Tow Center for Digital Journalism
A Tow/Knight Report

DATA JOURNALISM AND THE LAW

D. VICTORIA BARANETSKY

Columbia Journalism School

Funded by the Tow Foundation
and the John S. and James L. Knight Foundation

Acknowledgments

All of my gratitude goes to those journalists and lawyers who spoke to me as I formulated this report, and helped train me along the way. I'm also immensely grateful to Priyanjana Bengani, Samuel Thielman, Susan McGregor, Emily Bell, and the staff of the Tow Center for their tireless editorial support and endless patience. And to my family and friends, especially Amy Poppers and Emily Haydock, for their love and support in the revolutionary act of self-care.

August 2018

About the Author

D. Victoria Baranetsky is general counsel at the Center for Investigative Reporting, a fellow at the Tow Center for Digital Journalism, and former affiliate at Harvard's Berkman Klein Center for Internet and Society.

Contents

Executive Summary

While data journalism has been the hot topic of various recent reports, no studies have discussed how the changes in journalistic storytelling may create new legal considerations for journalists. This report aims to help journalists, lawyers, and academics understand the changes taking place in media law as a result of both the growing volume of data in our information economy, and the the seismic shifts occurring within journalism and technology. By examining developments in newsgathering law, the Freedom of Information Act, and laws involved in leak investigations, this research underscores worrisome shifts in the law, as well as gray areas where reform would strengthen the rights of a free press and journalists.

In its first part, the report looks at emerging concerns over data journalism projects that could trigger the Computer Fraud and Abuse Act (CFAA), 18 U.S.C. §1030(a)(4)[i] for scraping, a data collection technique that usually relies on automation—through bots, crawlers, or applications—to extract data from a website. As data collection becomes increasingly important for investigative journalists in particular, legal experts worry about civil and criminal penalties that exist under the statute—which has been described by some First Amendment advocates as unconstitutionally vague. In reviewing the history and case law of the CFAA in relationship to journalism, the research offers practical tips and various legal considerations on the issue.

Next, the report discusses troubling trends arising under the Freedom of Information Act (FOIA) in the digital age, as the amount of government information held in databases and government logs grows, and the need for transparency is crucial. Lastly, it reviews data's impact on laws affecting

i. This report uses legal notation for brevity's sake, and to allow the reader to locate and examine the laws for herself, should she so choose. In this case, the CFAA is located in Title 18 of the US federal law, or "U.S.C." for US Code, section 1030, paragraph a, subparagraph 4.

whistleblowers. In the past decade, we've seen more leak prosecutions in the United States than all those combined in the country's history. This, of course, occurs at a time when there is more information than ever before for whistleblowers to share.

Key findings

- While no journalists to date have been sued or prosecuted under the Computer Fraud and Abuse Act, there's evidence that stories have been hindered or held from publication for the threat of penalty. Under the statute, a person may be penalized for "unauthorized access" to data on a company's website by scraping data through bots, crawlers, or applications. While journalists have developed techniques and tools to sidestep potential liability, including piecemeal data extraction that goes unnoticed or crowdsourcing the public's help, the CFAA presents real obstacles to reporting a variety of important stories in the public interest.
- In the past decade, as the volume of US government-controlled data has increased, government agencies have experienced a swell in the number of FOIA requests. At the same time, reporters often see delays in processing requests, insufficient searches conducted by the agencies, government data equated with proprietary information leading to denial of access, and developing case law that prohibits access to government databases. In a number of recent instances, courts have upheld determinations by agencies that searching a government database amounts to producing a new document—which is prohibited under FOIA.
- While no journalist has been convicted under the Espionage Act, the statute includes provisions that could potentially be levied against journalists. Even more stifling than the policies and laws used to intimidate and silence whistleblowers in the digital age, though, is the degree to which government authorities seem preoccupied with journalists specifically. Now, possessing information obtained from confidential sources—a basic tenet of First Amendment doctrine—is potentially a prosecutable offense.
- In many ways, journalists' access to critical information is being restricted, either by the passive or explicit threat of criminal penalties,

de-prioritization in favor of corporate secrecy, or an inadequate legal understanding of technological advances.

While none of these shifts are totalizing—or irreversible—together they indicate a new direction and acceleration in our information economy that may have consequences for journalists. As more information is created, there is growing need for reporters to discern the importance of voluminous data dumps—and to uncover stories hidden in their details. Unlike anti-secrecy sites such as WikiLeaks, journalists review, analyze, and edit information to help citizens navigate the evolving information landscape. But the current state of journalism, which often sees the press intimidated by public officials and public figures, is presented with unique challenges in an oversaturated information economy where there are fewer resources and protections in place for journalists to discern the truth. It is time that we consider these subtle shifts as a hazard to the Fourth Estate itself.

Introduction

In 1961, legal scholar Alexander Meiklejohn famously wrote that the rationale for the First Amendment depended on citizens' ability to receive and use information relevant to democratic self-governance.[1] The crux of his statement was this: knowledge is power. Fifteen years later, scholar Thomas Emerson would rely on Meiklejohn's work to famously highlight the "vital importance in a democratic society of the right to know."[2] In his article, he explained how James Madison, the author of the First Amendment, asserted that "[a] popular government, without popular information or the means of acquiring it, is but a prologue to a farce or a tragedy; or perhaps both."[3] From there, Emerson continued, "A people who mean to be their own governors, must arm themselves with the power that knowledge gives."[4]

In view of this, asserting access to information seems paramount to self-governance. Every day reporters try to fulfill this duty through various mechanisms. But in our current environment, they are often competing with an unparalleled glut of information that readers absorb from the moment they wake up to the moment they power down their devices at night. One study by Northeastern University estimated that the size of the "digital universe" of data was 4.4 zettabytes in 2013—and is scheduled to jump to 44 zettabytes by 2020.[5] [ii] According to a *Forbes* magazine piece in 2015, "More data ha[d] been created in the past two years than in the entire previous history of the human race."[6] [iii]

This voluminous amount of information has led to massive shifts in the news industry for nearly a decade. Since about 2008, the explosion of data

ii. Note that this statistic does not underscore how amazingly large a zettabyte is.

iii. These changes are due to various causes, including greater access to technology, an increase in data-tracking technology, as well as vertical-to-horizontal shifts in various information industries. In addition to the increase in the quantity of data, the ability to access information has never been easier. By 2014, 90 percent of adult Americans owned a mobile phone. In fact, according to one 2014 study by the International Business Times, cellphones now outnumber people on Earth (Zachary Davies Boren, "Active Mobile Phones Outnumber Humans For the First Time," *International Business Times*, October 8, 2014, https://www.ibtimes.co.uk/there-are-more-gadgets-there-are-people-world-1468947).

journalism—defined as journalism that heightens the role numerical information plays in storytelling—is now a driving force in newsrooms around the country. Journalists are quickly learning how to obtain troves of data through electronic leaks, drones, and cutting-edge computer programs that sometimes require little more than the click of a button to access information. In other instances, journalists confronted with processing large swaths of information must employ complicated algorithmic and programming skills. Many larger news organizations have even built internal digital programs and tools to sort through these data swells and leaks—as was done with the Panama Papers.[iv]

As the Global Investigative Journalism Network reported in 2015, "After nearly 50 years of journalists using data, it is clear that data is not only a routine part of journalism, but also a driving force for stories."[7] A recent report by Google stated that 42 percent of reporters use data to tell stories regularly, and 51 percent of all news organizations in the United States and Europe now have a dedicated data journalist—a figure that rises to 60 percent for digital-only publications.[8]

While data has become integral to reporting the news, the quantity of data at large and the celerity with which it can spread have led to many journalistic concerns over protecting sources, the accuracy of published information, the inability to provide meaningful redactions, and journalistic

iv. Leaks will be used synonymously with whistleblowing in this report. Investigative reporter Seymour Hersh, among many others, argues that "leak" is a pejorative term, and "tip" or "whistleblowing" are better words. This report does not suggest that leak is a pejorative word. Whistleblowing and leaking, in this context, refer to "the act of providing confidential information to the public in a surreptitious way and without official authorization." Information, here, includes national security information relating to defense against transnational terrorism, which is protected by the three classification levels currently in place: Top Secret, Secret, and Confidential. This is different from planting, which is an authorized disclosure that plays a critical role in policymaking and communications, allowing the government to disseminate information, frame narrative, and gauge the reactions of both the people and Congress. Historically, a government employee or contractor would get in touch with members of the press and provide them with a tip, which journalists had to investigate and confirm prior to publishing. It is also worth noting that whistleblowers aren't unique to government employees. They also come out of private corporations and can have an equally big impact. For example, the Panama Papers consisted of 11.5 million leaked documents from Mossack Fonseca, a law firm in Panama that helped people open offshore accounts. However, different legal considerations apply to government whistleblowers than they do to private whistleblowers, such as the Espionage Act. This report focuses on the former.

liability. Meanwhile, corporations and governments, which hold much of the information that journalists are responsible for reporting on, are beginning to exercise stricter controls over their data—in many cases by asserting that it is proprietary information. Federal and local governments are also guarding their information by expanding exemptions under the Freedom of Information Act, and increasingly asserting privacy exemptions on behalf of individuals and corporations alike. Similarly, private companies are exerting stronger trade secret exemptions, with governments upholding those claims in even the most dubious circumstances.

While several reports have covered in depth what data journalism is and how to implement it at various institutions, no recent studies have discussed how this change in storytelling is subject to the legal landscape journalists must work within. Traditionally, media law concerns around newsgathering have been limited to questions about trespass, recording laws, access to illegally obtained material, and the potential for prosecution under the Espionage Act. These leading doctrines surely still apply, but there are shifts in case law and arguments that have yet to be fully explored in media law casebooks and conversations. This report is an attempt to tease out some of those new conversations and explore how various case law is being affected by our data addiction.

New Newsgathering Liabilities: The Computer Fraud and Abuse Act

Perhaps the most worrisome legal concern related to the increase in data collection is possible liability under the Computer Fraud and Abuse Act, 18 U.S.C. §1030(a)(4) for data "scraping," a data collection technique that usually relies on automation—through bots, crawlers, or applications—to extract data from a website. For investigative journalists, in particular, scraping has become "one of the most powerful techniques for data-savvy journalists who want to get to the story first, or find exclusives that no one else has spotted."[9] While gathering data from public resources and directly requesting data from corporations rarely puts journalists at risk of legal liability, trying to scrape corporate data can incur severe civil and even criminal penalties under the statute. Several reporters are on record describing stories that have been blocked because of legal concerns associated with the CFAA. The CFAA, then, presents real obstacles to reporting out stories important to the public's interest. Thus, as data journalism increasingly becomes "a good way of getting to the truth of things . . . in this post-truth era," as one data journalist told Google,[10] the need for further clarity around the CFAA increases.

Data scraping

"[W]eb scraping has become an increasingly useful tool for reporters who know how to code," wrote Nael Shiab, a reporter for the Global Investigative Journalism Network,"[11] in a 2015 article that cited several projects where reporters working a variety of beats wrote Python scripts to collect data.[12]

But even among those who don't specialize in coding,[13] scraping is a popular journalistic technique—and has been for years. A Knight Lab article in 2014 shared a list of five new web-scraping tools useful for journalists,[14] and in 2016 Sophie Chou of MIT Media Lab reported on its rise across the journalism industry.[15] For specialists, too, the field is growing. In 2016, Ricardo Brom, an Argentinian data journalist, alongside David Eads of NPR, Amanda Hickman of BuzzFeed, and Martin Burch of *The Wall Street Journal*, presented a panel on data scraping at NICAR16, a computer-assisted reporting conference run by Investigative Reporters and Editors.

Indeed, many essential stories from the past several years were born from data-scraping projects. In 2017, ProPublica journalists unveiled a project investigating Amazon's pricing algorithm. Using a "software program that simulated a non-Prime Amazon member," the team looked "at 250 frequently purchased products over several weeks," while also scraping product listing pages. What they found was that Amazon was prioritizing its own products and those from vendors that pay for its services, while often hiding the best deals from its customers.[16] The story was one of many highlighting increasing concerns around the use of algorithms in industry.

Some news organizations use scraping daily to keep their ongoing projects up to date. For instance, Reveal from the Center for Investigative Reporting (disclosure: where the author works) uses scraping to keep its data journalism project "The Lost & The Found" contemporaneous. Every night, Reveal uses Python to scrape a federal database of thousands of missing persons and unidentified bodies to see which cases are new, which have changed, and which have been removed or solved. Similarly for a story about the housing crash, Reveal wrote a scraper to get data on land con-

tract sales in Detroit after the government refused to disclose it unless an hourly fee was paid. The process took weeks.

The Atlanta Journal-Constitution has gone even further, developing multiple scrapers for a national investigation called "Doctors & Sex Abuse."[17] When its reporters' public record requests to medical boards and regulatory agencies in every state yielded very little return, the newsroom's data journalism team wrote multiple scripts that crawled the regulators' websites to obtain the board orders—which were hidden not by classification or court order, but by obscurity. The scrapers retrieved more than 100,000 disciplinary documents used in the yearlong investigation. The project was a finalist for the Pulitzer Prize for national reporting in 2017.

A recent blockbuster project by Julia Angwin at ProPublica relied on scraped data to report a piece that used the publication's Facebook Political Ad Collector extension for Google's Chrome browser. On its website, ProPublica says of this process: "The extension, which we call the Political Ad Collector, is a small piece of software that users can add to their web browsers. When a user logs into Facebook, the extension will collect the ads displayed on the user's News Feed and guess which ones are political based on an algorithm built by ProPublica."[18]

Since scraping can carry the threat of legal liability under the CFAA, various news organizations have built tools that effectively work around the potential legal implications. For instance, in this case, rather than scrape data from Facebook directly, ProPublica's Political Ad Collector explicitly informs users about the scope of data collected, and the users who opt in allow their computers to perform the data collection on ProPublica's behalf, minimizing (if not eliminating) ProPublica's liability under CFAA.[19]

The extension then collects political ads from users' Facebook pages and sends it to ProPublica, which helps "shine a light on political advertising on Facebook."[20] In keeping with the open nature of the project, ProPublica open-sources all the related data it collects.[21] Other, simpler techniques to avoid liability involve scraping sites or parts of the internet that are unlikely to trigger litigation. A recent Medium post by Lauren Glass, for example, describes how to scrape a Google search result.[22] Similarly, in its housing crisis project, Reveal first entered credit card information into the government database by hand before employing the scraper to ensure that

it followed government protocol but used the scraper to gather information efficiently.

Academics also scrape. Researchers from the Harvard Business School published an article focused on racial discrimination on Airbnb, gathering data through "inquiries to Airbnb hosts using web browser automation tools" that was all collected using scrapers.[23] The study found that people with distinctly African-American names were 16 percent less likely to have their rental requests accepted compared to identical guests with distinctly white names.[24]

While scraping is becoming a regular tool for people from many fields, governments and corporations oftentimes don't welcome it. "Web scraping activity often sets the interests of a website owner against those who want to harvest the data," noted one guide around best web scraping practices for publishers.[25] "Owners usually want to control, profit from, or leverage the data they have on their website," and journalists' excavation of this data threatens these resources.[26] Currently, governments and private companies store troves of data on people, their preferences, and their networks—personal and professional—that allow them to influence citizens and product users. It's common practice for companies to use algorithms that determine the content users see, while simultaneously allowing advertisers to micro-target very specific classes of people without users ever knowing about it. Recent research has also unveiled how algorithms can hold important sway over users through implicit bias.[27] [28]

Large companies and technology platforms hold a lot of control over the public through gathering its user data, and remain largely unregulated in the United States—often under the claim that the industry "self-regulates" and that the data companies collect should be considered a trade secret. Moreover, companies rarely make this data available to universities, researchers, or think tanks, limiting the ability of outside institutions to hold them accountable for misuse or educate users about the way they may be receiving influenced information. Since the Cambridge Analytica scandal and recent congressional hearings around Facebook's data breach, there may be change in this realm. More specifically, it is increasingly likely that companies will try to self-regulate and make various information even more difficult to access and review, triggering an ever greater need for reporters

and academics to continue their work advocating for more transparency and access to information as algorithms and data sets become distinctly determinative factors in our lives.

