'Information Pressures' and the Facebook Files:
Navigating questions around leaked platform data

Author: Dr. Rebekah Larsen
Postdoctoral Fellow, Oslo Metropolitan University, Norway
ORCID: 0000-0002-0884-6217
rebekahl@oslomet.no

ABSTRACT

Recently leaked Facebook data included material from the platform's internal social science research units. Given its character—and the likelihood of future whistleblowing—this case presents an opportunity for reflection on current and future academic engagement with leaked data. This contribution is grounded in the experiences of some members of a small, international team of journalism scholars, after being offered access to the leaked data by a media organization. It also describes other academic-involved initiatives—their similarities, differences, and efforts—working toward expanded access to and/or usage of this data. Rather than take a general or hardline stance on use of leaked data, this contribution emphasizes critical engagement—ethical, legal, practical—with the particulars of individual leaks. It also emphasizes the need for guidelines and mechanisms for working with leaked data, and the concomitant importance of transparency within academia about current practices. Such critical engagement should also include consideration of the positionalities, divisions, and marginalities of involved researchers and regions.

Introduction

In autumn 2021, the *Wall Street Journal* began publishing the Facebook Files (FF). This series, comprised of a limited run of articles and high-traffic podcasts, was curated from material revealed by an ex-Facebook employee. Frances Haugen was previously a member of the Facebook Civic Integrity unit, responsible for preventing election misinformation on the platform. This unit was created in the wake of a parade of PR disasters for the social media platform, including revelations about widespread Russian disinformation campaigns. The unit was essentially dissolved right after the 2020 US presidential election, implying to some that Facebook (the company now rebranded as Meta) considered election misinformation problems addressed (Edgerton, McLaughlin, and Forden 2021). (Tellingly, this was before the 6 January 'Stop the Steal' capitol riots in the US). Working with the Wall Street Journal (WSJ), Haugen revealed a trove of documents including research and discussions from Facebook's in-house social science group, and interactions on the company-wide, internal Facebook feed-esque platform (Vaidhyanathan 2021). The leaked data was also described as containing confidential memos, chat logs, and slide decks for presentation to senior management. According to journalists, "The documents offer perhaps the clearest picture thus far of how broadly Facebook's problems are known inside the company, up to the chief executive himself" (*Wall Street Journal* 2021).

The orchestrated release of the leaked data for maximal impact[1]—first via the WSJ and to US Congress, and then in batches to an uneasy consortium of news outlets (Kantrowitz 2021; Ingram 2021)—prolonged the duration of the leak in news cycles. It has also resulted in the data—all the thousands of somewhat-redacted pages—being widely distributed among journalist organizations, from tech blogs to legacy media. But at the time of this writing and despite the nature of the data, academics still do not have similar access. This leaked Facebook data and its distribution—its creation in a social science-related arena, its privacy implications, the manner of its release—invite reflection on how social sciences can, do, and should approach leaked platform data.

This commentary is a grounded exploration of some researchers' perspectives and experiences in navigating this data of uncomfortable provenance but potential public interest. It is also informed by calls within the academy, such as the one issued by Savage and Burrows (2007) around the 'Coming Crisis of Empirical Sociology', for a 'radical mixture' of newly critical reflection and approaches: "Rather than seeking refuge in our own, internal debates, this involves casting our net wide, critically engaging with the extensive data sources which now exist…" (890). Such 'critical engagement' must entail consideration of inevitable tradeoffs—for immediately involved actors and arguably non-consenting data subjects, but also generally for academia and its positionality (Donovan 2020). Wrangling with the problems surrounding leaked data is not a new predicament, but one that must be confronted more often—and with the specifics and 'criteriality' of each case in mind (Darnton 2021).

Background

In 2007, Savage and Burrows warned of a 'coming'[2] empirical crisis for social scientists, spearheaded by what could be described as *knowing capitalism*—or the analogous, trendier *surveillance capitalism* (Zuboff 2019)*.* One focus of their essay is the increasing marginalization of academic social scientists as information technology companies create, wall off, mine, and manipulate increasingly central sites—and

---

[1] This impact has also been furthered via ongoing global tours by whistleblower Frances Haugen, notably trained in public speaking and working with NGO Whistleblower Aid (Satariano and Isaac 2021; Nobel Peace Center 2022).

