# Policy Memos

## Canadian Commission on Democratic Expression

### Learning Session 2: Should we mandate financial and other disclosures around advertising and, if so, what are the challenges in doing so?

Thursday, Oct. 14, 2021 | 1:00 p.m. – 2:30 p.m. ET (UTC -4:00)
Virtual event via Zoom

### Abstract of session

Users around the world are being targeted with paid advertising on social media platforms, often without knowing how or when they are being targeted, or who is paying for them. Political advertising in particular, has come under mounting scrutiny for interfering with elections and democratic processes around the world. While countries like France and Canada have laws in place to protect elections from being influenced by political ads, user targeting, paid disinformation campaigns, and political advertising is an ongoing problem for dominant platforms who offer minimal meaningful transparency on the advertising activity, reach, and revenue on their platforms. While some platforms maintain archives or "ad libraries" of the political ads they run, there is a wide variation of descriptions and access to that data, making it difficult for publics and regulators to assess. In 2020, the European Commission unveiled the Digital Services Act, a draft law that focuses on systemic transparency, including mandating disclosure of all online advertising, including their key targeting parameters.

**Policy questions:**

What financial data should be shared and with whom?
Should platform companies be mandated to disclose information about political ads or all advertising on their platforms? What should be included in associated financial disclosure reports?
What improvements could be made to existing platform ad libraries to facilitate public access, searchability, and standardization across platforms?

My name is Catherine Armitage and I'm a policy advisor at AWO[1]. We are a new organisation working across the world to shape, apply and enforce data rights. We work with international organisations, governments, companies, NGOs, universities, philanthropy and individuals.

Before I joined AWO 5 months ago, I had spent my career in the advertising industry. I worked in digital marketing at the world's biggest advertiser, Procter & Gamble, and for the past 5 years I was Director of Policy at the global trade body for advertisers.

Today, I'm going to talk about three reasons why digital advertising is broken and what we need to focus on in order to fix it.

Online advertising - and the surveillance advertising business model - has become a proxy for a lot of things that are wrong with the way the digital economy functions today.

These include discrimination, disinformation, the undermining of democracy and manipulation of the public debate, the amplification of hate speech, racism, violence and content that harms our children.

The truth is that the personalised advertising – sometimes too personal – we see online is the most visible part of some much bigger underlying issues. These issues are:

1) The power of Big Tech.

2) Lack of accountability and transparency around how these platforms make money.

3) People having no control over companies turning every detail of their lives into profit.

All of this feeds into the problems I mentioned earlier, but these are the issues that urgently need to be tackled. If not, we might manage to relieve some of the symptoms –ads that follow us around the internet, for example - but we won't be able to treat the disease.

The dynamics of what Shoshana Zuboff calls the 'human futures market' are shifting quite dramatically at the moment.[2]

The entire digital advertising industry is in the process of re-writing the rules on data collection. How data can be collected, when and – most importantly - by whom.[3] [4] This might be the story of a 'data arms race' where companies double down on data collection in order to compete and survive.

---

[1] https://www.awo.agency/

[2] Zuboff, Shoshana. The Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power. 2019.

[3] In April 2021, Apple introduced 'App Tracking Transparency' (ATT) which prompted users to opt in to cross-site tracking by apps. So far, around 80% of users have chosen not to be tracked, significantly limiting the amount of data that can be collected by app owners and third parties in an iOS environment. See: 'User Privacy and Data Use - App Store' (*Apple Developer*) <https://developer.apple.com/app-store/user-privacy-and-data-use/> accessed 6 October 2021, Lazuik E, 'IOS 14 Opt-in Rate - Weekly Updates Since Launch' (*Flurry*, 25 May 2021) <https://www.flurry.com/blog/ios-14-5-opt-in-rate-idfa-app-tracking-transparency-weekly/> accessed 6 October 2021.