And while researching through data and scraping techniques have become the most efficient way to keep a check on publicly influential institutions, the specter of legal liability is palpably present. Several independent journalists and newsrooms accessing data not available through companies' APIs have indeed declined to publish stories for fear of liability under the CFAA.

Traditional and new legal doctrines

Before understanding more about the CFAA, it is important to note that private industry has and continues to use a variety of legal strategies to deter journalists and other researchers from collecting data. Traditional doctrines including trespass, copyright, fraud, and breach of contract are among the first recourses for threatened companies. But since these legal claims are not as tenable in the online space—or as caustic—they are therefore not as deterrent.

Trespass, for instance, has been asserted in cases of web crawling. However, the idea of trespass has ostensibly become a "norm of the Web."[29] [v] Thus, courts have been reluctant to compare a robot that "enters" a website to an unauthorized intruder who enters a "brick and mortar" location.[vi] As noted by the Electronic Frontier Foundation in an amicus brief, "Open access is a hallmark of today's Internet, and one of the main reasons the Internet has become our 'modern public square.' " Even more consequential is that trespass generally requires some harm to a property in order to have damages. This is a key difference between digital and physical unauthorized access. While stepping onto someone's property and altering their belongings may conceivably cause harm, in the case of web crawling, bots rarely cause harm to the business by crawling for data (with the exception of sometimes slowing traffic to the site).[vii]

v. See Ebay, Inc. v. Bidders Edge, Inc, 100 F. Supp, 2d 1058 (N.D. Cal 2000).
vi. Ibid.
vii. See Ebay, Inc. v. Bidder's Edge, Inc, 100 F. Supp, 2d 1058 (N.D. Cal 2000) (stating "[i]f BE's activity is allowed to continue unchecked, it would encourage other auction

Similarly, copyright claims are rarely viable. For instance, in *Ticketmaster Corp. v. Tickets.com, Inc.*, a California federal district court doubted the potential misappropriation claims of data under the Copyright Act.[viii] Ticketmaster alleged that Tickets.com wrongfully used software to scrape information from Ticketmaster's website. The court found that because copyright law protects neither facts nor compilations of facts,[ix] Ticketmaster could not claim copyright infringement. The court emphasized the danger of protecting factual data under copyright claims, since it is normally open to public use. Moreover, given that a great deal of the content posted on websites (such as social media sites and online retailers) is created by users, platforms are rarely able to assert copyright claims on their own behalf—because copyright is held by the creator of the content and various sites are merely custodians of that information, not the creators.

Still, some tactics of redress against scraping work better than others. Corporations have successfully used contract claims as a defense against web scraping, typically by treating the site's terms of service as a contract and prohibiting the act therein. In *Register.com, Inc. v. Verio, Inc.*, the defendant developed a scraping program to harvest publicly accessible WHOIS domain registrant email addresses; the plaintiff asserted that its terms of use had been violated.[x] That claim was successful, but in other cases terms of use that purport to restrict the use of data, but which do not require an "individualized step to limit access are unenforceable in many jurisdictions."[xi] Some have also argued that scraping does not trigger a

---

aggregators to engage in similar recursive searching of the eBay system such that eBay would suffer irreparable harm from reduced system performance, system unavailability, or data losses"); but see Oyster Software, Inc. v. Forms Processing, Inc., No. C-00-0724, 2001 WL 1736382, at *11-*13 (N.D. Cal. Dec. 6, 2001) (declining to dismiss trespass claim).

viii. See Ticketmaster Corp. v. Tickets.com, Inc., 2000 U.S. Dist. LEXIS 4553 (C.D. Cal. Mar. 27, 2000), aff'd 248 F.3d 1173 (9th Cir. 2001); Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

ix. See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 344, 348 (1991).

x. 356 F.3d 393, 395-96 (2d Cir. 2004)

xi. See Nicholas A. Wolfe, "Hacking the Anti-Hacking Statute," N.Western J. L. and Tech., 13, 2015 (citing Ticketmaster Corp. v. Tickets.com, Inc., 2000 U.S. Dist. LEXIS 4553 (C.D. Cal. Mar. 27, 2000), aff'd 248 F.3d 1173 (9th Cir. 2001); Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991)

contractual arrangement precisely because a bot can neither understand nor agree to the terms of service.

Facing the overall inconsistency of these protections, corporations have taken steps to assert claims under the Computer Fraud and Abuse Act, often called a blunt and outdated "anti-hacking" statute. The CFAA is a law passed in 1986 that imposes criminal as well as civil liability on "whoever . . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer"—which includes all computers connected to the internet.[xii] The statute, inspired by the 1980s film *War Games*, was intended to penalize hackers for breaking and entering into another person's computer.[xiii] The first edition of the statute narrowly criminalized breaking into a computer to obtain national security secrets, consumer credit, or financial information, or to "hack" into government computers.[xiv] In contrast, the current version leaves open criminality for a much broader set of actions that fall under any kind of "unauthorized access."

While no journalist has ever been convicted of either civil or criminal charges under the CFAA, the act has been used to prosecute journalistic sources who have uncovered information in the public interest. In March 2013, security researcher Andrew Auernheimer (also known as "Weev") was sentenced to 41 months in prison for revealing an AT&T security hole by obtaining the personal data of more than 100,000 iPad users from the company's publicly accessible website—not by "hacking," but by deducing how the site presented information—and then sharing the information with Gawker.[30] [31] [32] The same year, journalists at Scripps News who found sensitive phone customer records of TerraCom and YourTel America users through a Google search and made the weakness public by reporting on the issue were dubbed "hackers" by the company and threatened with a CFAA violation in a letter.[33]

xii. 18 U.S.C. §§1030(a)(2)(C), 1030(e)(2)(B)

xiii. United States v. Valle, 807 F.3d 508, 525 (2d Cir. 2015) (citing H.R. Rep. No. 98894, 1984 U.S.C.C.A.N. 3689, 369192, 369597 (1984); S. Rep. No. 99432, 1986 U.S.C.C.A.N. 2479, 2480 (1986)) (stating the statute was meant "to address 'computer crime,' which was then principally understood as 'hacking' or trespassing into computer systems or data.")

xiv. See Comprehensive Crime Control Act of 1984, Pub. L. No. 98-473, 98 Stat. 2190, codified at 18 U.S.C. §1030(a)(1)(3).

Civil liberties advocates, such as the Electronic Frontier Foundation, have long argued that confusion around the application of this law could have a chilling effect on the research of journalists and academics. On similar grounds, the ACLU filed suit last year in federal court in the United States' District of Columbia, challenging the constitutionality of the CFAA on behalf of a variety of academics, as well as First Look Media, the parent company of investigative news outlet The Intercept. In its complaint, the ACLU argued that the law criminalizes research that might break a site's terms of service by simply using its public-facing data for research, and that this type of penalty chills reporting and academic study because it threatens any number of uses of public data without making clear which ones the law considers legitimate.

Language of the CFAA

The essential unknown under the CFAA is what constitutes accessing a computer "without authorization." Currently, the federal circuit courts are split on this issue. Four appeals courts (including the First, Fifth, Seventh, and Ninth Circuit) have broadly interpreted the statute to include violations of a corporation's terms of use policies. This broad definition frames "authorization in terms of the computer owner's intentions, expectations, and contractual or agency relationships," meaning that a broad spectrum of behavior may be subject to criminal liability as long as a company lists the infraction in its terms of service. Under this interpretation, a company need only prohibit "scraping" or "data collection" in its terms of service to trigger CFAA protections.

However, some circuits have tried to limit this broad liability without explicit direction. For instance, in a Ninth Circuit case, *Power Ventures*, which the Supreme Court recently denied for cert, the Court of Appeals clarified that a violation of the terms of use of a website plus some other additional factor—which has not been specified—could ostensibly establish liability under the CFAA. The "additional factor" could be as simple as refusing to comply with a cease and desist letter, which would demonstrate that the complainant had proactively revoked access and that the infringer was on notice. But as Judge Stephen Reinhardt warned in his dissenting

opinion in a prior Ninth Circuit opinion called *Nosal*, there still stands the overall concern that solely relying on a company's policy "would base criminal liability on system owners' access policies," which creates a myriad of problems, given that it would effectively allow tech companies to pass unconstitutionally vague criminal laws by fiat.[34] Other courts have also noted that the imprecise meaning of "without authorization . . . has proven to be elusive," making it an unconstitutional way to define criminality.[xv]

By contrast to the Ninth Circuit, a minority number of appeals courts, including the Second and Fourth Circuits, go even further, narrowly construing the CFAA as an anti-hacking statute that only penalizes access if it amounts to "breaking and entering" a computer—as was initially intended by Congress. In 2015, the Second Circuit held in *United States v. Valle* that a narrow interpretation of the CFAA is "consistent with the statute's principal purpose of addressing the problem of hacking, i.e., trespass into computer systems or data"[xvi] Similarly, in 2012 the Fourth Circuit in *WEC Carolina Energy Solutions LLC v. Miller* stated it could not "contravene Congress's intent by transforming a statute meant to target hackers into a vehicle for imputing liability to workers who access computers or information in bad faith, or who disregard a use policy."[xvii] Under these more narrow interpretations, a company "authorizes" use of data stored on its computers simply by providing a means of access. Legal scholars, like Orin Kerr, have analogized password authorization to the idea of a key to an apartment that can be shared with third parties.[35] Both passively allow outsiders access to the company, negating criminality.

Two of the most recent legal decisions involving web scraping suggest a favorable future for its allowance under the CFAA. Decided in the Northern District of California in late 2017, *hiQ Labs v. LinkedIn*[xviii] "signals a shift in the way courts may be viewing attempts to restrict data scraping, giving web scrapers some arrows in their legal quiver to get back against recent opinions condemning scraping."[36] The dispute arose from hiQ's business model: The company scrapes and analyzes information from LinkedIn and

xv. See EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 582 n.10 (1st Cir. 2001).
xvi. 807 F.3d 508, 526 (2d Cir. 2015)
xvii. 687 F.3d at 207
xviii. No. 17-CV-03301-EMC, 2017 WL 3473663 (N.D. Cal. Aug. 14, 2017)

markets the resulting insights to clients interested in information about their own employees, such as whether workers are looking for a new job. In May 2017, LinkedIn sent hiQ a cease and desist letter, demanding the company immediately stop collecting data from its public profiles. HiQ then filed a declaratory action asking the court to state that it had not violated and would not violate any laws with its scraping.

The court decided in favor of hiQ, reasoning that outlawing such access would effectively make it illegal to view public information. This type of penalty could "lead to perverse consequences, effectively allowing websites to criminalize access to [a private company's] public site on the basis of discrimination or anti-competitive intent." The court further explained that if "merely viewing a website in contravention of a unilateral directive from a private entity would be a crime," the law would become, in effect, "the digital equivalence of Medusa."

The opinion is a strong aberration from other, more meek Ninth Circuit case law, which in applying this "unauthorized access" language had found that parties accessing computers after an express revocation of permission (such as a cease and desist letter) had violated the CFAA. In the *Power Ventures* case, a three-judge panel said the company violated the CFAA because it knew it wasn't authorized to access Facebook's computers after receiving the cease and desist letter. However, hiQ differs factually in at least one significant way: the LinkedIn data isn't password-protected, unlike the information accessed by Power Ventures. Regardless, the impending influence of hiQ seems palpable.[xix]

xix. This position was recently upheld by the Ninth Circuit in a case dealing with California's computer crime law. In *Oracle v. Rimini*, the federal court of appeals held that a company could not be held criminally liable for accessing Oracle's website in a manner that violated the website's terms of use. This particular case history is noteworthy because it dates back to a 2012 decision that held a violation of Oracle's terms of use is not a crime under the CFAA. Hoping to circumvent this ruling, Oracle then turned to the state computer crime statutes of California and Nevada—to enforce the company's terms of use. In this case, Rimini violated Oracle's provisions by using automated scripts to download data from the website in bulk, instead of downloading each file individually. Oracle sent Rimini a cease and desist letter and while Rimini did stop collection temporarily, it soon resumed using the scripts. Although Oracle sued Rimini and won on the district court level under both the California and Nevada computer crime statutes, on appeal, the Ninth Circuit reversed the lower court's rulings, stating that when "a computer user has permission and authorization to access and use the computer or data" it does not become a crime simply because the user then accesses or uses the information

Indeed, a similar case decided this year also appeared to follow suit. In 2016 the ACLU filed its lawsuit, *Sandvig v. Sessions*, in federal court in Washington, D.C., challenging the constitutionality of the CFAA—based in part on the fact that it violated plaintiffs' First Amendment rights.[xx] There, the ACLU represented scholars and a media organization The Intercept, who were scraping information to identify and understand how racism is implemented in algorithms. The action charged that 1032(A)(2)(c) of the CFAA, which allows terms of service to determine criminal violations under the CFAA, hinders important research, and is not only vague but also violates the First Amendment. In its complaint, the ACLU argued that criminalizing a violation of a term of service has a chilling effect on researchers and journalists by thwarting them from collecting data. The complaint states: "Refraining from conducting their research, testing, or investigations constitutes self-censorship and a loss of Plaintiffs' First Amendment rights." Prior to the decision being issued, Esha Bhandari, an attorney for the case, explained that "not being able to record information that is otherwise available to the public" is a clear First Amendment concern.

In the court's decision, which it issued March 30, 2018, during the writing of this paper, the district court denied in part and granted in part the government's motion to dismiss, allowing the case to proceed and address the merits of one of the ACLU's First Amendment claims. United States District Judge John Bates wrote:

> First, scraping plausibly falls within the ambit of the First Amendment. "The First Amendment goes beyond protection of the press and the self-expression of individuals to prohibit government from limiting the stock of information from which members of the public may draw."[xxi] The Supreme Court has made a number of recent statements that give full First Amendment application to the gathering and creation of information. Additionally, six courts of appeals have found that individuals have a First Amendment right to record at least some matters of public interest in order to preserve and disseminate ideas.

While this language does not directly address the question of the First

in a manner the website owner does not like. This finding likely signaled the positive decision in the Ninth Circuit opinion in *hiQ v. LinkedIn*.

xx. Sandvig v. Sessions, No. 16-1368 (D.D.C. Mar. 30, 2018)

xxi. First Nat. Bank of Boston v. Bellotti, 435 U.S. 765, 783 (1978)

Amendment, it opens the door to a First Amendment challenge to the CFAA, as it is the closest a court has come to saying that research done through scraping may be protected free speech, and therefore the CFAA is unconstitutional.