[2] This crisis has arguably now arrived, according to the authors, via the rise of 'big data' (Burrows and Savage 2014).

insights—of sociality for profit (Savage and Burrows 2007). They describe this marginalization in terms of access, but also influence in concepts, frameworks, and aims. The industrial social sciences can be described as 'parallel' to academia but missing the "ethical constraints which condition the work of academic researchers"; moreover the "glorious flourishes of the cultural turn, do not—with a few exceptions—speak to the workaday needs and interests" of Facebook and other companies (887 – 888). This commentary thus functions as a call for reflexive response to this marginalization—while considering implications for the public and inequalities in knowledge production.

For example, academics could contribute, "in an era when journalism is retreating from detailed social investigation", by creating information subsidies for better public framing, policies with "specific specialized expertise" (895) and attention toward ongoing ethical issues such as consent in the digital age (Donovan 2020). This potentiality is exemplified by the Facebook leak, where much of the data was collected and analyzed within the academically parallel but closed worlds of 'commercial sociology' or the industrial social sciences (Savage and Burrows 2007; Baritz 1960). The type and origin of this data thus contrasts with that from other high-profile leaks over the last decade. (These include the Panama Papers, the leaked records of 214,000 offshore companies (Obermaier and Obermayer 2017), and the Wikileaks releases over the last decade-plus, allegedly from public sector sources.) Social scientists with specialized, various expertise are arguably well-positioned—and should be able—to contribute to evaluation and contextualization for public knowledge (Calhoun 2009). Media researchers—particularly journalism scholars—are uniquely positioned as 'knowledge brokers'; they have potential expertise in the material being leaked, in journalistic and policy-making structures, and in environments through which the material is being sorted and framed (Weber 2021; Wahl-Jorgensen 2014). Journalism studies is populated with many 'hackademics', or former or side-practicing journalists (Harcup 2011). In addition to their subject expertise, they are arguably positioned to best navigate difficult access and conversations with the press. This would allow both types of actors to benefit from the expertise and resources of the other with minimal expenditure of time, credibility, and other resources.

Even in the narrow context of timely reporting around this case, the impacts of ongoing lack of access are significant. One social scientist, acting as a journalistic resource for a particular outlet, likened the process to a game of '20 Questions'. Journalists would ask this academic to verify claims in the leaked data, sight unseen—which the researcher was unable to do, without access. The frustration was no doubt felt on both sides. During a podcast interview in 2021, Jeff Horwitz (the WSJ journalist who broke the Facebook Files stories), exclaimed, "Dear god, these are people with PhDs asking for access. Let's get them the documents. They shouldn't be waiting for our leftovers" (Jong-Fast and Levy 2021). Calls for intergroup watchdogs, even in the present, exclude academia despite its offerings (Esser and Neuberger 2019); this most recent leak exemplifies positionally what academia could potentially offer to a larger project of 'realizing democratic functions' in the modern media environment—if we are able to recognize and grapple with internal and external barriers.

The Facebook leak also represents a unique opportunity for insight into a social media platform that has been reticent at best and misleading at worst in how they have been sharing data. Academics have been understandably skeptical of the research data and findings provided by Facebook in the past. Just weeks before the Facebook Files revelations, stories broke in the press about the significantly flawed datasets Facebook had shared with academics—and potential retraction of much published work—via the Social Science One initiative (Alba 2022; Timberg 2021). Moreover, what has been reported about the leak shows that Facebook has been well aware of many of the issues and topics external researchers and policymakers wish to explore. Moreover, the platform has been dedicating resources toward closed, internal research on those same topics but for PR and profit-motivated reasons (Frenkel and Kang 2021). This leaked data is thus particularly interesting to some researchers because of the piercing of the corporate black box and pushing against corporate-funded 'administrative' academic research (Abhishek

2021). Companies such as Facebook are less likely to be able to rely on claims of ignorance when their internal work shows them addressing the issues under question.

