investigator overturns long standing journalistic principles and burdens valuable newsgathering.

B. *The Markup* Uses Data Scraping Technology to Gather Information for its News Reporting.

Data scraping is the second common practice of data journalism used by *The Markup* that is imperiled by the Eleventh Circuit's overly broad reading of the CFAA. "Data scraping" is a term of art that refers to an automated process to extract content from a webpage using a specialized software tool or piece of computer code. It allows journalists to observe vast

quantities of data at scale. A well-known example of data scraping is the Internet Archive's Wayback Machine, which uses software to "crawl" through the world wide web by systematically visiting webpages at regular intervals and automatically downloading copies of any publicly available information. The result is a database that allows users to go "back in time" to see what websites looked like in the past, even after their owners update their content. [5]

In the journalism context, reporters routinely use automated data collection technology to assemble and review large datasets of publicly available information. The technology has two crucial benefits. The first is that it allows journalists to gather and organize data on a scale that would be practically impossible to do manually. While it would be theoretically possible for a reporter, for instance, to collect the same information manually from every website, scraping allows journalists to gather information on a scale that is commensurate with the gigantic amounts of information available. The second crucial function of scraping is to allow journalists to observe and understand trends. In the past, a journalist might carry out a consumer survey in the real world or systematically visit public record offices across the country to pull certain types of records.

---

[5] *See* Wayback Machine, INTERNET ARCHIVE, https://archive.org/web/. Every webpage that is copied by the Internet Archive's software (known as a "web crawler") is placed into an online database, where it can be searched by the time the webpage was copied and the URL address from which it was taken. *See Wayback Machine General Information,* INTERNET ARCHIVE, https://help.archive.org/hc/en-us/articles/360004716091-Wayback-Machine-General-Information.

Scrapers are tools that vastly expand journalists' ability to observe trends that can serve as the basis for news reporting, often in real time. As entities increasingly use data to widen their own audiences and customer bases, scrapers take advantage of the same technology to gather information. While it was possible for Ida B. Wells to gather the information she needed to expose lynching by physically visiting public records offices, scrapers make it possible for reporters to adapt the tradition of data-driven journalism to the demands of newsgathering in the digital realm.

As the President of *The Markup* put it at the 2018 Salant Lecture on Freedom of the Press at Harvard University, "[w]e need the media's truth telling role to be deciphering the hidden systems that result in observable facts." Nabiha Syed, President, *The Markup*, Salant Lecture on Freedom of the Press at the Harvard Kennedy Center (Nov. 14, 2018), https://shorensteincenter.org/nabiha-syed-2018-salant-lecture-freedom-press/. Scraping allows journalists to translate the data that websites make available to the public into a scale and form that can be observed critically.

A centerpiece of *The Markup*'s work is gathering large amounts of data by automated means, and its journalists thus routinely design computer code that collects information automatically or through systematic processes. *The Markup* – like most other data journalists – takes precautions so that its automated data retrieval applications do not intentionally access and collect information that would not available to an authorized human user. *The*

Markup also endeavors to design its scraping codes to avoid violating terms of service or circumventing technological restrictions on access. Finally, in an effort to maintain transparency and avoid any appearance of impropriety in the use of automated data gathering technology, *The Markup* publishes the code it uses to analyze data for its investigations and explains how that code operates in its "Show Your Work" articles.

A prime example of the unique possibilities of data scraping is *The Markup*'s forthcoming "Blacklight" project. This privacy inspection tool uses software designed by *The Markup* to ascertain the ways in which a website is secretly collecting data about its visitors. Blacklight works by using automated techniques to imitate the process of a human visiting a website. Blacklight launches a software program that visits a website and probes it to determine what types of data are being collected. Essentially it is the mirror image of the usual process whereby a website takes information to track a user; here, *The Markup*'s software examines a website and reports back what information is being taken. This tool – which *The Markup* will make freely available to the public – shows users how websites track them using their own personal data. In addition to allowing members of the public to use the tool, *The Markup* also intends to itself use Blacklight as the basis for investigative reporting on how website operators use surveillance technology to track users. Without automated scraping technology, this project would be impossible and the insights it can offer the public about how they are being tracked online would never come to light.

A broad reading of the CFAA would imperil this newsgathering by *The Markup* and similar newsgathering by other news organizations that seek to collect data in bulk using scrapers in order to inform the public. For example,

- Red Feed/Blue Feed, a project of the *Wall Street Journal*, which used data scraped from Facebook to compare the types of articles served to Facebook users across the political spectrum. John Keegan, *Blue Feed, Red Feed*, WALL STREET JOURNAL (Aug. 19, 2019), https://graphics.wsj.com/blue-feed-red-feed/.