[4] Google announced last year that it will phase out third-party cookies in the Chrome browser, thereby limiting the amount of companies that can track user activity across the web. See: Schuh J, 'Building a More Private Web: A Path towards Making Third Party Cookies Obsolete' (*Chromium Blog*) <https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html> accessed 8 October 2021, Goel V, 'An Updated Timeline for Privacy Sandbox Milestones' (*Google*) <https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones/> accessed 6 October 2021.

But it could also be the start of a new chapter for the industry. Anyone who – like me – has sat through endless industry debates about the "cookie-less world" and – like me – have even written papers on it[5], know that a central theme tends to be the need for advertisers and publishers to re-evaluate the central role of data in what is essentially a broken system. This actually could be a great thing for independent media, society and even democracy too.

Today, Google and Facebook are the biggest advertising companies in the world. They make nearly all of their money from advertising.[6] And they are also the two companies with the most access to our data. Nearly everything we do online can be tracked in some way – and when you look at which companies are doing most of that tracking, it's Google and Facebook.[7]

Google does this by controlling the access points to the internet – browsers and mobile operating systems.[8] Antitrust complaints and investigations have shown how this approach deprives users of choice – especially when it comes to how their data is collected and used.[9]

Facebook collects data by controlling and tracking the places where we spend most of our time online. Last week, Frances Haugen showed how this drives the company to prioritise user engagement – and therefore profits - over the public good.[10]

Often, this data is being collected without people's full knowledge or agreement. And there are no limits to what data can be collected. In a market where revenue is directly linked to the ability to collect data, this creates a perfect storm.

These companies are also increasingly opaque about the way they use this data for advertising. Advertisers have no choice but to rely on their algorithms, trusting that their marketing campaigns will work and their money won't be wasted. And trusting that their money isn't fuelling division, hate, abuse or eroding people's privacy.

But, when you talk to advertisers, you often get the feeling they have as much control over where they advertise as consumers do over sharing their data. Facebook and Google are so dominant that very few companies can afford not to advertise with them.

Combined with a policy context where digital advertising is under increased scrutiny, the market feels ripe for reform. Behind closed doors – and sometimes publicly – many advertisers and

---

[5] The Future of Data-Driven Marketing. World Federation of Advertisers, 2021, https://wfanet.org/knowledge/item/2021/03/10/WFA-report-The-future-of-data-driven-marketing

[6] Google and Facebook respectively generated 80% and 98% of their revenue from digital advertising last year. 'Alphabet Year in Review 2020' (Alphabet) <https://abc.xyz/investor/static/pdf/2020_alphabet_annual_report.pdf?cache=8e972d2>; 'Facebook, Inc. Form 10-K' (2020) <https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/4dd7fa7f-1a51-4ed9-b9df-7f42cc3321eb.pdf>.

[7] Appendix G to 'Online Platforms and Digital Advertising - Market Study Final Report' (Competition & Markets Authority 2020), https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf.

[8] Schmidt DC, 'Google Data Collection' (Digital Content Next 2018), https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf.

[9] State of Utah and others v Google LLC [2021] United States District Court, Northern District of California, San Francisco Division Case No. 3:21-cv-05227, <https://ag.ny.gov/sites/default/files/utah_v_google.1.complaint_redacted.pdf>; Commission Decision of 18.7.2018 [2018] European Commission Case AT.40099, <https://ec.europa.eu/competition/antitrust/cases/dec_docs/40099/40099_9993_3.pdf>; United States of America and others v Google LLC (United States District Court for the District of Columbia), <https://www.justice.gov/opa/press-release/file/1328941/download>, First Amended Complaint for Injunctive Relief: Epic Games, Inc v Google LLC and others [2021] United States District Court, North District of California Case No. 3:20-CV-05671-JD. <https://storage.courtlistener.com/recap/gov.uscourts.cand.364325/gov.uscourts.cand.364325.165.1.pdf>.