Finally, the Facebook leaks are likely not the last of their kind, given the growing precedence of leaks, especially from big tech employees, over the last decade (Sun 2021). In academic and public circles alike, there is now widespread acceptance of how massive platforms shape modern information flows, often in hidden, profit-motivated ways with negative impacts (Benvegnù et al. 2021). As public awareness of platform harms has been steadily growing, potential whistleblowers inside these platforms are not immune this kind of 'situational determinant' or nudge (Schmolke and Utikal 2018). Facebook's knowledge and lack of initial substantive action around elections, misinformation, and Russian interference provide a good example; this, among other sociopolitical issues impacted by the platform, spurred Haugen to go public and many other employees to leave.[3] Another reasoning relates to repercussions. There is a potential lessening of blacklisting and increasing of protections for defecting tech employees (Clayton 2021; OECD 2016); the source of the leak might also impact the willingness of other actors to engage with leaked material, including academics. For example, there was an initial noted absence of extensive academic engagement around the WikiLeaks releases, partially because of feared government repercussions (O'Loughlin 2016). However, in reviewing 20 peer-reviewed and editorial political science journals, Darnton (2021) found significant scholarly usage of leaked Wikileaks material by 2018, and even increasing over time. Another reason for increasing leaks has to do with affordances of information technologies being used by digitally-skilled populations (journalists and tech workers) to facilitate whistleblowing (Lazar 2022; Di Salvo 2021). The handling of the recent Facebook leak has also been described as a roadmap example and has prompted development of resources such as the Tech Worker Handbook for whistleblowing.[4]

Academic efforts around the leaked Facebook data

The character of the leaked data, the positionality of researchers, and the likelihood of more leaks set the stage for more grounded examination of academic engagement with the Facebook data. As the weeks and months tick by, full access is still widely limited to the original journalists, and the Facebook leak has largely disappeared from news cycles since November 2021. The following section describes perspectives and experiences of multi-disciplinary academics in small teams and larger coalitions navigating access to this leaked data. I first describe my own team's experiences and considerations—legal, ethical, and practical. I then briefly describe and compare some of the other academic-involved initiatives aimed at making the Facebook data more accessible—both to academics and the wider public.

Navigating Team Access

I am a postdoc within a small Norwegian-American team examining misinformation around elections, specifically exploring how newsrooms approach misinformation via technology, practices, education, and partnerships. We focus especially on factcheckers, including the field-defining Facebook factchecking partnership that arose in the wake of the 2016 US elections. I became interested in the Facebook leak for its wider implications around transparency, accountability, and platform power around misinformation. I

---

[3] The FF leaked data also included several internal posts from employees leaving the company based on grounds such as the company "having a net negative influence on politics" (Satariano and Isaac 2021) and other forms of disillusionment (Levy 2021). Also, in April 2021, The Guardian published material leaked from another Facebook analyst that received much less attention; this leak also detailed the platform's inaction around harmful disinformation campaigns and harassment outside the global North (Soni 2021).

[4] https://techworkerhandbook.org/

followed the breaking news stories, paying particular attention to those that might intersect with our work. In the first of the WSJ revelations concerning XCheck[5], I ran across one line mentioning platform-mandated decision reversals on fact-checked content (Horwitz 2021). Weeks later, the news outlet *Gizmodo* published a small blogpost with a document from the leak. It contained screenshots of internal, Facebook-wide discussion about a factchecking decision reversal in 2018 (Cameron and Wodinsky 2021). I messaged the bylined journalist over Twitter, asking about any other similar documents they might have uncovered. The journalist replied, generously offering our team full access to the data received from the whistleblower. In addition to describing security and data protection liabilities, Gizmodo also cautioned our team about potential cyberattacks if we publicly held such material (Marczak et al. 2021).

At the time of this publication, six months later, we still have not engaged with the data. The initial team meeting after being offered access was one characterized by uncertainty. Initial concerns were expressed around ethics and data protection liability. Given that most of our team is based in Norway and funded by the Norwegian government, our work is largely subject to the EU's General Data Protection Regulation (GDPR)—a regulatory regime that is substantially more robust and expansive than analogous protections in the USA. The leaked Facebook data contains significant amounts of identifiable personal information—names, titles, images—of employees (particularly lower level) and random platform users (whose profiles and personal data were used in internal slide decks and conversations). Beyond concerns of data protection and consent, some leaked material arguably also poses danger to public interests if broadcast.[6] Though the Haugen team conducted a first-pass redaction effort, but there is still a significant amount of material that needs to be further redacted. Thus, one main task of the different groups trying to facilitate access to the data (described later in this commentary) is improved redaction.[7]

Focusing simply on regulatory compliance, there are built-in exemptions for research in the GDPR. However, that does not mean the process of obtaining an exemption is quick or straight-forward. Our team was first offered access in early November 2021. Within two weeks, we had requested approval to examine the data from the NSD (Norwegian Centre for Research Data, which approves research practices in most Norwegian universities). The NSD discussed our project's request internally in December 2021 but could not reach a conclusion and asked our team instead to contact The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH). This 12-person committee provides evaluation for difficult individual cases; there was one member of the committee that could speak to our case. In mid-March 2022, NESH declared that 'in principle', it was ethically justifiable to conduct research on this data—despite questions of provenance, data protection, and platform private property—given its potential 'benefit to society'. However, NESH explicitly noted it is our responsibility, as researchers, "to concretize and prove the social benefits" outweigh ethical concerns—a potentially difficult task when we are unable to examine the data. Moreover, the committee returned the question of legality to our home institution, which (as of publication) has yet to come to a decision.