- *ProPublica* has accumulated a database of up to 3 million tax returns from tax-exempt organizations, which it has made available to the public to search for free. *See* Ken Schwencke et al., *Nonprofit Explorer*, PROPUBLICA, https://projects.propublica.org/nonprofits/.

- Public interest organizations have used scraping to make public documents more freely accessible to the public, like judicial decisions on PACER. *See, e.g.*, Free Law Project, https://free.law/recap/.

In all of these cases, the data being collected by the scrapers remains available to any authorized user of the service from which it was obtained, and does not require journalists to circumvent technological barriers in order to gain access (*i.e.*, hacking). The data scraping employed is merely a mechanical means of collecting and analyzing information that would be equally accessible to any other user. In other words, it

is a tool that allows journalists to see and understand what is happening online – at a scale and with an acuity that is not possible by looking at isolated anecdotes.

But the scraping may not perfectly adhere to terms and conditions required by each website, either because perfect adherence would be impossible with such a large number of websites or because the terms of service of a website could change from one day to the next to purport to prohibit data scraping without a corresponding change to the relevant technological protocols. Under the Eleventh Circuit's broad reading of the CFAA, however, this inadvertent overreach might qualify as "misuse" of a computer system that triggers criminal liability. An endorsement of this interpretation by this Court would thus have an enormous chilling effect on the automated newsgathering techniques that have created so many exciting opportunities for data journalists seeking to create and analyze large datasets of publicly available information.

In sum, data journalists like *The Markup* rely on their ability to gather information from websites that would be available to any other legitimate user, often by means of automated technology. These basic and fundamentally important reporting techniques would be criminalized under the government's and Eleventh Circuit's interpretation of the CFAA.

III. The First Amendment Protects the Online Newsgathering Practices of Data Journalists.

Petitioner has explained (Pet. Br. 17-36) why the court of appeals' interpretation of the CFAA is wrong as a matter of statutory interpretation. If there were any question, though, the First Amendment would prohibit that reading of the CFAA because it violates the rights of journalists to gather and report on publicly available information. The CFAA cannot, consistent with the First Amendment, make it a crime for *The Markup* or any other news organizations like *The Markup* to gather publicly available information from a website that they have lawfully accessed. Such a result is prohibited by venerable decisions from this Court protecting the right to gather and report upon lawfully obtained information.

A. Routine Newsgathering Activities Cannot Be Criminalized Under the First Amendment Just Because They Were Conducted Online.

News reporting is crucial to the proper functioning of our self-governing democracy. Accordingly, the First Amendment provides journalists with broad protections against governments and individuals who seek to thwart legitimate newsgathering activities. While there are certainly limits on how far a reporter can go to report a news story, this Court has time and again made clear that journalists cannot be penalized for gathering and publishing publicly available, truthful information on a matter of public interest – even in the face of laws requiring the information to remain secret. An interpretation of the CFAA that

enables website owners to make it a crime for journalists to observe and report on information that any other user could see would clearly violate these deeply entrenched First Amendment principles. A precedent effectively foreclosing newsgathering online would also violate decisions of this Court urging "extreme caution before suggesting that the First Amendment provides scant protection for access to vast networks" available through "the modern Internet." *Packingham*, 137 S. Ct. at 1736.

The decisions of this Court establishing the rights of journalists to gather publicly accessible information on a matter of public interest were developed in the context of shoe-leather journalism, but are no less applicable here. In *Cox Broadcasting Corp. v. Cohn*, a reporter covering a rape-murder trial "learned the name of the victim from an examination of the indictments which were made available for his inspection in the courtroom." 420 U.S. 469, 472-73 (1975). He then broadcast the woman's name in a television news report, which indisputably violated a Georgia statute that prohibited journalists from publicizing the name of a rape victim. *Id.* But application of the state privacy law violated the First Amendment, this Court held, because "[o]nce true information is disclosed in public court documents open to public inspection, the press cannot be sanctioned for publishing it." *Id.* at 496. *See also Oklahoma Publ'g Co.*, 430 U.S. at 310 (holding that a state court cannot "prohibit the publication of widely disseminated information obtained at court proceedings which were in fact open to the public"); *Garrison v. Louisiana*, 379 U.S. 64, 74-75 (1964) ("Truth may not be the subject of civil or criminal

sanctions where discussion of public affairs is concerned," because such speech "is the essence of self-government.").