[10] C-SPAN, 'Facebook Whistleblower Frances Haugen Testifies before Senate Commerce Committee' (5 October 2021) <https://www.youtube.com/watch?v=GOnpVQnv5Cw>.

publishers will tell you that the current system is broken. As Big Tech rewrite the rules of the game, maybe some will decide they don't want to play anymore. And maybe partnerships will emerge to explore new ways of doing things. This could be a path to creating a digital advertising ecosystem which is more respectful, fairer and more sustainable. And that creates a genuinely solid basis to fund independent media and support democracy. If that's the case, policymakers should be listening carefully - guiding where necessary - and participating in this journey to ensure that this vision becomes a reality.

# Moving Beyond Voluntary Transparency

Laura Edelson, NYU Cybersecurity for Democracy

Over the past several weeks, the public has learned a great deal about a variety of user safety concerns on Facebook's platforms. The specifics vary, but one thing is becoming increasingly clear: user safety appears to be inherently in tension with user engagement. Now that we know this, what can we do about it?

Across the globe, legislators and regulators are looking for solutions to protect users from online hate and misinformation. Engagement-based content promotion algorithms seem to be connected to a variety of problems, but exactly what should be done is harder to say. Should such algorithms be limited or banned entirely? Should platforms be prohibited from collecting the interaction data fuels such algorithms? What responsibilities should platforms have to limit the spread of harmful content? Right now, the answer to all these questions is: we don't have enough information to say. In order to come up with more concrete solutions, we need data.

For several years, researchers who study platforms for user-generated content have been focused on the voluntary transparency measures that Facebook in particular has been proposing. But something else that has become abundantly clear over the last several months is that voluntary transparency has failed. Platforms have resisted calls to make certain kinds of data available, such as impressions, that would allow people to study the platform itself, instead of just how the platform is used. And Facebook has moved swiftly to shut down researchers like my team, who seek to secure streams of data that Facebook does not control and operate independent research programs. In order to make progress towards real solutions, we need governments to act to require transparency so that researchers, journalists, and the public have access to the data we need. This doesn't need to come at the expense of user privacy. Most of the data that researchers need are already public, so the platforms just need to give academics and journalists better access to it. Right now, we need three things:

First, Universal Digital Ad Transparency is long overdue. Last year, I and nearly a dozen other researchers issued a call for platforms to adopt this voluntarily. The biggest digital ad platforms haven't yet stepped up, so they should be required to make all the ads they run publicly available in a machine-readable format. We will soon be publishing a draft proposal that spells out the technical specifications needed in detail.

Second, a researcher safe harbor law would be invaluable in protecting researchers who engage in direct collection of data from platforms. Platforms currently use their terms of service as a cudgel to shut down legitimate, and privacy-conscious projects. Most academic researchers simply can't afford to be sued by a major platform. This is why multiple projects that appear to be entirely legal and ethical have been shut down by cease and desist notices. The passage of a law such as this would not directly give researchers access to data, but it would clarify the legality of a great deal of work by independent researchers.

Third, platforms should be required to give researchers and journalists access to public content on their platforms. Several platforms, including Twitter and Facebook, already have tools in place for businesses to access this data, but Facebook in particular has been reluctant to grant researchers access to the tools they provide to content creators, and Twitter's transparency product, the Firehose API, is priced beyond the reach of most academic researchers. And there are important platforms, such as TikTok, that offer no tools of this kind at all.

Unfortunately, I simply no longer believe that incentivizing platforms to voluntarily make the data we need transparent will work - the last several years have demonstrated that clearly. And platforms have shown that they can't stop the spread of harmful content by themselves. But the good news is, they don't need to. Many researchers are ready to help, but we need governments to introduce strong digital platform transparency and accountability legislation so that we can have the data we need to do our jobs. This transparency can be the first step toward meaningful, publicly accountable solutions.

**Ethan Zuckerman**, Associate Professor of Public Policy, Information and Communication, University of Massachusetts, Amherst (USA)

# Briefing for The Canadian Commission on Democratic Expression on the challenges of studying the social media platforms (Facebook, YouTube, Twitter and others), October 11, 2021

**Dear Commissioners:**

Thank you for the opportunity to address this distinguished body on a critically important topic. As we are having this conversation, Facebook, the most visible social media company, is weathering a wave of media scrutiny. A whistleblower has revealed that the company has commissioned, and later ignored, research reports that indicate its tools can be harmful to individuals (young women on Instagram in particular) and societies as a whole (through the amplification of angry and polarized viewpoints.) Rather than address those specific research studies, I want to raise the question of why we are so reliant on whistleblower information to interrogate harms associated with Facebook and other platforms.