---

[5] One of the biggest revelations from the leaked data was the Facebook internal system for "exempting high-profile users", including politicians and celebrities, from content and behavior moderation. Known as XCheck (or Cross-Check), this system essentially 'whitelisted' millions of accounts, granting them *de facto* immunity if harassing others, inciting violence, and spreading misinformation (Horwitz 2021).

[6] One researcher, who I spoke to on background, has been involved in efforts to better redact and release the data. This person described documents in the data detailing how to "run a criminal enterprise on Facebook for human trafficking" and "how to run a money-laundering operation". They also noted that the leaked data also contained specific textual patterns (regexs) that can identify "horrible and dangerous content" currently on the platform.

[7] As one researcher with access noted, under anonymity, "These documents are not secure. Even after the redactions for Congress, there are identifiable people in there who are not Facebook employees, who are not Facebook executives. I genuinely believe that figuring out how these documents could be responsibly redacted and released is in the public's best interest."

Data protection compliance and research ethics aside, our initial discussions brought up other practical and far-ranging concerns. Ours is a relatively short-term project with funding tied to specific work packages and outputs; one question was if the leaked Facebook data would further those outputs. More and more, academia is structured in this way, for maximum and pre-determined output that can preclude more open agendas (Fasenfest 2021). The precarity of academics also plays into these concerns, where publishing in peer-reviewed journals is an inescapable metric of success. One of the first concerns brought up about this data was *if* it could be published. We reached out to one of the major publishers of communications and sociology journals, inquiring about the ability to publish any findings from this data. Their response indicated that while they would be willing to consider publications on a case-by-case basis, the potential legal and liability ramifications were significant and troubling. In reviewing published material around Wikileaks, Darnton (2021) argues that a general lack of clear publishing guidance has contributed to academics using but not properly disclosing leaked data in their work.

We then informally approached a legal expert based in the US, who described the publisher's response as the "most conservative". They also pointed out that US law might offer more avenues for publication on constitutional grounds around free expression—if it can be shown that publishing findings would best serve public interest. The party who leaked the information may be liable, but typically not the publisher.[8] The point for this commentary is two-fold. First, publisher reticence is part of a larger lack of readiness of the academic infrastructure for navigating leaked data. Second, it takes substantial resources and potential sacrifices (time, scarce funding, positionality to potentially forego publication count and encounter legal and reputational opposition) for any individual academic or small team to take on uncertainties—let alone navigate multiple legal jurisdictions. For a related example, I spoke with Dr. Emma Briant, a propaganda and information warfare scholar. Her academic work was key evidence in the US Congress, EU and British parliamentary inquiries into the Cambridge Analytica leak and the resultant scandal implicating Facebook. Her involvement in revealing wrongdoing, sparked by her feelings of "public responsibility as an academic", has led to ongoing difficult threats to her career and personal life (Briant 2021; Morgan 2018). She argued, in our discussion, that the Cambridge Analytica case shows both failing systems that allowed private firms to exploit academics, and a lack of institutional and legal protections for academics revealing wrongdoing and researching the powerful—
even where they are working in the context of parliamentary or governmental inquiries. If from a less-resourced or less prestigious institutions, the access and self-censorship effects would arguably be even greater (boyd and Crawford 2012). In the case of the Facebook leak, this is despite an urgent need for expertise on and greater coverage of issues outside the US and Europe (Scheck, Purnell, and Horwitz 2021).

In addition to liability and regulatory issues, our team discussions also identified various practical and logistical questions with the leaked data. Some cannot be resolved until we are able to look at the data and get a sense of its character, and what methods and resources would be used to store and analyze it in the context of our research remit. Some questions were around diversion of resources, given the structure and expected outputs of our project—could we afford to take what might be an unfruitful detour, expending substantial social and economic capital just to *look* at the data? (In speaking with another researcher who had also been navigating access, they voiced similar concerns, describing the data as being "either full of bombshells or duds…we don't know.") Another practical concern was expressed by some team members around future and ongoing access: what would this do to some team members' relationships with platforms? More explicitly, if we worked with this data, would they be able to continue sourcing interviews from contacts at platform and tech companies for their research agendas?