In a subsequent case, two newspapers learned the identity of a juvenile charged with a shooting "by monitoring the police band radio frequency" and dispatching reporters to the scene, who "obtained the name of the alleged assailant simply by asking various witnesses, the police and an assistant prosecuting attorney." *Smith v. Daily Mail Publ'g Co.*, 443 U.S. 97, 99 (1979). Both newspapers published the name of the alleged gunman and were indicted under a West Virginia statute that made it a crime to publish the name of a minor accused in a juvenile proceeding. *Id.* at 100. Noting that the newspapers were well within their rights to rely "upon routine newspaper reporting techniques to ascertain the identity of the alleged assailant," this Court held that "if a newspaper lawfully obtains truthful information about a matter of public significance then state officials may not constitutionally punish publication of the information, absent a need to further a state interest of the highest order." *Id.* at 103-04 (holding that the state's interest in the confidentiality of juvenile proceedings was not sufficient to overcome First Amendment protections).

Perhaps most instructive is *Florida Star v. B.J.F.*, 491 U.S. 524 (1989). In that case, the Duval County Sheriff's Department inadvertently placed an incident report containing the full name of a rape victim in the precinct pressroom, where it was accessed by a newspaper reporter. *Id.* at 526. The police did "not restrict access to either the pressroom or to the reports

made available therein" (*id*. at 527), but the pressroom did display "signs making it clear that the names of rape victims were not matters of public record, and were not to be published." *Id*. at 546 (White, J., dissenting). The newspaper ultimately reported the rape victim's name and was sued successfully by the victim under a Florida statute that prohibited disclosure. *Id*. at 529. In reversing the jury's verdict, this Court observed "that it is a limited set of cases indeed where, despite the accessibility to the public to certain information, a meaningful public interest is served by restricting its further release by other entities, like the press." *Id*. at 535.

The Court also noted that the police department's failure to prevent access to the victim's name did not "make the newspaper's ensuing receipt of this information unlawful." *Id*. at 536. To the contrary, the fact that the reporter relied on "routine newspaper reporting techniqu[es]" after the government made the information publicly available "without qualification, can only convey to recipients that the government considered dissemination lawful." *Id*. at 538-39. In other words, once the reporter was authorized to access the rape victim's name in the press room, the government could not punish the newspaper for publishing it – even though the police should have kept it secret and even though there were signs in the pressroom clearly telling the reporter that the information was confidential. *See also Bartnicki v. Vopper*, 532 U.S. 514, 535 (2001) (holding that reporter who lawfully received and published contents of illegally wiretapped phone call was not liable and reaffirming earlier holdings that "state action to punish the publication of truthful information seldom

can satisfy constitutional standards") (quoting *Daily Mail*, 443 U.S. at 102).

These decisions all focus on the rights of news organizations to publish information, but the necessary corollary to the right to publish news is the right to gather information in the first place. And each of these cases tacitly affirms the core principle that the government cannot punish journalists for using "routine newspaper reporting techniques" to obtain truthful information involving a matter of public interest that has already been made publicly available, even if it was disclosed by accident or unlawfully. *Daily Mail*, 443 U.S. at 103-04.

This First Amendment protection applies even when a journalist technically exceeds the terms under which he or she was allowed to access the information. *See Florida Star*, 491 U.S. at 534-35. For precisely the same reasons, the First Amendment should protect data journalists – including reporters working for *The Markup* – who employ equally innocuous newsgathering methods to gather truthful publicly available data involving matters of public interest from websites, even when the acquisition of that information exceeds the terms under which they were allowed to access it.[6]

---

[6] This is not to say that website operators have no remedies against users of their websites that violate the terms of service. Indeed, causes of action for breach of contract and copyright infringement may remain available in appropriate cases. *HiQLabs, Inc. v. LinkedIn Corp.*, 938 F.3d 985, 1004 (9th Cir. 2019). But what is not permitted is an "interpretation of the CFAA's 'without authorization' provision" that would "turn a criminal hacking statute into a "sweeping Internet-policing

An expansive reading of the CFAA to criminalize journalists who violate a website's terms and conditions while in the process of gathering publicly accessible information would also run afoul of this Court's longstanding commitment to preserve "the vast democratic fora of the Internet." *Reno*, 521 U.S. at 868. "[C]yberspace … in general … and social media in particular" are "essential venues for public gatherings to celebrate some views, to protest others, *or simply to learn and inquire*." *Packingham*, 137 S. Ct. at 1735 (emphasis added). Given the fact that this Court has unequivocally extended the protections of the First Amendment into the digital realm, it simply cannot be the case that the First Amendment protections afforded to journalists gathering public information evaporate the moment they go online, if a website operator should so choose.