Moreover, while the court's opinion did not cite the press clause but instead the First Amendment "right to record," the court seemed to acknowledge the important impact the CFAA had on the news media by citing a foundational press case: Bartnicki v. Vopper, 532 U.S. 514, 527 (2001). In that case, a radio station that obtained an illegally taped conversation was not held liable for broadcasting the tape. The 2001 decision came to stand for the proposition that information illegally obtained by a source does not legally implicate a journalist who chooses to share that information—as long as the journalist had no hand in obtaining the data and the publication of the information was in the public interest. This principle is incredibly important for journalists and the news industry.[xxii]

On page 16 of the court's opinion in *Sandvig*, Judge Bates cites *Bartnicki* after writing "plaintiffs contend that they have the right, and the desire, to publish the results of their research, and that some sites' ToS prohibit them from doing so without prior permission or else employ anti-disparagement clauses." More specifically, it continued:

> The Supreme Court has made very clear that the right to publish falls within the core of the First Amendment's protections. See, e.g., Bartnicki v. Vopper, 532 U.S. 514, 527 (2001) ("As a general matter, 'state action to punish the publication of truthful information seldom can satisfy constitutional standards.' " (quoting Smith v. Daily Mail Publishing Co., 443 U.S. 97, 102 (1979)). Applying criminal sanctions for publishing original material that uses publicly available information, or for making negative statements about a website, triggers First Amendment scrutiny.

Through this langauge the court appears to underscore that information scraped by a confidential source—even criminally—and given to an outlet may be published without sanctions. This would, at the very least, shield

xxii. For instance, it ostensibly protected *The New York Times*'s Suzanne Craig when she posted about her interest in President Trump's tax returns, even if illegally obtained. Similarly, it continually protects news organizations which may obtain scraped information from sources, even sources who have violated the CFAA.

publications that obtain data from sources or hackers like Weev or others who may illegally scrape sites.

Legal and practical tips

While these recent decisions support some cautious optimism among data journalists and their attorneys, the still uncertain state of the law leaves journalists who scrape data, especially in violation of a site's terms of service, at risk of both civil and criminal liabilities. For that reason, there are a few tips to consider. It is important to emphasize that these are in no way exhaustive and that if you are dealing with a legal issue involving data scraping, you should speak to an attorney about the situation.

1. Legal tactics

In many circumstances, journalists using data collected by a source—even if done illegally—may escape liability under precedent from *Bartnicki v. Vopper*. Even before the *Sandvig* decision, several newsrooms changed how they collected data to create protection for the outlet under *Bartnicki*. For instance, where the public or a source collects the data, a journalist's liability for data scraping can be circumvented, as information gathering by the public is less noticeable and less likely to be prosecuted under the statute.

However, given the *Sandvig* decision, there is now an even stronger claim that the journalists analyzing and researching are protected under their right to record, a long-established First Amendment right. This legal avenue is especially valuable to the press, given that the news media has often been found to have strong protections to record even where there were extant privacy concerns. For instance, in 1995 the Seventh Circuit found that while journalists were possibly liable for breaking onto private property of an opthamologist to cover a story under theories of trespass, privacy concerns were distinct and likely unhindered.[xxiii] In essence, where a reporter

xxiii. Desnick v. Am. Broad. Cos., 44 F.3d 1345, 1351-53 (7th Cir. 1995) (finding that "there is no journalists' privilege to trespass," however, noting that the journalists "entered offices that were open to anyone expressing a desire for ophthalmic services and videotaped physicians engaged in professional, not personal, communications with

goes into circumstances as a normal citizen might, and then attempts to somehow record that information, they may be protected despite privacy claims.

As law professor Margot Kaminski ably discussed in a recent law review article, privacy laws that constrain recording can be cast as speech-targeting laws.[xxiv] And while the CFAA is not a privacy law per se, the same principles seem to apply. While data may be on private property of various online sites, First Amendment concerns may outweigh any privacy or other considerations, especially where there is a public interest in publication of that material. This is especially important as privacy concerns about data increase in response to the Cambridge Analytica scandal, perhaps making the CFAA more alluring to some.

Regardless, it is important to note that without attorneys and more protection from the CFAA, journalists who do not have a larger newsroom supporting them with tools are often unable to tell certain stories. For instance, one unnamed journalist for a major news outlet built a tool to scrape information from a Silicon Valley company. The outlet told the journalist he could not collect the data himself for fear of violating the CFAA. While the journalist could have gotten the information using a tool like the kind employed by ProPublica, the Center for Investigative Reporting, and Gizmodo, he was unable to do so as a freelancer without access to a similar forum. Moreover, the tool would not have been able to obtain all the important data that the journalist would have wanted to collect. Stories like this one are not uncommon.

2. Practical ideas

To address CFAA concerns, journalists, lawyers, and ethicists have discussed what can be done to limit liability when scraping is necessary.[37]

First, reporters should always pay attention to the terms of service of

strangers (the testers themselves)"); see also Deteresa v. Am. Broad. Cos., 121 F.3d 460, 466 (9th Cir. 1997) (finding two years later the Ninth Circuit granted newsgatherers greater protection where privacy concerns arose, suggesting that the privacy laws were no longer inscrutable laws of general applicability).

xxiv. Justin Marceau and Alan K. Chen, "Free Speech and Democracy in the Video Age," 116 *Columbia Law Review*, 991, 997–98 (2016) (arguing that the First Amendment protects recording even when it occurs in a privately owned space)

the website they're accessing to understand whether the company prohibits scraping, even if its data is accessible. When reviewing the terms of use, there often is a section discussing bots, scripts, or other methods for collecting information from the site. Journalists should ensure that they are fully aware of what precisely the terms of use prohibit. If they cannot determine what those terms are saying, they should reach out to an attorney.

Second, if the terms of use prohibit data scraping, reporters should first seek alternate sources for the information, bits and pieces of which may be available on other sites or via public records requests. Unfortunately, companies have increasingly used trade secret exemptions to thwart disclosure of their information via public record requests. And, as *The Atlanta Journal-Constitution*'s "Doctors & Sex Abuse" case shows, public records requests do not always recover sufficient information. Data may have to be obtained from some third party available to you.

However, if the information exists only on a company's site, it can be a good idea to contact the company and see if they will simply turn over the information. John Muyskens, a reporter at *The Wall Street Journal*, told the Tow Center that this approach not only often works but has helped journalists develop industry contacts. In some cases, this may even lead to a company representative proactively reaching out to the requesting journalist, knowing that they may report on the information. Of course, in some instances, journalists may be wary of putting a company on notice that they intend to scrape the site.

Regardless, whether they contact the company or not, reporters should always diminish the risk of injury to the company or other individuals through whatever methods possible. For example, building a tool that accesses the information in a way that performs individual downloads or piecemeal scraping might be less likely to trigger liability. Journalists should be careful not to build any tool that might overwhelm a company's servers. Additionally, if any sensitive data with privacy concerns is collected, journalists should be careful to redact and employ ethical journalistic standards, such as those advocated by the Society of Professional Journalists.

Perhaps most importantly, journalists should be good citizens and ensure that any data collection is done in the public interest. They should have

a clear reason why they need specific information, rather than going on fishing expeditions and scraping in the hope of finding a story. On that point, Nieman Lab recommends keeping "track of process, so that a step-by-step narrative of what steps were taken and why can be presented if necessary."[38]

Finally, media institutions may wish to look to groups like the Electronic Frontier Foundation for guidance. The EFF and similar groups have long offered guidance on the CFAA. Recently, the organization recognized the statute's negative effects on journalism, writing in an amicus brief that the court's reading of the CFAA "threatens to chill socially valuable research, journalism, and testing online, much of which is protected First Amendment activity."[39] The group argues that the "investigative techniques of these journalists and academic researchers sometimes require violating specific company prohibitions on certain activities, and are often adversarial to a company's business interests."[40]

Data and the Freedom of Information Act

Background

Arguably, no other tool is more powerful in bringing transparency to governance than the Freedom of Information Act. In 1966, Congress passed the Act in response to Cold War concerns about corruption and government secrecy.[xxv] At that time, the United States was only the second country in the world to adopt a freedom of information law, and the statute demanding government transparency was considered an indispensable tool for journalists seeking to shed light on government corruption. Indeed, journalists were at the forefront of the push that led to the Act's adoption.[xxvi] As some have stated, "FOIA was . . . designed largely by journalists, for journalists, and with the particular goal in mind that journalists would use access to government information to provide knowledge to the public."[xxvii]

Many aspects of FOIA make it a powerful statute. In addition to establishing a presumption in favor of disclosure, the Act permits any person—citizen or not—to request records, without requiring them to provide a reason or justification for the request. Moreover, FOIA only has nine exemptions that are meant to be "narrowly drawn," according to case law, in order to justify withholding under the statute. State and federal courts have consistently recognized FOIA as vital to promoting transparency.[xxviii]

xxv. See generally, H.R. REP. No. 89-1497, at 12 (1966), reprinted in 1966 U.S.C.C.A.N. 2418, 2429.

xxvi. See Margaret B. Kwoka, FOIA, Inc., 65 *Duke Law Journal*, 1361, 1369–71 (2016)

xxvii. Kwoka, FOIA, Inc., 65 DUKE L.J. at 1371.

xxviii. See, e.g., Nat'l Labor Relations Bd. v. Robbins Tire & Rubber Co., 437 U.S. 214, 242 (1978) (noting that "[t]he basic purpose of FOIA is to ensure an informed citizenry, vital to the functioning of a democratic society"); Kish v. City of Akron, 2006-Ohio-

Over the years, thousands of journalists have used FOIA and its state-level progeny[xxix] to uncover important stories critical to the public interest.[41] Stories have varied from how many cans of Ensure the government bought to force-feed detainees at Guantanamo Bay, to documents unveiling waterboarding techniques, to how in Obama's final year the US spent $36 million in records lawsuits.[42] [43] [44] Investigative journalists are especially reliant on the statute. Some have even built a career on it. For instance, as a senior investigative reporter at BuzzFeed, Jason Leopold has created a cottage industry profession around FOIA. He has even been named a "FOIA terrorist" by several agencies, which amounts to an immense honor among transparency advocates.

While FOIA has always been a valuable tool, its use has increased markedly as the intrigue and value in data has escalated. To date, the total number of requests made on all agencies has increased by over a third from a decade ago, jumping from 557,825 in 2008 to nearly 800,000 in 2016. This increased use is not limited to journalists, however. Reporters account for only a small share of FOIA requests (approximately 7.6 percent according to a 2017 study by FOIAMapper.com).[45] Instead, the majority of FOIA requests come from a diverse ecosystem of "lawyers, nonprofits, academic researchers, hospitals, political committees, hedge funds and corporations."[46]

The mechanization of FOIA has even been realized by artists and activists. For instance, activist Parker Higgins wrote a script that lets him automatically send a FOIA request for the FBI file of every public figure listed in *The New York Times*'s obituary pages. (It's probably fair to say that this has contributed to Mr. Higgins being on the FBI's Vexsome Filer list, which also was revealed by a FOIA request.)[47] Even more notably, Michael Best, an activist from the transparency advocacy organization MuckRock, which requests, analyzes, and shares government documents to

1244, 17 (noting that "[s]uch statutes . . . reinforce the understanding that open access to government papers is an integral entitlement of the people, to be preserved with vigilance and vigor").

xxix. Today, in addition to the federal FOIA, all 50 states and the District of Columbia have adopted an open record law, and more than 100 countries around the world have also followed suit. See, e.g., Herald Publ'g Co. v. Barnwell, 351 S.E.2d 878, 881 (S.C. Ct. App. 1986) (noting that all 50 states have adopted open records laws).

promote government transparency, wrote a script automatically requesting files on nearly 7,000 dead FBI employees and released that script to permit others to create similar projects. In a quote for Motherboard, Best said, "I think simple tools like that can be turned into 'weapons of mass transparency.'"[48]

Unfortunately, the abundant and vigorous exercise of the statute, coupled with both the increase in the amount of data created by the government and corporate incentives behind acquiring data, has led to problems with the current implementation of the statute. Many academics, including David Pozen at Columbia Law School, have taken to critiquing these problems with the legislation by highlighting that the system creates an antagonistic posture between public and citizen, and doesn't focus on other places of interest like corporate accountability.[49] [50] Similarly, faculty members at Syracuse University tracking FOIA have produced a number of studies finding the less than satisfactory efficacy of FOIA. Attorney Mark Zaid, a member of the National Archives and Records Administration, said that while "there are a lot of excellent FOIA officers," there are some that don't have the resources; then others like those from the CIA "go out of their way to be difficult."[51] [52] [53]

Given these criticisms, in 2015 Congress held a hearing about the declining efficacy of the statute, discussing delay, the misuse of certain Exemptions, the lack of responsiveness, and the few resources allocated to FOIA offices. At the hearing, Representative Timothy Walberg (D-Mich.) noted that only 30 percent of all FOIA requests resulted in full disclosure in 2013.[xxx] The vice president and assistant general counsel of *The New York Times*, David McCraw, also spoke, discussing how delay often leads to litigation, which is a misuse of government. His statement highlighted that delay often occurs because of a "culture of unresponsiveness," referrals made to other agencies, and consultations with companies that submitted the information sought for in the request.[54] While the statute was subsequently amended in 2015, delays continue to be an issue.

Since the most recent administration's takeover, various agencies have re-

xxx. "Ensuring Transparency through the Freedom of Information Act (FOIA)," Hearing Before the Committee on Oversight and Government Reform, House of Representatives, June 2, 2015, Serial No. 114–80

moved public information from their websites and made filing requests more difficult (by requiring that they not be made by email) and stonewalling requesters.[55] [56] In 2017, the White House announced it will no longer voluntarily post its visitor logs, breaking with precedent of other presidencies. And in March of this year, reviewing the Trump administration's compliance with FOIA, the Associated Press reported that roughly one out of five FOIA requests is completed fully; other requests were only partially completed or denied. Moreover, the AP noted that the 2017 annual Office of Information Policy report on FOIA—which provides statistics on requests and fulfillment—has not yet been processed.[57]

Response to this inaction has been tremendous. Activists from MuckRock have suggested holding the Trump administration accountable by filing more FOIA requests, and starting the "FOIA the Trump Administration" project and a Slack channel "to share ideas and get help with your requests."[58] [59] A 2017 article in CJR suggests that requests are no longer sufficient, and recommends most requesters also file an appeal.[60] Meanwhile, a review by Politico identified 55 public records lawsuits since Trump's inauguration.[61] Still, while there is little consensus on the changes needed, most agree that at least some further overhaul of the law is required to address the changing issues, given the growing amount of data subject to FOIA.