---

[8] New York Times Co. v. United States, 403 US 713 (Supreme Court 1873).

All the issues and roadblocks described above—privacy and security, liability exposure, question marks around publishing, quality and relevance of potential findings, resource sacrificing, the potential damaging of platform-academic relationships—have hampered any progress for our team. Yet we are also in a better position than many to interact with this type of data source, given our positionality in terms of resources and networks. Other academic projects could conceivably have other issues or questions, or lack the capacity to examine them as we did. Also, these issues at scale create a situation of unreadiness for the types of data coming from platforms—the legal, career, and administrative costs are too much for individual researchers or teams to navigate alone. All the issues point to the importance of more transparent, developed, and even institutionalized pathways for accessing and working with such data.

Wider academic-involved efforts around the Facebook leak

I began reaching out to researchers and institutions in the platform transparency and accountability space. Given the challenges encountered by our team, it seemed optimal to join forces—and perhaps contribute our subject expertise—to any other groups working with the Facebook data. In the weeks following the leak, academics began publicly arguing for access. One group of academics likened the situation to a pharmaceutical company only releasing scientific material to government officials and journalists (Sanderson, Nagler, and Tucker 2021). Another group of academics, primarily journalism and media scholars, published a piece in The Atlantic citing researcher access to this data as a matter of public interest in terms of countering platform power; they also disclosed that they were working on initiatives to share data (Benesch 2021). Whistleblower Frances Haugen shared with the New York Times that she did intend to offer academic 'writers' access, particularly those from the Global South with subject matter expertise (Smith 2021). When I reached out to a member of the whistleblower's team for details, I was told that a US-based academic institution had obtained access from Congress, and they would be "making [the data] publicly available for research purposes at some point this year". (There has been no public announcement from the whistleblower's team or this group of academics as of the time of publication.)

Later in 2021, two efforts were announced with aims to make the leaked data more publicly available and accessible via improved redaction. First, a group of Harvard University research labs began a project with an immediate aim was to launch an "innovative tool" for collaborative analysis of the data, mainly among vetted researchers and journalists; wider public access is a later goal. Their research with the data also includes aims to "measure and assess the de-identification of personal names found in images".[9] I contacted the Harvard-based initiative in early December 2021 via email. They encouraged us to pursue access with Gizmodo instead, but gave some insight into their planned timeline and work goals. They were aiming to use "internal users and close collaborators," and automated tools, to review, redact, and organize the data. They then envisioned batch releases to vetted researchers. At the time of this commentary's publication, the Harvard group has shared some visualizations of the data on their website but has not made any of the data more publicly available.

Second, a partnership of journalists and academics arose, aiming for a fully public release. It was spearheaded by the online outlet Gizmodo, one of the news outlets to receive early access to the data from Haugen's team. (This is also the news outlet that originally offered access to our team.) On 22 November 2021, Gizmodo announced via a blog post that it would be rolling out batches of the leaked data to the general public, with the help and review of an academic committee. This committee was described by Gizmodo as "a small group of independent monitors…establishing guidelines for an accountable review of the documents prior to publication" (Cameron, Couts, and Wodinsky 2021). When I spoke with an involved academic in March 2021, they described these 'monitors' as a deliberately multidisciplinary

---

[9] https://fbarchive.org/about

"committee of academics, from computer science to journalism." However, when we spoke in March 2021, the Gizmodo staff had gone on strike just as they were about to push the first batch of documents out to the academics for approval. The strike was resolved within the first week; at the time of this publication, Gizmodo has published two batches of documents on their website. The first focuses on the American presidential 2020 election and subsequent capitol riot; the second 'looks under the hood' of the Facebook news feed (Cameron, Wodinsky, and DeGuerin 2022).