More fundamentally, the Eleventh Circuit's broad reading of the CFAA undermines this Court's efforts to preserve the openness of the Internet as a source of information. Indeed, this Court has urged "extreme caution" before adopting any rules that would inhibit "access to the vast networks" of the Internet because "we cannot appreciate yet its full dimensions and vast potential to alter how we think, express ourselves and define who we want to be. The forces and directions of the Internet are so new, so protean, and so far reaching that courts must be conscious that what they say might be obsolete tomorrow." *Packingham*, 137 S. Ct. at 1736. But the Eleventh Circuit's broad interpretation of the CFAA would effectively deprive

---

mandate" that would give website operators broad powers to criminalize routine journalistic activities. *Id*. at 1000-01.

journalists of the ability to engage in routine newsgathering online. Such a reading would also concentrate more power in the government and online platforms by giving them an unprecedented veto power – criminal sanctions – to punish anyone attempting to investigate their activities. As society increasingly migrates online, it is crucial to keep the Internet open not just to the companies that profit from it but also to the news organizations that seek to tell the public how it works. Banning *The Markup* and other journalists from collecting and analyzing publicly accessible information would make it impossible to observe activity on the Internet, which in turn would make it impossible for reporters to serve their vital oversight function.

In short, a broad reading of the CFAA that criminalizes basic reporting practices used by data journalists to obtain publicly accessible information would clearly violate the First Amendment and must be rejected.

B. The CFAA Should Be Interpreted to Avoid Violating the First Amendment

Ultimately, this Court can avoid the confrontation between the CFAA and the First Amendment by interpreting it, as petitioner urges, not to criminalize any access by authorized users even if they exceed terms of service. "[I]t is a well-established principle governing the prudent exercise of this Court's jurisdiction that normally the Court will not decide a constitutional question if there is some other ground upon which to dispose of the case." *Bond v. United States*, 572 U.S. 844, 855 (2014). Here, the

constitutional question can be avoided simply by giving the CFAA its most rational reading.

That was the course taken by the District Court for the District of Columbia in *Sandvig v. Barr,* 2020 WL 1494065, at *1. In that case, academic researchers tested job websites by creating "profiles for fictitious job seekers" in order to determine whether job rankings are "influenced by race, gender, age, or other attributes." *Id*. The researchers knew that their fake profiles would violate the terms and conditions of the hiring websites and brought a pre-enforcement challenge, arguing that the CFAA violated their First Amendment rights. Judge Bates recognized that an interpretation of the CFAA that criminalized academic research would raise First Amendment concerns, but sidestepped this constitutional conflict by interpreting the CFAA to "not actually criminalize plaintiff's proposed conduct – namely violating website's terms of service." *Id*. at *7. The court proceeded to interpret the statue and rejected the Eleventh Circuit's holding that a person violates the CFAA by obtaining "information for a non-business reason" in excess of their authorized access. *Id*. at *12 (quoting *United States v. Rodriguez*, 628 F.3d 1258, 1263 (11th Cir. 2010). Judge Bates noted that "consumer websites' terms of service do not provide adequate notice for purposes of criminal liability" because these "protean contractual agreements are often long, dense, and subject to change." *Id*. at *10, *13. The court ultimately held that "violating public websites' terms of service … does not constitute a CFAA violation under the 'exceeds authorized access' provision." *Id*. at *13.

* * * * *

In sum, the CFAA cannot be interpreted to criminalize routine newsgathering activities without violating the First Amendment. But this constitutional clash can be avoided simply by giving the CFAA its rational interpretation, which takes into account the ways in which information is gathered online. The expansive interpretation of the CFAA that the government advances here would criminalize entirely lawful methods of gathering information on the Internet and have a devastating chilling effect on data journalism. Adopting petitioner's interpretation of the statute – which is consistent with the plain meaning of the text, reflects the reality of how newsgathering works online, and does not violate the constitution – avoids this improper result.

CONCLUSION

For the foregoing reasons, the decision of the Court of Appeals should be reversed.

Respectfully submitted,

Katherine M. Bolger*
John M. Browning
DAVIS WRIGHT TREMAINE LLP
1251 Avenue of the Americas
New York, NY 10020
(212) 489-8230
katebolger@dwt.com
jackbrowning@dwt.com

David M. Gossett
DAVIS WRIGHT TREMAINE LLP
1919 Pennsylvania Ave. NW
Washington, DC 20006
(202) 973-4200
davidgossett@dwt.com

Nabiha Syed
THE MARKUP
900 Broadway, Suite 202
New York, NY 10159
(347)894-3746
nabiha@themarkup.org

* Counsel of Record

Counsel for Amicus Curiae

July 8, 2020