Influx of corporate requests and little corporate transparency

Congress created FOIA to benefit the public, journalists, and academics, but the Act has long been used as a competitive tool for gain by corporations.[62] As early as 2006, legal scholar Mark Fenster wrote that "frequent FOIA requesters include businesses that seek the records of competitors for commercial motivations."[63] Despite common opinion, Fenster and others have found that federal agencies received only a small percentage of FOIA requests from journalists.[64] In fact, Fenster explained, as the commercial importance of data has grown, so too has the corporate venture of FOIA. He added, "One unintended consequence of transparency, then, is the transfer of wealth between corporations inside, as well as outside of, the United

States when competitors and foreign governments obtain information about American industry that is submitted to or collected by the government."[65]

While FOIA was never meant to necessarily exclude corporations—stating that a FOIA request "may be made by 'any person,' a broad term that . . . includes individuals, partnerships, corporations, associations, or public or private organizations other than an agency"[66]—the trend of using it for corporate gain has intensified in recent years as data becomes more valuable to corporations. In 2016, for example, legal academic Margaret B. Kwoka emphasized how "corporations, in pursuit of private profit, have overrun FOIA's supremely inexpensive processes and, in so doing, potentially crowded out journalists and other government watchdogs from doing what the law was intended to facilitate: third-party oversight of governmental actors."[67] Kwoka's study revealed "a cottage industry of companies whose entire business model is to request federal records under FOIA and resell them at a profit, which distorts the transparency system even further." In conjunction, with the growing interest in data, there has been an increasing number of public records created in government agencies and privatized arms of government. For instance, in a report issued by the Obama administration, it mentions that there is more public information than ever before.[68]

These trends have had several negative effects on journalists utilizing FOIA. First, more data and more requests have created increasing delay for journalists looking for information. This is especially the case as government search capabilities are generally inferior to search abilities in the private sector. And while the statute suggests that journalists should be privileged with "expedited processing" and fee waivers, journalists have said anecdotally that the delay to FOIA has become increasingly pervasive.[69] This problem is especially an issue for freelance journalists without access to lawyers willing to sue on their behalf, as they are otherwise unlikely to get a response from an agency.[70]

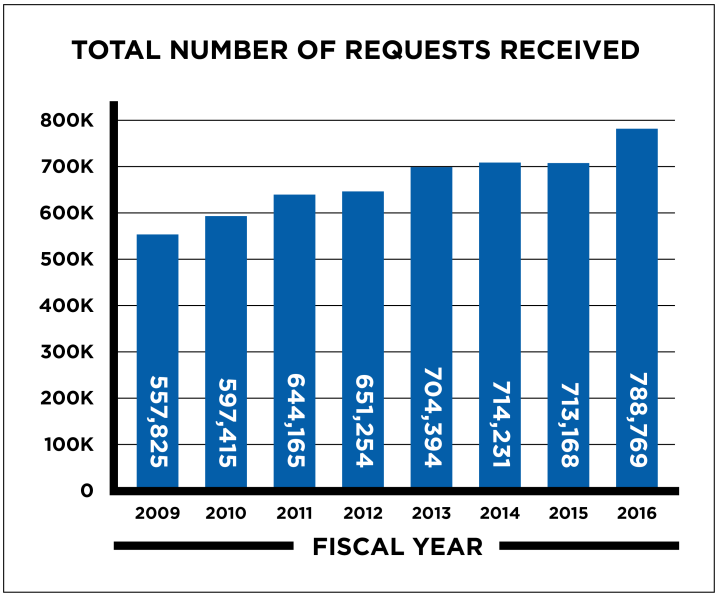
Second, the increased value of data has encouraged governments to begin asserting proprietary ownership over documents they hold, demanding that journalists pay a fee for producing digital information that is inherently public. Newsrooms have been asked to pay tens of thousands of dollars to complete public records requests.

Lastly, the increase in data production, in combination with the privatization of government responsibilities, has led to government control and possession of corporate information that those businesses are hesitant to share. Governments have shown increasing warniness around disclosing private information, leading to a litany of reverse-FOIA actions (i.e., lawsuits brought by corporations under FOIA to shield ostensibly "private" information from disclosure).[71]

Delay problems

Under FOIA, an agency is required to respond to a request in 20 days, unless the agency claims that special circumstances qualify for an extension. Delays in FOIA requests have been a perennial issue, but they have been further exacerbated as the number of documents increases exponentially. During President Obama's tenure, backlogs in FOIA across all agencies more than doubled, increasing from 77,377 in 2009 to 159,741 in 2014.[72] In 2012, a Bloomberg News test of the administration's FOIA practices found that 19 of 20 agencies failed to disclose public information within the FOIA's 20-day time limit, and only eight of 57 agencies processed requests within the timeframe.[73]

In part, these delays are a direct result of the rising number of FOIA requests. From 2009 to 2014, the number of requests increased from 557,825 to 714,231, while the number of full-time government employees assigned to handle them decreased from 4,396 in 2011 to 3,838 in 2014.[74] In the 2017 OIP Report, the government disclosed that the number of requests inflated even further to 788,769. Since the new president has taken office, this trend has not improved. The DOJ, which receives a large number of the requests every year, broke a new record in 2017 with 82,088 requests, compared to 73,103 in 2016 and 67,783 in 2015.[75] This steady increase in requests is in line with the overall numbers, which show approximately 230,000 more FOIA requests were filed in 2016 than in 2009.[76]



Source: Summary of Annual FOIA Reports for Fiscal Year 2016. Office of Information Policy, US Department of Justice.

Problems with search

Under FOIA, to achieve the statute's core purpose of disclosure, an agency must perform an "adequate search" for all responsive records.[xxxi] It "must show that it made a good faith effort to conduct a search for the requested records, using methods which can be reasonably expected to produce the information requested."[xxxii] A court will apply "a 'reasonableness' test" to assess whether an agency's search for responsive records was adequate.[xxxiii] This reasonableness test is supposed to be "consistent with congressional intent tilting the scale in favor of disclosure."[xxxiv] However, in demonstrating the adequacy of the search, an agency may simply rely upon "reasonably detailed, nonconclusory affidavits submitted in good faith." With

xxxi. Founding Church of Scientology v. NSA, 610 F.2d 824, 837 (D.C. Cir. 1979)
xxxii. Oglesby v. Dep't of the Army, 920 F.2d 57, 68 (D.C. Cir.1990)
xxxiii. Campbell v. U.S. Dep't of Justice, 164 F.3d 20, 27 (D.C. Cir. 1998)
xxxiv. Steinberg v. U.S. Dep't of Justice, 23 F.3d 548, 551 (D.C. Cir. 1994) (citing Weisberg v. Dep't. of Justice, 745 F.2d 1476, 1485 (D.C.Cir.1984))

increasing data, this minimal requirement has become especially delinquent and unreasonable.

Various requesters have seen the impact of this insufficient standard in recent years, as challenges to the adequacy of an agency search have failed. For instance, in *Judicial Watch, Inc. v. U.S. Dep't of Def.*, a court found it was not enough for a requester to bring "purely speculative claims about the existence and discoverability of other documents" even where the likelihood of the documents was nearly certain.[xxxv] In the case, the requesting organization sought images of the May 2011 raid on Osama bin Laden's compound in Abbottabad, Pakistan. The requester argued that the secretary of defense's office should have been searched because he was likely to possess the responsive images. Despite the near certain likelihood of the documents, the court rejected this argument as "bald conjecture."

Even more illustrative was a case (which the author of this report worked on) involving the FBI's underwhelming search. In that case, the nonprofit the Freedom of the Press Foundation sued the Justice Department under the Freedom of Information Act for rules determining how the agency employs National Security Letters (NSLs) to investigate journalists and sources—the controversial tool the FBI uses to conduct surveillance without any court approval.[77] The rules were so certain to exist that a coalition of three dozen news organizations, including *The New York Times* and the Associated Press, demanded that the agency release them.[78] In response to Plaintiff's FOIA request, the FBI conducted a search of its records and identified 302 pages of records even though the agency did not yield the requested document containing the actual rules.[xxxvi] However, just weeks into the case being filed, The Intercept published a leaked document titled "Appendix G" that contained the FBI's secret rules for targeting journalists' NSLs. The Intercept's publication of Appendix G, which the government did not disclose, brought the government's search squarely into question.[79]

In that case, Judge Haywood Stirling Gilliam Jr. for the United States District Court for the Northern District of California held that the gov-

xxxv. No. 11-890(JEB), 2012 WL 1438688 at *6 (D.D.C. Apr. 26, 2012)

xxxvi. See Dkt. Freedom of the Press Foundation v. United States Department of Justice Doc. 51 No. 30–1 (Hardy Decl.) 1721 & Exs. H, I.

ernment fulfilled its burden to search. The court admitted that the agency "must show that its search for responsive records was adequate . . . " and that the agency should not succeed "if a review of the record raises substantial doubt, particularly in view of well-defined requests and positive indications of overlooked materials."[xxxvii] Still, the court ruled in favor of the FBI, stating courts hold "limited institutional expertise on intelligence matters," making them particularly reliant on "[a]ffidavits submitted by an agency to demonstrate the adequacy of its response." Ultimately, it found "[b]ecause Plaintiff's speculation cannot overcome the FBI's reasonably detailed, nonconclusory declarations, the Court holds that the FBI has conducted an adequate search for responsive records." The court effectively bowed to the agency's claims that it conducted a reasonable search, even in these dubious circumstances.

In the course of another case involving the FBI, *Shapiro v. Department of Justice*, plaintiff Ryan Shapiro, an MIT doctoral student, also became a suspect of the FBI's search techniques. In this case, Shapiro filed 58 FOIA requests to which the FBI gave him 42 "no records" responses, meaning that it told him no records existed that were responsive to his requests. After Shapiro did some investigating, he alleged that the FBI's software used to search and respond to FOIA requests was developed in 1995 and is so old "that it doesn't even have a graphical user interface, meaning no mouse or icons."[xxxviii] Shapiro said the software performs such rudimentary searches that they are "akin to searching a card catalog at a library—when better technology is readily available." He further alleged that the FBI's main records database, the Central Records System, isn't itself searchable, and that bureau personnel use a much more powerful search software—commissioned at tremendous taxpayer expense—which its Office of Information Policy refuses to use to conduct FOIA searches.

In 2016, Shapiro successfully sued the FBI over its use of the two-decades-old database. In his complaint, he cited a 2012 report in which former FBI Director William H. Webster called ACS "the FBI's most outdated system."[80] As documents continue to multiply, it will be important

xxxvii. Hamdan v. United States Department of Justice, 797 F.3d at 771
xxxviii. Shapiro v. United States Department of Justice, No. 17-5122 (D.C. Cir. 2018)

that more regimented search standards are imposed, and that people like Shapiro continue to advocate for more robust search tools.

Equating disclosure of documents with profit

As data has increasingly gained importance in our society, government officials have begun to treat public information as proprietary data—as theirs to sell and profit from. In researching this report, many journalists told us that they have been asked to pay for an agency's release of documents under FOIA. FOIA does account for the possibility of some costs, however, those costs are reserved for efforts such as copying records.[xxxix] In addition, FOIA contains a fee waiver provision requiring that records "shall be furnished without any charge . . . if disclosure of the information is in the public interest.[xl] Still, agencies have attempted to charge for the value of the data itself.

Perhaps the best example is a FOIA lawsuit filed in federal court in the District of Columbia, *Yanofsky v. United States Department of Commerce*, in which Quartz reporter David Yanofsky requested from the United States Department of Commerce two databases containing information about foreign travel to and from the United States. The data, which is of immense value to the agency, was not subject to the FOIA, it claimed. According to the complaint, Yanofsky was told that FOIA was inapplicable and he would have to purchase the records for $173,775.

Reporters Committee for Freedom of the Press, the attorneys representing Yanofsky, put a finer note on the point, stating, "Simply put, the government can't hold public information for ransom." In their brief filed with the court, they wrote, "Indeed, it should go without saying that the mere fact that government information may have a commercial value does

xxxix. §552(a)(4)(A)(ii)(II), (a)(4)(A)(iv)(II) (stating "representatives of the news media are entitled to a fee benefit and may be charged only for duplication costs, not including the first one hundred pages of duplication")

xl. See also Michael Russo, "Are Bloggers Representatives of the News Media Under the Freedom of Information Act?," 40 *Columbia Law Review*, 225, 228 (Winter 2006) ("By favoring journalists, with their professional skills and broad distribution networks, FOIA subsidizes those requests most likely to inform the public.").

not mean that government agencies are permitted to charge more than the cost of dissemination for access to such information. In fact, a wide variety of commercially valuable government databases are made available to the public for free."[xli] On March 30, 2018, the district court decided in favor of Yanofsky and determined the agency must follow FOIA's fee-setting and fee-waiver provisions rather than charging exorbitatant fees. Still, the case is a harbinger of the way agencies are beginning to see information the same way that corporations view it: by dollar value instead of as a means of "ensur[ing] an informed citizenry."[xlii]

Stymieing corporate transparency: Reverse FOIA actions

Perhaps most concerning is not just that corporations have overused FOIA as a monetization tool, but that they have also employed it as a shield against transparency. Corporations hold a privileged position under FOIA, as they are not subject to the same scrutiny as government agencies. Unlike some transparency laws abroad, the Act does not require businesses to disclose information, despite their growing power in our societal structure. Corporations have taken this privileged position to employ so-called "reverse FOIA procedures." Originally developed by the federal courts in the 1970s, these actions allow private actors (or businesses) that submit supposedly proprietary information to an agency to file suit in hopes of thwarting disclosure and stopping information from being revealed to a requester. To be especially cautious, government agencies have begun employing this process, giving corporations notice and opportunity to object to requests before disclosing the information to the public.

For instance, The Center for Investigative Reporting recently filed a complaint challenging an agency's determination that corporations can shield information they have provided to the Equal Employment Opportunity Commission. Silicon Valley companies, which are federal contractors, must file reports to the agency that contain their diversity statistics in or-

xli. Yanofsky v. Department of Commerce, Complaint, Dkt. 21, 27. (2016)
xlii. Judicial Watch, Inc. v. Rossotti, 326 F.3d 1309 (D.C. Cir. 2003) (internal quotations omitted)

der to ensure their compliance with federal law. The agency has asserted that companies can withhold their consent from releasing the information by qualifying the diversity statistics as "trade secrets." In essence, the agency and various Silicon Valley companies have argued that their diversity statistics are somehow confidential business information. In a similar circumstance, a CJR article disclosed that Facebook attempted to create an arrangement with public officials that required it be given notice about any relevant public records requests an agency received at least three days before it responded to them.[81] These circumstances are especially troubling, as private companies are increasingly being contracted to take over public responsibilities.

One possible solution to corporations' ability to block the release of their information—while not comprehensive—would involve heightening the requirements for transparency. In a recent law review article, David Pozen discusses the potential for requiring corporations to meet disclosure requirements similar to those of government agencies.[82] While this proposal has been suggested previously, given the increasing power of Silicon Valley companies and their recent receptivity to accountability discussions, it appears that the time for reform is upon us.[xliii] As Louis Brandeis often repeated, "Sunlight is said to be the best of disinfectants."[83]

Search of a database as creating a new document

A FOIA request can be made for any agency record. The requester may also specify the format in which he or she wishes to receive the records. That said, FOIA does not require agencies to create new records or to conduct research, analyze, or answer questions when responding to requests.[xliv] FOIA only requires disclosure of extant records, which the agency has decided to create for its own reasons. However, what constitutes creating a

xliii. See also Roy Peled, "Occupy Information: The Case for Freedom of Corporate Information," 9 HASTINGS BUS. L.J. 261 (2013).
xliv. See, e.g., Kissinger v. Reporters Committee for Freedom of the Press, 445 U.S. 136, 152 & n.7 (1980) (FOIA "does not obligate agencies to create" records); NLRB v. Sears, Roebuck & Co., 421 U.S. 132, 162 (1975).

new record is uncertain. In fact, Congress did not even attempt to define this in the statute. Unfortunately, many questions around what constitutes a document are becoming critical as agencies move to electronic documentation. As the DOJ writes:

> Indeed, over the course of the past decade, federal agencies have increasingly "computerized" and automated their records systems—sometimes entirely replacing their conventional files, sometimes supplementing them, with more sophisticated "electronic" systems. At many agencies, use of computer databases as a primary means of records maintenance has become the rule rather than the exception. At others, such automation is just beginning to take hold. Overall, throughout the federal government, agencies are now moving with increasing swiftness toward the creation of a new "electronic record" environment in which they can operate with high-tech speed and efficiency.[84]

Given this circumstance, it is especially important that agency records be extended to documents such as databases. However, in the same report, when agencies were asked: "Does the FOIA require agencies to provide requested records in the particular forms (or database formats) specified by requesters?" nearly 60 percent answered in the negative. Increasingly, FOIA requesters have been denied access to databases, being told that to conduct a search would require the agency to create a new document.