Opinions, sensitivities, openness, and approaches diverged between different research groups and fields (including our own). These divergences are around research aims, resources, openness to collaboration, concerns around consent, expansiveness of releases, team composition, transparency, and how we viewed our work benefitting the public—directly or indirectly. These divergences are not inherently problematic, and a key part of heterogeneous academic inquiry. However, they can also unfortunately feed into the problem of academic and institutional unreadiness—a lack of widely shared guidelines, processes, or standards—for working with leaked data. The leaked Facebook data has led to multiple groups, publicly and behind the scenes, simultaneously expending scarce resources but with certain shared general aims: to redact and curate the leaked material for public benefit, via answering their own research questions and/or facilitating public access. There was also a shared sense of inevitable public release—and thus a sense of urgency and maybe even competition—between different groups. One researcher, in conversation, described it as a sense of immense 'information pressure'—a dynamic they believed would eventually push the documents wholly into the public sphere, with or without curation or interpretation.

Conclusion

This brief commentary is aimed at grounding and furthering discussion around working with leaked data in academia. It does not take a hardline, general stance; leaked data needs to be approached based on context. In focusing on the recent Facebook data leak, I have tried to transparently describe some of the issues faced by our team and others—ethical (privacy and security, public responsibilities, consent), legal (data protection and liabilities), practical (what tools and approaches to employ, resource allocations), and pragmatic (potential impacts on research agendas, careers, and professional relationships). Each of these arenas deserve more work and discussion within academia, predicated on more transparency between academics. This is despite the fact that such data of uncomfortable or unusual provenance can be (and has been, for this early career researcher) a sensitive and challenging topic to broach. But this topic will continue to arise, particularly in our era in platformization—and given academia's potential role in 'realizing democratic functions' in the modern media environment. Thus, as a concluding point, **we need a clearer picture of and more transparency around how different researchers are currently approaching leaked data; this is an important step toward the development of necessary guidelines, standards, and even mechanisms for engaging with this type of data.**

Positionalities (and marginalities) must also be central to discussion around leaked data. boyd and Crawford (2012) wrote of 'a new kind of digital divide' around big data sets from platforms, based on who gets to create, collect, and analyze—and who is left out. One obvious divide touched on throughout this commentary is between platforms and academics. Savage and Burrows (2007) focused on marginalization in terms of academics being shut out from industrial sites of knowledge production. But academia must address larger inequalities in knowledge production in the way we try to access—and use—platform data, leaked or not (Donovan 2020). This commentary described questions faced by a small research team and larger coalitions—but all these initiatives were based at well-funded, prestigious American and European universities. This is despite the leaked data including significant findings impacting the Global South. **This is another point for further discussion: if public benefit is a key reason for academia's engagement with leaked data, there is a substantial need to address larger**

inequalities and marginalities in academic knowledge production. Without the inclusion of more perspectives, academic evaluations and practices will be insufficient and potentially harmful, to our work and to impacted publics.

References

Abhishek, Aman. 2021. "Overlooking the Political Economy in the Research on Propaganda." *Harvard Kennedy School Misinformation Review*, April. https://doi.org/10.37016/mr-2020-61.

Alba, Davey. 2022. "Facebook Sent Flawed Data to Misinformation Researchers." *The New York Times*, February 24, 2022, sec. Technology. https://www.nytimes.com/live/2020/2020-election-misinformation-distortions.

Baritz, Loren. 1960. *The Servants of Power: A History of the Use of Social Science in American Industry*. The Servants of Power: A History of the Use of Social Science in American Industry. Middletown, CT, US: Wesleyan University Press. https://doi.org/10.1037/11283-000.

Benesch, Susan. 2021. "Nobody Can See Into Facebook." *The Atlantic*, October 30, 2021, sec. Ideas. https://www.theatlantic.com/ideas/archive/2021/10/facebook-oversight-data-independent-research/620557/.

Benvegnù, Carlotta, Niccolo Cuppini, Mattia Frapporti, Milesi Floriano, and Maurilio Pirone, eds. 2021. *Platformization and Its Discontents*. Vol. 120. 4 vols. South Atlantic Quarterly. Duke University Press. https://www.dukeupress.edu/platformization-and-its-discontents.

boyd, danah, and Kate Crawford. 2012. "CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Briant, Emma. 2021. "3 Years Ago My #CambridgeAnalytica and #SCL Evidence Was Published by @CommonsDCMS Fake News Inquiry. You Cannot Know How Hard, How Terrifying It Was. 3 Years on I Still Live, Write, Publish in Constant Threat but I Grow Stronger, Fight Harder with a Little Help from My Friends." Tweet. *Twitter*. https://twitter.com/EmmaLBriant/status/1384658866655408128.

Burrows, Roger, and Mike Savage. 2014. "After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology." *Big Data & Society* 1 (1): 2053951714540280. https://doi.org/10.1177/2053951714540280.