The answer is very simple: it's extremely hard to study Facebook and other platforms in ways that would give us believable information about the platforms' effects on individuals, groups and societies as a whole. With my colleague, Dr. Elizabeth Hansen Shapiro, I led an investigation of this topic for a set of US-based foundations called Netgain who were interested in helping academic, journalism and advocacy partners investigate questions of the platforms' influence on democratic societies. We interviewed over thirty research teams to understand what questions they were trying to answer and why answering many questions about social media platforms is so difficult.

Here are some reasons why research is so challenging, particularly for researchers seeking to analyze large set of data:

**Most platforms significantly limit how much public data a researcher can collect.** YouTube limits registered research users to 100 searches per day[1], which means a task like compiling a list of extremist videos based on search terms could require a researcher weeks to complete, and weeks more any time you attempted to update it. Twitter, which is admirable in many of its approaches to research access, forbids researchers from calculating the total number of tweets posted per day or estimating usage of the service.[2] While this is intended to prevent researchers from reporting how big Twitter is compared to other services, it also prevents a researcher from calculating how prevalent or rare a type of content is on Twitter – we can detect tweets containing hate speech posted in a given day, but not the number of tweets posted that day, giving us the numerator, but not the denominator.

---

[1] Researchers are given 10000 API "credits" per day on YouTube. Some operations, like listing videos in a user's channel, cost 1 credit. Searches – i.e., list data for videos that mention "nazis" – cost 100 credits. See https://developers.google.com/youtube/v3/determine_quota_cost for details on the credit system.

[2] See https://developer.twitter.com/en/developer-terms/agreement-and-policy, specifically the section on Twitter Performance Benchmarking.

**Many types of data are simply inaccessible to researchers**. Most content posted to Facebook is shared with Facebook friends, not the general public, and cannot be accessed through the site's research tools, even in aggregate. Information sensitive to business operations, like data on ad targeting or on content moderation decisions, is generally inaccessible across platforms, even though such data is central to many questions about platform influence on elections and platform control over speech.

**Researchers are actively hindered from collecting their own data**. Responsible and careful researchers have responded to these constraints by developing their own data sets. These unauthorized data sets are sometimes collected by "scraping" publicly accessible web data, much as search engines develop their own indexes of the web (Media Cloud, Pushshift). Others recruit panels of web users and ask them to volunteer demographic data and data from their web browsers (The Markup's Citizen Browser, Mozilla's Project Rally). Facebook recently retaliated against the NYU Ad Observatory, which uses a data donation method to collect political ads, changing their system to break Ad Observatory's tools and suspending Facebook accounts for the involved researchers. Facebook claimed it was protecting user privacy as required under a US FTC consent order. The acting director of the FTC's Consumer Protections Bureau intervened with a stern public letter, noting "the consent decree does not bar Facebook from creating exceptions for good-faith research in the public interest. Indeed, the FTC supports efforts to shed light on opaque business practices, especially around surveillance-based advertising."[3] Facebook has not restored the researchers' accounts.

**We cannot trust data from the platforms.** The platforms would prefer that researchers use data provided by the platforms instead of collecting our own. The Facebook Ad Observatory researchers collected ad data specifically because they found many political ads were missing from the data Facebook released to its public ad archive. Researchers who applied to work with Facebook data through an academic/industry partnership called Social Science One waited over a year for data access. Then a researcher discovered that Facebook had only released half the data it had promised. Researchers believed they were working with data from all US Facebook users, but Facebook had only released data for users with known political affiliations, rendering the data useless for most political research purposes.[4]

In our interviews, we found no research team who believed they had access to the data they needed to study key social science questions. When we asked a senior researcher involved with a high-profile industry/academic partnership whether he believed researchers should rely on data from the platforms, his answer was an unambiguous no: "We need as many ways of collecting data as there are platforms, and probably more."