Recent cases have discussed this issue. In *National Security Counselors v. CIA*, the court found that when responding to a request for "aggregate data" an agency "need not create a new database or reorganize its method of archiving data," also noting that "searching that database does not involve the creation of a new record."[xlv] The court continued by saying that "sorting a pre-existing database of information to make information intelligible does not involve the creation of a new record . . . [I]t is just another form of searching that is within the scope of an agency's duties in responding to FOIA requests." However, the court did note that a request for a "listing or index of a database's contents that does not seek the contents of the database, but instead essentially seeks information about those contents, is a request that requires the creation of a new record, insofar as the agency has not previously created and retained such a listing or index."

xlv. Nos. 11-443, 11-444, 11-445, 2012 WL 4903377 (D.D.C. Oct. 17, 2012) (Howell, J.)

On the state level, an Illinois court on appeal reversed a lower court in part, holding that the data requested within the agency's database (e.g., zip codes) were public records under the state FOIA law because such data is clearly information recorded during a government function performed by the public body. The court determined that running a search for records did not constitute creating a new record, just as looking through a file cabinet for a file does not create a new record. It likewise warned that as the use of electronic records grows, so does the possibility of an agency trying to shield public information from review by merely storing it in a database.

Decided in the opposite direction, the Superior Court in California recently ruled under the California Public Records Act (CPRA) that the anonymization of records is tantamount to the creation of a new record. The Reporters Committee for Freedom of the Press filed an amicus brief in the case, *Sander & The First Amendment Coalition v. State Bar of California et al.*, arguing that the ruling "jeopardizes the public's ability to access anonymized data from government databases, a vital source of information about the workings of government." It continues that the "holding will, as a practical matter, put many government databases beyond the reach of the CPRA [because a] number of California government entities subject to the CPRA collect large amounts of data that may include private information that could necessitate redaction or deletion before disclosure under the Act."[85] Furthermore, legally speaking, the holding goes against the rule of the CPRA given that "when a record contains both exempt and non-exempt information, the Act requires 'deletion' of portions of the record that are exempt and that the public be afforded access to non-exempt portions of the record."[xlvi]

More recently, the author of this report filed a FOIA lawsuit on behalf of The Center for Investigative Reporting against the Bureau of Alcohol Tobacco Firearm and Explosives (ATF), after the agency claimed that a search of its database storing gun trace information would be producing a new document.[86] In that case, the court ruled in favor of the ATF, stating that the request for "statistical aggregate data" from the database amounted to a new document. There, the court held, "Plaintiff's request

xlvi. (Gov. Code §6253)

for statistical aggregate data derived from the FTS database requires a compilation of data points—all firearms traced to former law enforcement ownership since 2006—and seeks information that does not currently exist."[87]

While many courts have acknowledged that searching and anonymizing a database is not creating a record, questions around what a record is and rules around analyzing that database are sure to continue. And holding that searches of electronic databases are tantamount to creating a new record is sure to create new levels of government secrecy as more public information is held in these formats.

Data and Whistleblowing

In 2014, President Obama issued a report in which he signaled concern arising from the revolutionary shift in the "volume, variety, and velocity" of information growing in our society—especially within the walls of government.[88] The report presaged a variety of hot topics involving government data and algorithms, including implicit bias and discrimination, artificial intelligence, and attention deficits. However, hidden in the report, in a section titled "Insider Threat and Continuous Evaluation," he suggests another issue: concern over "a string of troubling breaches and acts of violence by insiders who held security clearances, including Chelsea Manning's disclosures to WikiLeaks . . . and the most serious breach in the history of U.S. intelligence, the release of classified National Security Agency documents by Edward Snowden."[89] The report continues, "We must ensure the big data analytics powering continuous evaluation are used in ways that protect the public as well as the civil liberties and privacy rights of those who serve on their behalf."[90]

This understated section in the report highlights that as data grows, so should the Fourth Estate's wariness over the safety of its sources and journalists. This concern appears legitimate, arising in large part from the growing number of leak prosecutions in recent years. In the 100 years since the Espionage Act of 1917 passed, there have been a grand total of only 14 prosecutions targeting leakers. The majority of these prosecutions have only come in the last decade during the boon of data: eight under President Obama's administration, and one in the first six months of President Trump's administration. In fact, more leak prosecutions occurred under the Obama Administration than under all other previous administrations combined.

This is exceptionally troubling given that confidential sources confiding in journalists is the foundation for some of the most democratically important and honored news reporting in our country. Canonically, more than

150 articles authored by Bob Woodward and Carl Bernstein were based on unnamed sources. More recently, based on reporting from confidential sources, *The New York Times* was able to reveal that the National Security Agency had been monitoring phone calls and email messages without seeking approval from federal courts.[91] Several other news organizations reported on the use of harsh interrogation tactics against terrorism suspects."[92] And since the beginning of the Trump administration, confidential sources have been central to reporting on issues involving our democracy.

Even more troubling perhaps than these worries around sourcing is the increasing discussion and legal theory developing for imprisoning not just sources, but journalists. President Obama's Department of Justice was first to utilize the legal argument in a complaint asserting that a journalist, James Rosen, was a co-conspirator to his source's actions.[93] While the charges were later dropped, the legal theory seems to have been resurrected, at least in conversation, within President Trump's administration. Since President Trump assumed office, Attorney General Jeff Sessions has stated that criminal investigations into the sources and journalists are up 800 percent.[94] When asked if he would imprison a journalist, Sessions responded he might if it "makes things easier."[95] In addition, the administration has warned government employees that leaks will be penalized. While the administration's efforts to clamp down on leaks are not a deviation of precedent, its strong insinuations that it will prosecute not only leakers, but also news organizations and journalists, is an aberration from decades of custom.[96] Additionally, other actions like arrests, equipment seizures, impersonation of journalists, subpoenas, and possible tracking of journalists have also arisen as concerns.[97]

Leaks are integral to democracy, however, as they help keep the government in check. As Daniel Ellsberg, the whistleblower in the Pentagon Papers case, recently put it: "Can you really have democracy, in a real sense, with the government having the final voice and the total voice as to what citizens shall know about what they're doing, and whether they're telling the truth and whether they're obeying the law? I would say no."[98] This view has been echoed by famed legal scholars and First Amendment attorneys, such as Alexander Bickel who presaged that while the government must "guard mightily" to protect its information, the press must oppose

and contest this force and publish everything "newsworthy" to preserve the equilibrium of information in our democracy. The following explores the complex nexus between government, leakers, journalists, and the law in a landscape where leakers have more data than ever before at their fingertips. It investigates new concerns around leaks, the current administration's approach to them compared to previous administrations, the laws in place that can be used to penalize journalists and leakers, as well as how leakers and journalists can protect themselves (and their sources) in a world where big data leaks are becoming commonplace.

New risks with leaks

Whistleblowers have always played a central role in newsgathering. In the famed 1971 Pentagon Papers case, the Nixon administration tried to prevent *The New York Times* and *The Washington Post* from publishing classified documents regarding the Vietnam War. In an affidavit submitted for the case, which established the right of a news publication to print classified documents leaked to a newspaper despite a prior restraint, *New York Times* Washington Bureau Chief Max Frankel explained that leaks are intrinsic to journalism and are the "currency" of American democracy.[xlvii] Similarly, legal scholars like Alexander Bickel have by implication underscored the important role leaks play to preserving the role of the press protected under the First Amendment.[99] Bickel argued that the Pentagon Papers ruling vindicated the "disorderly situation" or the contest between the government and the press for information and "the presumptive duty of the press . . . to publish, not to guard security or to be concerned with the morals of its sources."[100] Today, however, because of the abundance of data, changes in the legal landscape, and the shifting attitude toward the press, everyday circumstances cause confusion and hinder the reporter-source relationship.[101] Before discussing the legal implications, it is important to note several recent developments in leaking. The first is the size of leaks and their technological implications; the second is the legal guidelines used by the DOJ and the changes within them.

xlvii. Affidavit, United States v. New York Times Co. et al., U.S. District Court for the Southern District of New York, Civil File 71-2662, RG 21

1. The size of leaks

In 2010, WikiLeaks published 400,000 pages or 1.7 GB of classified materials, comprised of documents, videos, and diplomatic cables sent by former soldier Chelsea Manning. Subsequently, the site began near continuously publishing leaked documents from various sources. Just three years later, Edward Snowden leaked 1.7 million documents about the NSA's surveillance programs to *The Guardian*. While the industry could not have imagined a more copious leak, in 2016 several news organizations published stories on the Panama Papers, the largest leak in history consisting of 11.5 million documents and a staggering 2.6 terrabytes of data.

The size of these leaks has various practical and legal implications. Unlike the 7,000-page Pentagon Papers leaked in 1971, these "big data" leaks can no longer be manually perused—or, for that matter, manually handed over to newsrooms in cardboard boxes. Instead, the documents are often transmitted electronically to newsrooms through special security tools, and then searched with dedicated tools to determine if any of it is in the public interest and worth publishing. In addition to the mere difficulty of transferring large files of data, there are also various legal surveillance and technological tactics in place that allow the intelligence community and law enforcement to access electronic communication records and ascertain what information has been shared with journalists. For that problem, tools like SecureDrop, an open-source whistleblower submission system used by media institutions, have been employed to securely transfer documents without revealing a source's identity.

However, SecureDrop is not a panacea. For instance, digital communications leave footprints, allowing the the source, recipient, or both, to be identified, making it easier for authorities to initiate prosecution proceedings. The most recent example of this arose with Reality Winner, who leaked documents to The Intercept last year.[102] In this case, The Intercept published the documents—which were watermarked and traceable—online, leaving Winner easily susceptible to prosecution. Some newsrooms have avoided this problem by creating new documents altogether before publishing so that any watermarks or signage is completely removed. Additionally, newsrooms have always employed traditional tactics like redactions, verification, and anonymous sourcing.

Alongside the concern over revealing sources, the size of these leaks makes it harder to ensure proper redaction of libelous or private information that a newsroom can be held liable for publishing. As a basic tenet of defamation law, anything a newsroom publishes—even if it is contained in documents—can trigger liability. While redactions work, they require much more tenacity than ever before. For instance, while a judge recently ruled that BuzzFeed may have protection for publishing the Trump dossier, the lawsuit shows how the lack of a redaction may lead to possible liability and expensive litigation costs.

Due to the difficulty in maneuvering large leaks, there have even been cases where though news organizations have opted not to publish the leaked data in raw form, stories about its existence—which wouldn't result in legal liability (because the information wasn't published)—still had potential to cause harm to individuals. For instance, when Luke Harding and David Leigh published their book, *WikiLeaks: Inside Julian Assange's War on Secrecy*, they disclosed the passphrase that Assange had used to encrypt the leaked embassy cables, assuming it was no longer in use and that the files were not available publicly. Unbeknownst to the journalists, even though WikiLeaks had removed the files from its servers, they ended up on BitTorrent and were accessible to the public. Not only did these files include details like the queen of the Netherlands' personal phone number, they also contained the names and contact information of informants in Israel, Jordan, Iran, and Afghanistan, some of whom had to be relocated as a result of this breach.

Cases like this one demonstrate how even if journalists follow protocol and are not hasty about publishing content without thinking through the consequences, human folly can catch up with them and put others—and their newsrooms—at risk.

2. Possible changes to the DOJ guidelines

In the event of leaked information being published by a news outlet, the government is bound to immediately take action to find the source and stop the transmission of information. In order to identify the whistleblower, the government can initiate various processes, including bringing in the FBI and CIA. Other questionable tactics have also been employed. For instance,

past administrations have surreptitiously wiretapped journalists to identify their potential sources. President Kennedy initiated Project Mockingbird, which allowed the government to eavesdrop on the communications of journalists who covered leaks; President Nixon had people surveilling journalists; and during President Obama's administration, the phone records of journalists were seized. These schemas offend our notions of First Amendment press protections. Thus, to protect against investigative tactics that would obstruct the First Amendment, the Justice Department has created various rules—the most well-known are commonly referred to as the Media Guidelines—to be followed.

For over 40 years, outlined in its own policy, the Department of Justice has implemented a series of procedures for its law enforcement when investigating members of the press with respect to the First Amendment rights of the news media.[xlviii] The Guidelines were created in 1972 during the Nixon presidency to protect against the chilling effect that investigative tools have on the "reporter-source relationships and therefore on the flow of information to the press."[103 xlix] Overall, the Guidelines announce three procedures federal officers must meet before they can, as a "last resort" after "all reasonable alternative attempts" have been taken, subpoena information about the press.[l] First, the Guidelines require advance authorization by the attorney general before issuance. Second, they recommend the agency provide "reasonable and timely notice" to the affected news media member of the request before issuance. Third, they provide that notice opens the door to possible judicial review.

In August, Attorney General Jeff Sessions said a new task force had been established inside the FBI to focus on leaks to the press and public. During this announcement, the attorney general also said he was reviewing the Media Guidelines that had been overhauled by the previous attorney general, Eric Holder, in 2014. The amendments made to the Guidelines in 2014 were done so after convening over 30 news organizations, as well as First Amendment groups and academics, to obtain insight from these

xlviii. 28 C.F.R. §50.10 (hereafter "Section 50.10 Guidelines" or "Guidelines")

xlix. David McCraw, National Security Letters and Leak Investigations, Just Security Jan. 20, 2016.

l. S 50.10 (a)(3) (describing subpoenas as a "last resort" and "all reasonable alternative attempts" must first be taken)

stakeholders when revising its policies. Reviewing and possibly changing these Guidelines to make it easier for the FBI to subpoena news media records is seriously concerning, especially given that the current rules are an expression of built consensus from various stakeholders. Moreover, these rules already have loopholes that create potential for obfuscating the spirit of the Guidelines in a data-driven landscape.

Since they were created, the Guidelines have been strengthened several times in response to various instances where the government improperly investigated the press. In 1980, they were fortified to protect journalists' telephone toll records held by third-party telephone companies after it was disclosed that the FBI obtained toll records of *The New York Times*'s Atlanta bureau and home telephone records of the bureau chief by subpoena. The FBI had also directed the telephone company not to notify the paper for a month and a half, thereby stymieing any possible challenge to the subpoenas.[li]

Most recently, in 2014 the guidelines were expanded to include additional procedures after public outcry about government investigations into Fox and AP reporters. But these changes appear to have left holes: Today, newsrooms use online applications and social media platforms rather than notebooks and pens that are kept in their offices, and the government can more easily subpoena digital materials from third parties than work products from a newsroom.[104] Similar to the events of 1980, in May 2013 the Associated Press reported that the government had surreptitiously seized two months of telephone records for at least 20 AP phone lines, including home and cellphone records used by journalists to communicate with sources.[105] [106] The events were a "massive and unprecedented intrusion" into newsgathering activities of the press, according to the AP.[107] Within the same week, Fox News correspondent James Rosen was identified in a search warrant as a "co-conspirator" to a leak, enabling the FBI to obtain

li. Policy with Regard to Issuance of Subpoenas to Members of News Media, Subpoenas for Telephone Toll Records of Members of News Media, and Interrogation, Indictment, or Arrest of, Members of News Media, 45 Fed. Reg. 76,436 (Nov. 19, 1980) (to be codified at 28 C.F.R. §50.10)

the journalist's personal emails in connection with the investigation.[lii] Fox News was "outraged" and called the incident "downright chilling."[108]

The events alarmed the news media, members of Congress, and the country. Fifty media organizations wrote a letter to Attorney General Eric Holder saying, "The nation's news media were stunned" by the event because "none of us can remember an instance where such an overreaching dragnet for newsgathering materials was deployed by the Department, particularly without notice to the affected reporters or an opportunity to seek judicial review."[liii] Members of Congress similarly voiced concern about the government's breaches.[liv] Even the president showed concern and directed Attorney General Eric Holder to "review existing Department of Justice guidelines governing investigations that involve reporters."[109]

Upon President Obama's instructions, a group of media organizations was convened by the Department of Justice to help revise and strengthen the Guidelines. A coalition of over 50 news media organizations led by the Reporters Committee for Freedom of the Press submitted comments.[110] Subsequently, Attorney General Holder "personally held seven meetings with approximately 30 news media organizations as well as with First Amendment groups, media industry associations and academic experts" and the Justice Department issued a public report outlining its revisions.[111] Among its revisions, the Justice Department underscored that the Guidelines would require additional authorized review, advanced notice, and judicial oversight. For one, they required approval from the attorney general for tools, such as court orders issued under the Stored Communications Act, 18 U.S.C. §2703(d), used to obtain records related to the news media.[lv] Notice to, and negotiations with, members of the news media whose records are sought, either directly or from third parties, was also highlighted.[112]

lii. Letter from Gary Pruitt, Exec. President & CEO, Associated Press, to Attorney General Eric Holder, Dep't of Justice (May 13, 2013)

liii. Letter from Reporters Committee for Freedom of the Press and 50 other news organizations, to Attorney General Eric Holder, Dep't of Justice (May 14, 2013), available at http://rcfp.org/x?xszk.

liv. Letter from Chairman Bob Goodlatte, House Judiciary Comm., and Chairman F. James Sensenbrenner, Jr., Crime, Terrorism, Homeland Security, and Investigations Subcomm., to Attorney General Eric Holder, Dept of Justice (May 29, 2013)

lv. See Office of Attorney General, Memorandum to All Departments January 14, 2015; see also §§50.10(a)(3), (c)(1), (c)(4)(i), and (c)(5)(i).