Calhoun, Craig. 2009. "Social Science for Public Knowledge." In *Academics as Public Intellectuals*, edited by Sven Eliaeson and Ragnvald Kalleberg. Cambridge Scholars Publishing.

Cameron, Dell, Andrew Couts, and Shoshana Wodinsky. 2021. "We're Making the Facebook Papers Public. Here's Why and How." News. Gizmodo. November 22, 2021. https://gizmodo.com/we-re-making-the-facebook-papers-public-here-s-why-and-1848083026.

Cameron, Dell, and Shoshana Wodinsky. 2021. "Yes, There Are More Facebook Papers. Yes, It's a Lot." News. Gizmodo. November 4, 2021. https://gizmodo.com/hey-kid-wanna-see-some-leaked-facebook-docs-1847936740.

Cameron, Dell, Shoshana Wodinsky, and Mack DeGuerin. 2022. "The Facebook Papers: The Algorithms That Control Your News Feed." Gizmodo. May 2, 2022. https://gizmodo.com/facebook-news-feed-ranking-algorithms-how-they-work-1848814459.

Clayton, James. 2021. "Silenced No More: A New Era of Tech Whistleblowing?" *BBC News*, October 11, 2021, sec. Technology. https://www.bbc.com/news/technology-58850064.

Darnton, Christopher. 2021. "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks." *American Political Science Review*, December, 1–16. https://doi.org/10.1017/S0003055421001374.

Di Salvo, Philip. 2021. "Securing Whistleblowing in the Digital Age: SecureDrop and the Changing Journalistic Practices for Source Protection." *Digital Journalism* 9 (4): 443–60. https://doi.org/10.1080/21670811.2021.1889384.

Donovan, Joan. 2020. "Big Data, Bigger Risks." *Harvard Kennedy School Misinformation Review*, January. https://doi.org/10.37016/mr-2020-006.

Edgerton, Anna, David McLaughlin, and Sara Forden. 2021. "Facebook Hobbled Team Tasked With Stemming Harmful Content." *Bloomberg.Com*, October 25, 2021. https://www.bloomberg.com/news/articles/2021-10-25/facebook-whistle-blower-files-reveal-internal-struggle-profit-vs-public-safety.

Esser, Frank, and Christoph Neuberger. 2019. "Realizing the Democratic Functions of Journalism in the Digital Age: New Alliances and a Return to Old Values." *Journalism* 20 (1): 194–97. https://doi.org/10.1177/1464884918807067.

Fasenfest, David. 2021. "Reflections on the Decline of Academia: Large Problems and Small Minds." *Critical Sociology* 47 (7–8): 1057–63. https://doi.org/10.1177/08969205211054607.

Frenkel, Sheera, and Cecilia Kang. 2021. *An Ugly Truth: Inside Facebook's Battle for Domination*. HarperCollins.

Harcup, Tony. 2011. "Hackademics at the Chalkface." *Journalism Practice* 5 (1): 34–50. https://doi.org/10.1080/17512786.2010.493333.

Horwitz, Jeff. 2021. "Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt." *Wall Street Journal*, September 13, 2021, sec. Tech. https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353.

Ingram, Mathew. 2021. "On the Facebook Papers Media Strategy." Academic Blog. Columbia Journalism Review. October 28, 2021. https://www.cjr.org/the_media_today/on-the-facebook-papers-media-strategy.php.

Jong-Fast, Molly, and Andy Levy. 2021. *The New Abnormal: How WSJ Reporter Jeff Horwitz Got the Facebook Whistleblower to Talk on Apple Podcasts*. The New Abnormal. https://podcasts.apple.com/us/podcast/unlocked-how-wsj-reporter-jeff-horwitz-got-the/id1508202790?i=1000540339073.

Kantrowitz, Alex. 2021. "I'm in the Consortium Possessing the Leaked Facebook Documents. Let's Dissolve It." News. *Nieman Lab* (blog). October 26, 2021. https://www.niemanlab.org/2021/10/im-in-the-consortium-possessing-the-leaked-facebook-documents-lets-dissolve-it/.

Lazar, Tamar. 2022. "Organizational Scandal on Social Media: Workers Whistleblowing on YouTube and Facebook." *Information and Organization* 32 (1): 100390. https://doi.org/10.1016/j.infoandorg.2022.100390.