---

[3] A letter from Acting Director Samuel of the FTC's Consumer Protection Division to Facebook, archived at https://www.ftc.gov/news-events/blogs/consumer-blog/2021/08/letter-acting-director-bureau-consumer-protection-samuel

[4] Craig Timberg, "Facebook made big mistake in data it provided to researchers, undermining academic work", September 10, 2021, Washington Post. Available at:https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/

Our report offered several recommendations for funders seeking to support research access to platform data, five of which I incorporated into a briefing for the journal Nature. I offer those five points here:

**Platforms should give researchers access to the data tools advertisers have access to**. Advertisers on Facebook have access to Crowdtangle, an internal analytics tool that many researchers are not able to access. Advertisers across Google's properties have access to ad targeting data that can be useful for understanding the makeup of Google's audience. It is routine for academics to pose as ad buyers to use these tools, but this technical violation of terms of service is another way platforms can block research they dislike

**Researchers must be allowed to pool and share data collected from the platforms**. This data might be obtained through the APIs provided by platforms, or might be generated through scrapers. So long as researchers are operating their tools responsibly and not affecting access to the sites, the public interest in understanding the operations of these platforms outweighs the costs to the platform operator in enabling this research.

**Users must be allowed to donate their data to public interest research projects.** The code behind these projects needs to be carefully reviewed, and those code audits should have the power to shut down projects if researchers are using data in a way other than they've advertised. Otherwise, platforms should not be able to shut down research carried out through data donation studies.

**We need a safe haven provision for public interest research.** Researchers who collect data from the platforms have reason to worry that their actions may violate the US Computer Fraud and Abuse Act, an especially vague piece of legislation passed in 1986, which has been used to criminalize legitimate research activities of computer systems. While recent caselaw has made clear that simply violating a website's Terms of Service in the course of research is not a crime, CFAA makes certain types of platform research risky for those involved and often leads to IRBs blocking important studies. An explicit protection of academics, journalists and activists accessing systems in the course of public interest research would be a critical step towards enabling this work.

**We need a powerful independent auditor of algorithmic fairness.** Many of the most important questions about social media platforms involve actions of their algorithms. Is some political content favored over others? Are search engine results politically biased? Are members of protected groups (racial or religious minorities) excluded from ads for job opportunities or housing? These algorithms are difficult, and sometimes impossible, to evaluate from the outside. We propose a professional audit body that can work within a company to test the fairness of its algorithms against a (yet to be created) set of industry standards, certifying a level of fairness much as a financial auditor uses proprietary information to certify that a publicly traded company is complying with generally accepted accounting principles (GAAP).

Some structures already exist to make these interventions possible. The European Data Governance Act, proposed last year, includes the concept of "data altruism", which could support the right to donate data. Existing collections of unauthorized platform data exist and are widely used in scholarly circles – establishing a right to maintain these platforms would be simply recognizing an

existing reality. What is most critical is a shift in understanding. Understandably spooked by the data breach that revealed millions of user records to Cambridge Analytica, Facebook and other firms are seeking ways to prevent their data from being technically accessed and used for nefarious purposes. In the process, they are limiting legitimate access to their data for public benefit.

The strategy of technical protection is not working. Clearview AI has scraped billions of images from social media sites in violation of Canadian privacy law, yet remains operational in the US with almost $40 million in venture capital funding. When we consider collection and use of data from social media platforms, **intent matters**. Data collected by responsible academics to understand some of the most pressing social issues of our time should be treated differently than data stolen from billions of unaware users to create a for-profit product.

*This briefing relies heavily on New Approaches to Platform Data Research, archived at https://drive.google.com/file/d/1bPsMbaBXAROUYVesaN3dCtfaZpXZgI0x/view and prepared in fall 2020 for the Netgain Partners.*