In February 2014, the new Guidelines were formalized. After convening a group one last time to review the amended Guidelines, they were revised again in January 2015 to clarify ambiguous language involving internal training requirements and updates to internal manuals for field officers.[113]

Despite this concerted effort, the Guidelines were not comprehensive and lacking in a way that is worrisome given our new data landscape.[114] The Guidelines remained silent as to NSLs, thereby possibly allowing the FBI to circumvent the Guidelines' requirements of authorized approval, notice, and judicial oversight for the news media—for instance, an agency may go to third parties to obtain journalists' records. Lack of authorization with NSLs is especially concerning because the government has previously reported the FBI abused use of similar tools, such as exigent letters, against the press.[115] In fact, Barton Gellman, *The Washington Post* reporter who covered the Snowden leaks, has been told that his phone records have been obtained via an NSL, meaning he spends a lot more of his time on "technical and operational security," providing potential sources with various options across different domains, along with relevant instructions: email, chat applications, or SecureDrop.[116] Moreover, it has been reported that an average of nearly 60 NSLs are being issued daily.[117]

Even more significant is that using NSLs the FBI may fail to provide the news media with proper notice, as required by the Guidelines, in analogous circumstances. Today, as journalists often use a variety of tools, apps, and calendars on their phones and computers, the agency can go to online providers like Google, Facebook, LinkedIn, Twitter, etc. (where reporters speak to their sources, keep their calendars, and source contacts) instead of to a news organization to obtain a journalist's information. Since many newsrooms today use Gmail for their email clients, it is more likely that NSLs will be issued on various internet companies like Google to obtain journalistic information; NSLs already have nondisclosure requirements or gag orders imposed on the third parties, making it easy to muzzle those parties and obviate the spirit of the Guidelines.[lvi] Indeed, in 2016 The Intercept published leaked secret portions of the Domestic Investigations and Operations Guide, which indicate that FBI agents have been allowed to obtain journalists' phone records with the approval of only two internal of-

lvi. See 18 U.S.C. §2709(d).

ficials.[118] It is worth noting that The Intercept also raised the question of whether other procedures may exist that circuvment the Guidelines, such as procedures involving the Foreign Intelligence Surveillance Act.[119]

Unfortunately, this seems likely. Attorney General Sessions recently stated that the Department of Justice has 27 leak investigations open, which is nine times as many investigations as last year. In November 2017, Marty Baron, the executive editor of *The Washington Post*, said, "It's possible that some *Post* reporters are ensnared in leak investigations right now, but *The Post* wouldn't know it unless there was a subpoena, a search warrant or 'an actual leak about the leak investigation.' "[120] Similarly, two media organizations have resorted to suing the Trump administration to find out what surveillance procedures have been adopted to target journalists. The Knight Institute and Freedom of the Press Foundation recently filed a FOIA lawsuit to have the agency disclose any revisions to the FBI Domestic Investigations and Operations Guide (known informally as the DIOG) that concern the use of secret "national security letters" and any other secret tools.

The Espionage Act: Penalties for the journalist and the source

Leaked information can frustrate and embarrass an administration, but that does not make leaks de facto illegal. There is currently no clear, black-and-white course of action available to the government—that is to say, no single statute that the government uses to prosecute. Instead, as a 2017 congressional research survey points out, "The legal framework is based on a complex and often overlapping set of statutes with provisions that differ depending on, among other factors, what information was disclosed, to whom it was given, and the intentions of the discloser."[121] Generally, under this framework, the leaker is more liable than the journalist or the news organization. Referring to this as the "source-distributor divide" in an article for *The Harvard Law Review* titled "The Leaky Leviathan," Columbia Law School professor David Pozen says, "The First Amendment has been construed to provide so little protection for the leaker and yet so much protection for the journalist who knowingly publishes the fruits of the

leaker's illicit conduct and thereby enables the very harm—revelation of sensitive information to the public and to foreign adversaries—that the leak laws were designed to combat."[122]

The traditional reporter-source divide: Pentagon Papers and Bartnicki

From a legal perspective, historically it has been generally assumed that a reporter would be protected by the First Amendment in a circumstance of leaking. Two landmark decisions reached by the Supreme Court, *New York Times Co. v. United States* and *Bartnicki v. Vopper*, seemed to provide copious protection.[123] [124] Indeed, in a congressional hearing University of Chicago law professor Geoffrey Stone told congressmembers that prosecuting journalists would outright violate the First Amendment. However, the notion that reporters are protected from prosecution has increasingly become open to debate.[125]

Perhaps the most important case establishing protection for journalists and their publishers comes from the 1971 decision *New York Times Co. v. United States*, in which the Supreme Court permitted *The Times* to print the Pentagon Papers—then-classified documents about the Vietnam War that the Nixon administration wanted to keep secret. The decision cemented the rule that prior restraints on publishing could only withstand the strictest standard if the government can show that the publication would cause "grave and irreparable" danger.[126] The watershed case made clear that only in dire circumstances where a clear injury would result could a publication be thwarted from sharing the news—even when obtained from a source. That said, several justices wrote separately suggesting that the Department of Justice could still prosecute the journalists after the story was published.

However, those stray remarks seemed to get clarified in perhaps the second most important case on leaking: the 2001 Supreme Court decision in *Bartnicki v. Vopper*. In this case, a radio commentator at a local station, Vopper, played a clandestinely taped conversation between a union negotiator and union president involving a teachers' strike in a Pennsylvania high school. While the recording was determined to be illegally intercepted, the Supreme Court ruled that "the statutes do not forbid the receipt of the

tape itself."[127] A concurring opinion clarified that even though the person who taped it had committed a crime, because Vopper had "engaged in no unlawful activity other than the ultimate publication of the information another had previously obtained," had "neither encouraged nor participated directly or indirectly in the interception," and because publication was in the public interest, no liability could ensue.[128]

These precedents have led most media lawyers to understand that the press is rather free to publish information, even if obtained illegally, so long as the journalist does not originally participate in acquiring the information, and its publication is in the public interest and would not cause irreparable harm. While the Bartnicki precedent has not yet been tested on any case involving national security or the Espionage Act, it is often paired in the minds of media attorneys with the Pentagon Papers to solidify the right to publish even the most sensitive government information—such as the recent Snowden leaks.

Even so, in the recent past there have been occasions when these protections have come into question, and the government has considered bringing forward charges against the distributors—news organizations or individual journalists—as well. For example, as previously mentioned, in May 2013 Fox News Washington correspondent James Rosen was named "at the very least, either as an aider, abettor and/or co-conspirator" in an espionage case involving a State Department security adviser. Additionally, both the previous administration and the current one have contemplated bringing espionage charges against WikiLeaks, which would set a dangerous precedent for other news media organizations.

What follows explores the legal options available to prosecute both the leaker and the journalist or news organization for different types of leaks. While possible statutes include the the Intelligence Identities Protection Act (1982), the Atomic Energy Act (1954), and various federal statutes codified under US Code (U.S.C.), the focus is on the Espionage Act (1917), the most likely option to be used.

Open questions under the Espionage Act

When America entered World War I, the Espionage Act was passed to thwart espionage activities, including collection and communication of sensitive national defense information that would benefit American enemies. Passed in 1917, the law's one-hundredth anniversary came in 2017. A century later, the Espionage Act seems to have turned into a broad anti-disclosure law akin to the UK's Official Secrets Act, which is used by the government to prevent any and all leaks pertaining to national defense. More specifically, it is irrelevant whether the leak is in public interest.

To date, 14 leakers have been prosecuted for violating the Espionage Act in the last 100 years, nine of whom have been charged in the last decade. Even in cases where the administration acknowledges that the unauthorized disclosure was in the public interest and had minimal real-world impact, the government remains adamant that the leaker must be penalized.

Thus far, though, no journalists or news organizations have been prosecuted, despite there being "nothing in the language of the law that prevents its use against a news organization."[129] However, as already mentioned, the government is getting more aggressive in dealing with leaks pertaining to national security, be it via labeling journalists as co-conspirators or attacking the integrity of news media itself.

Whether the *Bartnicki v. Vopper* ruling provides journalists absolute protection under the Espionage Act is unknown, but it should not be taken for granted for multiple reasons. First, prosecutions under the Espionage Act are for leaks involving classified government information, whereas *Bartnicki* involved private information. Second, the act is incredibly broad in language, making both the receipt of sensitive information and the retention of that information unlawful, not just its publishing. Lastly, the Espionage Act deals with national security, which is a continually growing category wherein the state secrets privilege grows in expanse every day. As detailed in the "Developments in International Media Law" bulletin published by the Media Law Resource Center in October 2017, the authors warn that concerns about the Espionage Act prosecuting journalists is not merely hypothetical.[130]

There are two main sections of this Act under which news organizations, journalists, and leakers can be indicted: Sections 793 and 798, which pose even more concern in our data-heavy environment.

Section 798 of the Espionage Act explicitly bars publishing, unlike any of the subsections of §793. The text reads: "Whoever knowingly and willfully communicates, furnishes, transmits, or otherwise makes available to an unauthorized person, or publishes, or uses in any manner prejudicial to the safety or interest of the United States or for the benefit of any foreign government to the detriment of the United States any classified information." Classified information here includes code, cipher, cryptographic systems, and communication intelligence activities, all pertaining to national security. Only one national security leaker has been charged for violating §798: Shamai Leibowitz. There is also a still-pending criminal complaint which charges Snowden with violating §798.[131]

Unlike 798, Section 793 of the Espionage Act is even more draconian because it "prohibits certain activities related to gathering, receiving, or transmitting national defense" even if the journalist does not publish the information as long as one is "not entitled to receive it." The section has eight subsections, six of which are relevant to leaks to news media, and three of which have been used to charge unauthorized disclosures. Section 793(a) prohibits obtaining information connected with national defense infrastructure that is owned by the United States, which is then "to be used to the injury of the United States, or to the advantage of any foreign nation," and §793(b) clamps down on obtaining or duplicating "any sketch, photograph, photographic negative. . . document, writing, or note of anything connected with the national defense." Perhaps even more stringent is §793(c) which implicates anyone or any organization that receives or attempts to receive any material related to national defense, provided the recipient knows that the material was sourced by violating other provisions of the Espionage Act. Daniel Ellsberg of Pentagon Papers fame is the only leaker charged with violating §793(c).[lvii] According to this language it ap-

lvii. §793(d) incriminates anyone with security clearance who "willfully communicates, delivers, transmits or causes to be communicated, delivered, or transmitted or attempts to communicate, deliver, transmit or cause to be communicated, delivered or transmitted the same to any person not entitled to receive it" material pertaining to national security. Daniel Ellsberg, Samuel Morison, Lawrence Franklin, Stephen Jin-Woo Kim, Jeffrey

pears unclear whether simply obtaining documents through SecureDrop could be construed a violation.

However, §793(e) is considered by some to be the section that poses the greatest and broadest risk to journalists, as it targets those who have "unauthorized possession of, access to, or control over" material related to national defense, and who subsequently "wilfully communicates, delivers, transmits or causes to be communicated, delivered, or transmitted, or attempts to communicate, deliver, transmit or cause to be communicated, delivered, or transmitted" among those not entitled to receive it. There is little in the language of this section to limit application of the section to journalists who did not themselves take documents from the government. Thus, this makes *Bartnicki*'s possible defense application even more inapplicable.

Historically, the courts seemed to side with news organizations responsible for publishing the sensitive material, lessening concerns around §793(e). In *New York Times Co. v. United States*, Justice Douglas said that the word "communicates" in §793(e) does not encompass publishing by news media; if it did, the subsection would explicitly say so, as "Congress was capable of, and did, distinguish between publishing and communication in the various sections of the Espionage Act."[132] Moreover, some precedent suggests that "reason to believe [the leaked documents] could be used to the injury of the United States"[lviii] is a stringent requirement. In *United States v. Rosen*, many feared that the case, involving two lobbyists, could become a precedent for the prosecution of journalists under the Act. The government eventually decided to abandon its case after a federal judge held that the prosecution would have to show beyond a reasonable doubt that the lobbyists acted with a certain "mens rea" or "bad purpose" to disobey or disregard the law, and that disclosure of the documents could harm the United States.[lix] The district court for the Eastern District of Virginia in-

---

Sterling, and John Kiriakou have all been charged with violating this subsection of the Espionage Act. Further, there's a pending criminal charge against Snowden for violating this subsection. §793(f) focuses on gross negligence, where the person who has lawful access to the national security material "permits the same to be removed from its proper place of custody or delivered to anyone in violation of his trust, or to be lost, stolen, abstracted, or destroyed," and fails to disclose it having knowledge of the breach.

lviii. 18 U.S. Code 793

lix. United States v. Rosen, 447 F. Supp. 2d 538 (E.D. Va. 2006)

sisted that §793(e) "'requires the government to demonstrate the likelihood of the defendant's bad faith purpose to either harm the United States or to aid a foreign government."' Thus, while Daniel Ellsberg, Samuel Morison, Lawrence Franklin, Thomas Drake, Chelsea Manning, Jeffrey Sterling, and James Hitselberger have all been charged with violating §793(e), no news organizations associated with those leaks have been prosecuted.

In addition to the high mens rea requirement for §793, there is a thought that the law is too unclear for the government to bring charges. "The Espionage Act is so vague and poorly defined in its terms, that it's hard to say exactly what it does and does not cover," said Steven Aftergood, director of the Project on Government Secrecy for the Federation of American Scientists.[133] This ambiguity in the law would make any prosecution under the Act stand on very shaky ground. An unsuccessful prosecution against a journalist would likely threaten its efficacy at deterring leakers. While Justice Harlan called the Act "singularly opaque" and other academics have called it "incomprehensible," the likelihood that a claim would have been brought against a news organization was very limited—that is, at least, until the most recent presidential election.[134]

Recently, the consideration of charges brought against a news organization is not unthinkable. While the Justice Department under President Barack Obama actively decided not to charge WikiLeaks for revealing government information because it was too similar to a news organization,[135] the current Justice Department has indicated to prosecutors that it is reconsidering the case. *The Washington Post* reported last year that the administration was drafting a memo considering the possibility of a prosecution.[136] While scholars like Professor Stone believe that prosecuting journalists would violate the First Amendment, there is no way of knowing what is in store for the future.