Levy, Steven. 2021. "Facebook Failed the People Who Tried to Improve It." *Wired*, October 25, 2021. https://www.wired.com/story/facebook-papers-badge-posts-former-employees/.

Marczak, Bill, John Scott-Railton, Siena Anstis, Bahr Abdul Razzak, and Ron Deibert. 2021. "Breaking the News: New York Times Journalist Ben Hubbard Hacked with Pegasus after Reporting on Previous Hacking Attempts." Citizen Lab Research Report No. 145. University of Toronto. https://citizenlab.ca/2021/10/breaking-news-new-york-times-journalist-ben-hubbard-pegasus/.

Morgan, John. 2018. "'Unethical Not to' Submit Brexit Interviews to MPs, Says Academic." Times Higher Education. April 22, 2018. https://www.timeshighereducation.com/news/unethical-not-submit-brexit-interviews-mps-says-academic.

Nobel Peace Center. 2022. *Is Facebook Undermining Democracy and Creating Harm?* https://www.youtube.com/watch?v=CcC3VPqQqBw.

Obermaier, Frederik, and Bastian Obermayer. 2017. *The Panama Papers: Breaking the Story of How the Rich and Powerful Hide Their Money*. Simon and Schuster.

OECD. 2016. "Committing to Effective Whistleblower Protection." Paris: Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/governance/committing-to-effective-whistleblower-protection_9789264252639-en.

O'Loughlin, John. 2016. "The Perils of Self-Censorship in Academic Research in a WikiLeaks World." *Journal of Global Security Studies* 1 (4): 337–45. https://doi.org/10.1093/jogss/ogw011.

Sanderson, Zeve, Jonathan Nagler, and Joshua A. Tucker. 2021. "Academic Researchers Need Access to the Facebook Papers." *Slate*, November 4, 2021. https://slate.com/technology/2021/11/facebook-papers-frances-haugen-academic-research-transparency.html.

Satariano, Adam, and Mike Isaac. 2021. "Facebook Whistle-Blower Brings Campaign to Europe After Disclosures." *The New York Times*, October 25, 2021, sec. Business. https://www.nytimes.com/2021/10/25/business/frances-haugen-facebook.html.

Savage, Mike, and Roger Burrows. 2007. "The Coming Crisis of Empirical Sociology." *Sociology* 41 (5): 885–99. https://doi.org/10.1177/0038038507080443.

Scheck, Justin, Newley Purnell, and Jeff Horwitz. 2021. "Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show." *Wall Street Journal*, September 16, 2021, sec. Tech. https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953.

Schmolke, Klaus Ulrich, and Verena Utikal. 2018. "Whistleblowing: Incentives and Situational Determinants." SSRN Scholarly Paper ID 3198104. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3198104.

Smith, Ben. 2021. "Inside the Big Facebook Leak." *The New York Times*, October 25, 2021, sec. Business. https://www.nytimes.com/2021/10/24/business/media/facebook-leak-frances-haugen.html.

Soni, Paromi. 2021. "A Tale of Two Facebook Leaks." *Columbia Journalism Review*, November 1, 2021. https://www.cjr.org/analysis/two-facebook-leaks.php.

Sun, Mengqi. 2021. "More Tech Whistleblowers Are Expected, Experts Say." *Wall Street Journal*, October 28, 2021, sec. C Suite. https://www.wsj.com/articles/more-tech-whistleblowers-are-expected-experts-say-11635413403.

Timberg, Craig. 2021. "Facebook Made Big Mistake in Data It Provided to Researchers, Undermining Academic Work." *Washington Post*, September 10, 2021. https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/.

Vaidhyanathan, Siva. 2021. "Facebook Has Just Suffered Its Most Devastating PR Catastrophe Yet." *The Guardian*, October 8, 2021, sec. Opinion. https://www.theguardian.com/commentisfree/2021/oct/08/facebook-frances-haugen-whistleblower-pr-catastrophe-yet.

Wahl-Jorgensen, Karin. 2014. "Is WikiLeaks Challenging the Paradigm of Journalism? Boundary Work and Beyond," 12.

Wall Street Journal. 2021. "The Facebook Files," October 1, 2021, sec. Tech. https://www.wsj.com/articles/the-facebook-files-11631713039.

Weber, Matthew S. 2021. "Advocating for Journalism Studies' Impact on Policymaking." In *Journalism Research That Matters*, edited by Valérie Bélair-Gagnon and Nikki Usher. Oxford University Press.

Zuboff, Professor Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Main edition. London: Profile Books.