Conclusion

During the composition of this report, threats toward the press have continued. In June 2018, Reality Winner pleaded guilty under the Espionage Act after being the first person prosecuted by the Trump administration on charges of leaking classified information. The same month, *New York Times* reporter Ali Watkins became the focus of a lurid investigation by her own paper into her relationship with Senate staffer James A. Wolfe, who was arrested on charges not of distributing classified information, but of lying to investigators. Watkins's telephone and email records were seized as part of the investigation, and Customs and Border Patrol Officer Jeffrey Rambo, who gained illicit access to documents pertaining to the investigation, accosted Watkins and tried to coerce her into informing him of the details of her colleagues work, including their sources. More recently, President Trump has repeatedly stated that a large percentage of the media is the "enemy of the people." These concerns are not limited to public officials. Elon Musk, CEO of Tesla, made disparaging comments about reporters at The Center for Investigative Reporting after the newsroom published a series of stories about his company, and said he would create a new site called "Pravda" where journalists could be rated according to truthfulness.

In many cases, President Trump's penchant for flamboyant rhetoric about the mainstream media has been cited as either tacitly approved-of or emboldening the public against the press, including Jarrod Ramos, who killed five people when he attacked the newsroom of *The Capital Gazette* in Annapolis, Maryland, with a shotgun.

Trump's tweets and speeches regularly deride established news organizations like *The New York Times*, *The Washington Post*, and especially CNN as "fake news." His language around leaks, which he has called the work of "traitors," is particularly troubling.[137]

This stance toward leaks has been a feature of Trump's administration since his first days in office. Soon after the inauguration, then-Press Secretary Sean Spicer made White House aides leave their mobile phones on a table to undergo a "phone check" in an attempt to ensure that the staffers

were not speaking to reporters. CNN reported that Spicer had the sign-off from the president before initiating these checks.

President Trump's contempt for the essentials of free expression came to further light in May 2017 after FBI Director James Comey leaked memos describing his encounters with the president to the press through Columbia Law Professor Daniel Richman. The memos revealed that during a February 2017 conversation, the president told the ex-FBI director to consider putting reporters in prison for publishing classified information and suggested approvingly that they might be raped there.

In May 2017, when *The Washington Post* reported on the president's son-in-law and senior advisor, Jared Kushner, President Trump sent three tweets that read:

> It is my opinion that many of the leaks coming out of the White House are fabricated lies made up by the #FakeNews media.
>
> Whenever you see the words 'sources say' in the fake news media, and they don't mention names . . .
>
> . . . it is very possible that those sources don't exist but are made up by fake news writers. #FakeNews is the enemy!

Two months later, in response to a story about Attorney General Jeff Sessions, President Trump tweeted, "These illegal leaks, like Comey's, must stop!" Finally, in January 2018, a White House memo by Chief of Staff John Kelly banned the use of personal electronic devices, including mobile phones and smart watches, by White House staff to "to protect White House information technology infrastructure from compromise and sensitive or classified information from unauthorized access or dissemination." According to one unnamed senior White House official, the new policy is "designed to prevent leaks and ensure the productivity of the people who work here."[138]

Within 10 days of the president suggesting that his attorney general was weak on leaks, Sessions announced there were currently "three times as many leak investigations as were open at the end of the Obama era" in a news conference held the first week of August, a move that the president commended.[139]

In April 2017, *The Washington Post* reported that federal prosecutors are contemplating bringing criminal charges against WikiLeaks for the 2010 leaks to the organization by Chelsea Manning, and that arresting Assange was a priority.[140] That same month, then-CIA Director Mike Pompeo called WikiLeaks "a non-state hostile intelligence service";[141] by August of last year, the Senate had tried to insert the exact same language about the anti-secrecy organization into a boilerplate appropriations bill.

Continuously lambasting the press and its sources inflicts damage in at least two ways: First, imposing restrictions and inculcating fear among government workers, including with a lockdown of devices used inside the White House, can stymie speech, especially as the growing number of communications enabled by technology are increasingly kept secret by easy deletion, end-to-end encrypted messaging, and new policies requiring stricter secrecy.

Second, framing leaks as traitorous and journalists as the "enemy" has serious consequences for our democratic order. As RonNell Anderson Jones and Lisa Sun wrote in a recent article about the subject, "Undercutting the watchdog, educator, and proxy functions of the press through enemy construction leaves the administration more capable of delegitimizing other institutions and constructing other enemies—including the judiciary, the intelligence community, immigrants, and members of certain races or religions—because the viability and traction of counter-narrative is so greatly diminished."[142]

Lastly, as data becomes ever more difficult to collect, review, and distribute, framing the press as the enemy undermines even the tenuous grip the public can be expected to have in the current, confusing environment. Labeling news "fake" in this oversaturated and cacophonous atmosphere undercuts the ability of the press to investigate, clarify, and report in ways that add to our understanding of truth.

Citations

1. Alexander Meiklejohn, "The First Amendment Is an Absolute," *The Supreme Court Review* (1961): 255–57.

2. Thomas I. Emerson, "Legal Foundations of the Right to Know," *Washington University Law Quarterly* (1976): 1–24.

3. Ibid.

4. Ibid.

5. Mikal Khoso, "How Much Data Is Produced Every Day?" Northeastern University "Level" blog, May 13, 2016, http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/.

6. Bernard Marr, "Big Data: 20 Mind-Boggling Facts Everyone Must Read," *Forbes*, September 30, 2015, https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#764a266b17b1.

7. Nael Shiab, "Web Scraping: A Journalist's Guide," Global Investigative Journalism Network, August 11, 2015, https://gijn.org/2015/08/11/web-scraping-a-journalists-guide/.

8. Simon Rogers, "The State of Data Journalism in 2017," Google News Initiative, September 18, 2017, https://www.blog.google/topics/journalism-news/data-journalism-2017/.

9. Paul Bradshaw, *Scraping for Journalists (2nd Edition)* (British Columbia: LeanPub, 2017), 462.

10. Rogers, "The State of Data Journalism in 2017."

11. Shiab, "Web Scraping: A Journalist's Guide."

12. Python (programming language), Wikipedia, https://en.wikipedia.org/wiki/Python_(programming_language).

13. Paul Bradshaw, *Scraping for Journalists (2nd Edition)*.

14. Shelly Tan, "Five Data Scraping Tools for Would-Be Data Journalists," Knight Lab, March 20, 2014, https://knightlab.northwestern.edu/2014/03/20/five-data-scraping-tools-for-would-be-data-journalists/.

15. Sophie Chou, "To Scrape or Not to Scrape: Technical and Ethical Challenges of Collecting Data Off the Web," Storybench, April 2016, http://www.storybench.org/to-scrape-or-not-to-scrape-the-technical-and-ethical-challenges-of-collecting-data-off-the-web/.

16. Julia Angwin and Surya Mattu, "Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn't," ProPublica, September 2016, https://www.propublica.org/article/amazon-says-it-puts-customers-first-but-its-pricing-algorithm-doesnt.

17. Carrie Teegardin and Danny Robbins, "Still Forgiven," *The Atlanta Journal-Constitution*, 2017, http://doctors.ajc.com/still_forgiven/?ecmp=doctorssexabuse_microsite_nav.

18. Jeff Larson, Julia Angwin, and Jennifer Valentino-DeVries, "How We Are Monitoring Political Ads on Facebook," ProPublica, December 5, 2017, https://www.propublica.org/article/how-we-are-monitoring-political-ads-on-facebook.

19. Jeff Larson, Rob Weychert, Julia Angwin, and Jennifer Valentino-DeVries, "Political Ads on Facebook," ProPublica, December 4, 2017, https://projects.propublica.org/facebook-ads/.

20. Ibid.

21. ProPublica, "Facebook Political Ad Collector," Github, https://github.com/propublica/facebook-political-ads.

22. Lauren Glass, "Mastering Python Web Scraping: Get Your Data Back," Hackernoon, September 12, 2017, https://hackernoon.com/mastering-python-web-scraping-get-your-data-back-e9a5cc653d88.

23. Benjamin Edelman, Michael Luca, and Dan Svirsky, "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," Harvard Business School, September 4, 2016, https://www.hbs.edu/faculty/Publication%20Files/16-069_5c3b2b36-d9f8-4b38-9639-2175aaf9ebc9.pdf.

24. Ibid.

25. Simon Alterman, "Web Scraping: Legal Issues and Best Practices for Publishers," Outsell, December 9, 2016, http://docplayer.net/60279697-Web-scraping-legal-issues-and-best-practices-for-publishers.html.

26. Ibid.

27. Sidney Fussell, "AI Professor Details Real-World Dangers of Algorithm Bias [Corrected]," Gizmodo, December 8, 2017, https://gizmodo.com/microsoft-researcher-details-real-world-dangers-of-algo-1821129334.

28. Will Knight, "Biased Algorithms Are Everywhere, and No One Seems to Care," MIT Technology Review, July 12, 2017, https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/.

29. Orin S. Kerr, "Norms of Computer Trespass," *Columbia Law Review* 116, no. 4 (2016): 1143–1163, https://columbialawreview.org/content/norms-of-computer-trespass/.

30. United States v. Andrew Auernheimer, Electronic Frontier Foundation, https://www.eff.org/cases/us-v-auernheimer.

31. Matt Brian, "Andrew 'Weev' Auernheimer Sentenced to 41 Months for Exploiting AT&T IPad Security Flaw," The Verge, March 18, 2013, https://www.theverge.com/2013/3/18/4118484/andrew-weev-auernheimer-sentenced-att-ipad-hack.

32. Ryan Tate, "Apple's Worst Security Breach: 114,000 IPad Owners Exposed," Gawker, June 9, 2010, http://gawker.com/5559346/apples-worst-security-breach-114000-ipad-owners-exposed.

33. Jonathan D. Lee, "Re: Cease and Desist/Mitigation of Harm/Preservation of

Evidence Request," April 30, 2013, https://www.documentcloud.org/documents/701519-response-from-jonathan-lee.html.

34. United States of America, Plaintiff-Appellee v. David Nosal, Defendant-Appellant, July 5, 2016, https://cdn.ca9.uscourts.gov/datastore/opinions/2016/07/05/14-10037.pdf.

35. Orin S. Kerr, "Cybercrime's Scope: Interpreting 'Access' and 'Authorization' in Computer Misuse Statutes," *New York University Law Review* 78, no. 1596 (2003), https://ssrn.com/abstract=399740.

36. Shepard Goldfein and James Keyte, "Big Data, Web Scraping and Competition Law: The Debate Continues," *New York Law Journal* 258 (2017).

37. SIGCHI Research Ethics, "Do Researchers Need to Follow TOS?" Medium, November 30, 2017, https://medium.com/sigchi-ethics-committee/do-researchers-need-to-follow-tos-f3bde1950d3c.

38. Caroline O'Donovan, "Hacking in the Newsroom? What Journalists Should Know About the Computer Fraud and Abuse Act," Nieman Lab, March 3, 2014, http://www.niemanlab.org/2014/03/hacking-in-the-newsroom-what-journalists-should-know-about-the-computer-fraud-and-abuse-act/.

39. Electronic Frontier Foundation, "Computer Fraud and Abuse Act Reform," https://www.eff.org/issues/cfaa.

40. Electronic Frontier Foundation, "Nosal v. U.S. Cert Petition," June 5, 2017, https://www.eff.org/document/nosal-v-us-cert-petition.

41. The National Security Archive, "FOIA in the News—2004–2006," 2017, https://nsarchive2.gwu.edu/nsa/foia/stories.htm.

42. Jason Leopold, "Gitmo Spent $300,000 on Liquid Supplements While Denying a 'Mass Hunger Strike'," Vice, August 18, 2014, https://news.vice.com/article/gitmo-spent-300000-on-liquid-supplements-while-denying-a-mass-hunger-strike.

43. Jason Leopold and David Mack, "A Bunch of CIA Contractors Got Fired for Stealing Snacks from Vending Machines," BuzzFeed, June 21, 2017, https://www.buzzfeed.com/jasonleopold/cia-vending-thefts?utm_term=.gfpJAzADN#.gfpJAzADN.

44. Ted Bridis, "Obama's Final Year: US Spent $36 Million In Records Lawsuits," March 14, 2017, https://apnews.com/0b27c4d4b23b436d805328694e58c605/obamas-final-year-us-spent-36-million-records-lawsuits.

45. Max Galka, "Who Uses FOIA?—An Analysis of 229,000 Requests to 85 Government Agencies," foiamapper.com, March 13, 2017, https://foiamapper.com/who-uses-foia/.

46. Ibid.

47. MuckRock, "FBI Vexsome Filer List (2016)," January 20, 2016, https://www.muckrock.com/foi/united-states-of-america-10/fbi-vexsome-filer-list-2016-23522/#file-123991.

48. Joseph Cox, "An Activist Wrote a Script to FOIA the Files of 7,000 Dead FBI Officials," Motherboard, 2016, https://motherboard.vice.com/en_us/article/kb7w5w/an-activist-michael-best-wrote-a-script-to-foia-the-files-of-7000-dead-fbi-officials.

49. David E. Pozen, "Freedom of Information beyond the Freedom of Information Act," *University of Pennsylvania Law Review* (2017): 1097–1158.

50. Adam M. Samaha, "Government Secrets, Constitutional Law, and Platforms for Judicial Intervention," *University of California Law Review* (2006): 909.

51. Mark Zaid, https://markzaid.com.

52. National Archives, https://www.archives.gov.

53. Greg Munno, "Key Agencies Flub Response to Simple Request," The FOIA Project, April 24, 2015, http://foiaproject.org/2015/04/24/agencies-rated/.

54. "Testimony of David McCraw on Behalf of The New York Times Company before the Committee on Oversight and Government Reform," United States House of Representatives on the Freedom of Information Act, June 2015, https://oversight.house.gov/wp-content/uploads/2015/06/McCraw-NYT-Statement-6-2-FOIA.pdf.

55. Chris Mooney and Juliet Eilperin, "EPA Website Removes Climate Science Site from Public View After Two Decades," *The Washington Post*, April 29, 2017, https://www.washingtonpost.com/news/energy-environment/wp/2017/04/28/epa-website-removes-climate-science-site-from-public-view-after-two-decades/?noredirect=on&utm_term=.e9b96e54efe5.

56. First Amendment Watch, "The Limits of Transparency and FOIA Under Trump," March 14, 2018, http://firstamendmentwatch.org/2018/03/14/limits-transparency-foia-trump/.

57. Ted Bridis, "US Sets New Record for Censoring, Withholding Gov't Files," Associated Press, March 12, 2018, https://apnews.com/714791d91d7944e49a284a51fab65b85.

58. MuckRock, "FOIA the Trump Administration," https://www.muckrock.com/project/foia-the-trump-administration-85/.

59. MuckRock, "Join the MuckRock Slack Channel," https://www.muckrock.com/slack/.

60. Jonathan Peters, "'Always Appeal,' and More Pro Tips from a Dozen FOIA Experts," CJR, March 17, 2017, https://www.cjr.org/united_states_project/foia-expert-pro-tips-sunshine-week.php.

61. Emily Holden, "Anti-Secrecy Lawsuits Soaring against Pruitt's EPA," Politico, February 26, 2018, https://www.politico.com/story/2018/02/26/pruitt-epa-secrecy-lawsuits-environment-355842.

62. Gregory L. Waples, "The Freedom of Information Act: A Seven-Year Assessment," *Columbia Law Review* 74, no. 895 (1974): 958.

63. Mark Fenster, "The Opacity of Transparency," *Iowa Law Review* 91, no. 885 (2006): 894.

64. Mark Tapscott, "The Right Move on Freedom of Information," The Heritage Foundation, April 16, 2005, https://www.heritage.org/civil-society/commentary/the-right-move-freedom-information.

65. Fenster, "The Opacity of Transparency."

66. Office of Information Policy, "Department of Justice Guide to the Freedom of Information Act," Department of Justice, 2009, https://www.justice.gov/oip/doj-guide-freedom-information-act-0.

67. Margaret Kwoka, "FOIA, Inc.," *Duke Law Journal* 65, no. 1361 (2016): 1376–1381, https://dlj.law.duke.edu/article/foia-inc-kwoka-vol65-iss7/.

68. Barack Obama, "Remarks by the President at the National Defense University," https://obamawhitehouse.archives.gov/the-press-office/2013/05/23/remarks-president-national-defense-university.

69. Office of Government Information Services, "Understanding FOIA Delay," National Archives, October 4, 2017, https://foia.blogs.archives.gov/2017/10/04/understanding-foia-delay/.

70. Jonathan Peters, "'Always Appeal,' and More Pro Tips from a Dozen FOIA Experts."

71. Office of Information Policy, "Department of Justice Guide to the Freedom of Information Act," Department of Justice, 2009, https://www.justice.gov/oip/doj-guide-freedom-information-act-0.

72. Office of Information Policy, "Summary of Annual FOIA Reports for Fiscal Year 2014," Department of Justice, June 2015, https://www.justice.gov/sites/default/files/oip/pages/attachments/2015/05/01/fy_2014_annual_report_summary.pdf.

73. Jim Snyder and Danielle Ivory, "Obama Cabinet Flunks Disclosure Test with 19 in 20 Ignoring Law," Bloomberg Businessweek, Septtember 27, 2012, http://www.bloomberg.com/news/articles/2012-09-28/obama-cabinet-flunks-disclosure-test-with-19-in-20-ignoring-law.

74. Office of Information Policy, "Summary of Annual FOIA Reports for Fiscal Year 2014," Department of Justice, June 2015, https://www.justice.gov/sites/default/files/oip/pages/attachments/2015/05/01/fy_2014_annual_report_summary.pdf.

75. Ibid.

76. Ibid.

77. Trevor Timm, "We Just Sued the Justice Department Over the FBI's Secret Rules For Using National Security Letters on Journalists," Freedom of the Press Foundation, July 30, 2015, https://freedom.press/news-advocacy/we-just-sued-the-justice-department-over-the-fbias-secret-rules-for-using-national-security-letters-on-journalists/.

78. Trevor Timm, "Dozens of News Orgs Demand DOJ Release Its Secret Rules for Targeting Journalists with National Security Letters," Freedom of the Press Foundation, June 15, 2016, https://freedom.press/news-advocacy/dozens-of-news-orgs-demand-doj-release-its-secret-rules-for-targeting-journalists-with-national-security-letters/.

79. Cora Currier, "Secret Rules Make It Pretty Easy for the FBI to Spy on Journalists," The Intercept, June 30, 2016, https://theintercept.com/2016/06/30/secret-rules-make-it-pretty-easy-for-the-fbi-to-spy-on-journalists/.

80. Sam Thielman, "Justice Department 'Uses Aged Computer System To Frustrate FOIA Requests'," *The Guardian*, July 16, 2016, https://www.theguardian.com/politics/2016/jul/16/justice-department-freedom-of-information-computer-system.

81. Mya Frazier, "Big Tech's Bid to Control FOIA," CJR, February 2, 2018, https://www.cjr.org/business_of_news/facebook-amazon-foia.php.

82. Pozen, "Freedom of Information beyond the Freedom of Information Act."

83. Louis D. Brandeis, *Other People's Money and How the Bankers Use It* (New York: Frederick A. Stokes Company Publishers, 1914).

84. Office of Information Policy, "FOIA Update: Department Of Justice Report on 'Electronic Record' FOIA Issues, Part I," Department of Justice, January 1, 1990, https://www.justice.gov/oip/blog/foia-update-department-justice-report-electronic-record-foia-issues-part-i#N_1_.

85. **sander**.

86. Andrew Donohue, "Why We're Suing the Department of Justice for Gun Records," Reveal, November 25, 2017, https://www.revealnews.org/blog/why-were-suing-the-department-of-justice-for-gun-records/.

87. Ibid.

88. "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 1, 2014, https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

89. Ibid.

90. Ibid.

91. James Risen and Eric Lichtblau, "Bush Lets U.S. Spy on Callers without Courts," *The New York Times*, December 16, 2005, https://www.nytimes.com/2005/12/16/politics/bush-lets-us-spy-on-callers-without-courts.html.

92. David Johnston Scott Shane and James Risen, "Secret U.S. Endorsement of Severe Interrogations," *The New York Times*, October 4, 2007, https://www.nytimes.com/2007/10/04/washington/04interrogate.html.

93. Glenn Greenwald, "Obama DOJ Formally Accuses Journalist in Leak Case of Committing Crimes," *The Guardian*, May 20, 2013, https://www.theguardian.com/commentisfree/2013/may/20/obama-doj-james-rosen-criminality.

94. Betsy Woodruff, "Leak Investigations Rise 800% Under Jeff Sessions,"

The Daily Beast, November 14, 2017, https://www.thedailybeast.com/leak-investigations-rise-800-under-jeff-sessions.

95. Callum Borchers, "Sessions Says He Can't 'Make a Blanket Commitment' Not to Jail Journalists," *The Washington Post*, October 18, 2017, https://www.washingtonpost.com/news/the-fix/wp/2017/10/18/sessions-says-he-cant-make-a-blanket-commitment-not-to-jail-journalists/?utm_term=.75e4716665dd.

96. Ibid.

97. U.S. Press Freedom Tracker, https://pressfreedomtracker.us/.

98. Kit Roane, *Lies, Leaks, and Consequences: How Toughness on Leaks and a Free Press Link Trump, Obama, and Nixon*, https://www.retroreport.org/video/lies-leaks-and-consequences/.

99. Alexander M. Bickel, *The Morality of Consent* (New Haven and London: Yale University Press, 1975).

100. Ibid.

101. PBS Fronline, "Washington's Culture of Secrets, Sources and Leaks," February 13, 2007, https://www.pbs.org/wgbh/pages/frontline/newswar/part1/frankel.html.

102. Charlie Savage and Alan Blinder, "Reality Winner, N.S.A. Contractor Accused in Leak, Pleads Guilty," *The New York Times*, June 26, 2018, https://www.nytimes.com/2018/06/26/us/reality-winner-nsa-leak-guilty-plea.html.

103. "Testimony of David McCraw on Behalf of The New York Times Company before the Committee on Oversight and Government Reform."

104. Spencer Ackerman and Ed Pilkington, "Obama's War on Whistleblowers Leaves Administration Insiders Unscathed," *The Guardian*, March 16, 2015, https://www.theguardian.com/us-news/2015/mar/16/whistleblowers-double-standard-obama-david-petraeus-chelsea-manning.

105. Charlie Savage and Leslie Kaufman, "Phone Records of Journalists Seized by U.S.," *The New York Times*, May 13, 2013, https://www.nytimes.com/2013/05/14/us/phone-records-of-journalists-of-the-associated-press-seized-by-us.html.

106. Mark Sherman, "Gov't Obtains Wide AP Phone Records in Probe," Associated Press, May 13, 2013, https://www.ap.org/ap-in-the-news/2013/govt-obtains-wide-ap-phone-records-in-probe.

107. Ibid.

108. Ann E. Marimow, "Fox Calls DOJ Investigation of Reporter 'Downright Chilling'," *The Washington Post*, May 20, 2013, https://www.washingtonpost.com/local/fox-calls-doj-investigation-of-reporter-downright-chilling/2013/05/20/1236692e-c17f-11e2-bfdb-3886a561c1ff_story.html.

109. Obama, "Remarks by the President at the National Defense University."

110. "Letter from Reporters Committee to Attorney General Holder," The Reporters Committee for Freedom of the Press, June 21, 2013, http://rcfp.org/x?TpCu.

111. Eric Holder, "Statement of Attorney General Eric Holder on the Justice Department Report on Revised Media Guidelines," Office of Public Affairs, Department of Justice, http://1.usa.gov/1P1g5SJ.
112. Ibid.
113. Reporters Committee for Freedom of the Press, *Summary of Changes to the Attorney General Guidelines in January 2015*, 2015, https://www.rcfp.org/sites/default/files/docs/attorney_gen_guidelines_changes_2015.pdf.
114. Reporters Committee for Freedom of the Press and Karen Kaiser, "Summary of Changes to the Attorney General Guidelines," July 30, 2013, https://www.rcfp.org/sites/default/files/docs/RCFP%20summary%20of%20AG%20Report-1.pdf.
115. Office of the Inspector General, "A Review of the FBI's Use of National Security Letters: Assessment of Progress in Implementing Recommendations and Examination of Use in 2007 through 2009," Department of Justice, https://oig.justice.gov/reports/2014/s1408.pdf.
116. Barton Gellman and Ellen Nakashima, "U.S. Spy Agencies Mounted 231 Offensive Cyber-Operations in 2011, Documents Show," *The Washington Post*, August 30, 2013, https://www.washingtonpost.com/world/national-security/us-spy-agencies-mounted-231-offensive-cyber-operations-in-2011-documents-show/2013/08/30/d090a6ae-119e-11e3-b4cb-fd7ce041d814_story.html?utm_term=.19292e23759b.
117. "National Security Letters: FAQ," Electronic Frontier Foundation, https://www.eff.org/issues/national-security-letters/faq.
118. Cora Currier, "Secret Rules Make It Pretty Easy for the FBI to Spy on Journalists."
119. Ibid.
120. Brian Stelter, "Marty Baron on The Washington Post in the Trump Age," *Reliable Sources*, https://podcasts.cnn.net/embed/single/skin/erfysl/marty-baron-on-the.html?iid=EL.
121. "Criminal Prohibitions on Leaks and Other Disclosures of Classified Defense Information," Congressional Research Service, March 7, 2017, https://www.everycrsreport.com/files/20170307_R41404_d37dd9d3fd8e8a1f382b574e4b5681ef102f1ecf.pdf.
122. David E. Pozen, "The Leaky Leviathan: Why Government Condemns and Condones Unlawful Disclosures Of Information," *Harvard Law Review* 127, no. 512 (December 20, 2013), https://harvardlawreview.org/wp-content/uploads/pdfs/vol127_pozen.pdf.
123. New York Times Co. v. United States, Cornell Law School, Legal Information Institute, June 30, 1971, https://www.law.cornell.edu/supremecourt/text/403/713.
124. Bartnicki et al. v. Vopper, aka Williams, et al. Certiorari to the United

States Court of Appeals for the Third Circuit, Cornell Law School, Legal Information Institute, May 21, 2001, https://www.law.cornell.edu/supct/html/99-1687.ZS.html.
125. Ronnell Andersen Jones and Sonja R. West, "Don't Expect the First Amendment to Protect the Media," *The New York Times*, January 25, 2017, https://www.nytimes.com/2017/01/25/opinion/dont-expect-the-first-amendment-to-protect-the-media.html.
126. New York Times Co. v. United States, Cornell Law School, Legal Information Institute, June 30, 1971, https://www.law.cornell.edu/supremecourt/text/403/713.
127. Bartnicki et al. v. Vopper, aka Williams, et al. Certiorari to the United States Court of Appeals for the Third Circuit,
128. Bartnicki et al. v. Vopper, aka Williams, et al. Certiorari to the United States Court of Appeals for the Third Circuit, Cornell Law School, Legal Information Institute, May 21, 2001, https://www.law.cornell.edu/supct/html/99-1687.ZS.html.
129. David Greene, "As the Espionage Act Turns 100, We Condemn Threats against Wikileaks," Electronic Frontier Foundation, June 14, 2017, https://www.eff.org/deeplinks/2017/06/espionage-act-turns-100-we-condemn-threats-against-wikileaks.
130. "International Media Law Bulletin," Media Law Research Center, October 2017, http://www.medialaw.org/publications/mlrc-bulletin.
131. Definition of "unauthorized person", Cornell Law School, https://www.law.cornell.edu/definitions/uscode.php?width=840&height=800&iframe=true&def_id=18-USC-1618243585-317968083&term_occur=1&term_src=title:18:part:I:chapter:37:section:798.
132. New York Times Co. v. United States, Cornell Law School, Legal Information Institute, June 30, 1971, https://www.law.cornell.edu/supremecourt/text/403/713.
133. Emily Peterson, "WikiLeaks and the Espionage Act of 1917," Reporters Committee for the Freedom of the Press, https://www.rcfp.org/browse-media-law-resources/news-media-law/wikileaks-and-espionage-act-1917.
134. Edgar Harold and Benno C. Schmidt Jr., "Curtiss-Wright Comes Home: Executive Power and National Security Secrecy," *Harvard Civil Rights-Civil Liberties Law Review* 21, no. 2 (1986): 349–408, https://heinonline.org/HOL/LandingPage?handle=hein.journals/hcrcl21&div=14&id=&page=.
135. Sari Horwitz, "Julian Assange Unlikely to Face U.S. Charges Over Publishing Classified Documents," *The Washington Post*, November 25, 2013, https://www.washingtonpost.com/world/national-security/julian-assange-unlikely-to-face-us-charges-over-publishing-classified-documents/2013/11/25/dd27decc-55f1-11e3-8304-caf30787c0a9_story.html?utm_term=.eec423bb8f31.

136. Matt Zapotosky and Ellen Nakashima, "Justice Dept. Debating Charges against WikiLeaks Members in Revelations of Diplomatic, CIA Materials," *The Washington Post*, April 20, 2017, https://www.washingtonpost.com/world/national-security/justice-dept-debating-charges-against-wikileaks-members-in-revelations-of-diplomatic-cia-materials/2017/04/20/32b15336-2548-11e7-a1b3-faff0034e2de_story.html?noredirect=on&utm_term=.fc81f4a8813f.
137. Rebecca Morin, "Trump Says Leaks Are Exaggerated but Vows to Track Down 'Traitors'," Politico, May 14, 2018, https://www.politico.com/story/2018/05/14/trump-leaks-traitors-white-house-586041.
138. Katherine Faulders, "White House Memo: Kelly Bans Personal Phones, Smart Watches from West Wing," ABC News, January 10, 2018, https://abcnews.go.com/Politics/white-house-memo-kelly-orders-west-wing-staff/story?id=52264274.
139. Charlie Savage and Eileen Sullivan, "Leak Investigations Triple under Trump, Sessions Says," *The New York Times*, August 4, 2017, https://www.nytimes.com/2017/08/04/us/politics/jeff-sessions-trump-leaks-attorney-general.html.
140. Matt Zapotosky and Ellen Nakashima, "Justice Dept. Debating Charges against WikiLeaks Members in Revelations of Diplomatic, CIA Materials."
141. "Director Pompeo Delivers Remarks at CSIS," Central Intelligence Agency, April 13, 2017, https://www.cia.gov/news-information/speeches-testimony/2017-speeches-testimony/pompeo-delivers-remarks-at-csis.html.
142. RonNell Andersen Jones and Lisa Grow Sun, "Enemy Construction and the Press," BYU Law Research paper, March 8, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2929708.