

2021

A FOIA for Facebook: Meaningful Transparency for Online Platforms

Michael Karanicolas
mkaranicolas@gmail.com

Follow this and additional works at: <https://scholarship.law.slu.edu/lj>



Part of the [Law Commons](#)

Recommended Citation

Michael Karanicolas, *A FOIA for Facebook: Meaningful Transparency for Online Platforms*, 66 St. Louis U. L.J. (2021).

Available at: <https://scholarship.law.slu.edu/lj/vol66/iss1/4>

This Article is brought to you for free and open access by Scholarship Commons. It has been accepted for inclusion in Saint Louis University Law Journal by an authorized editor of Scholarship Commons. For more information, please contact [Susie Lee](#).

A FOIA FOR FACEBOOK: MEANINGFUL TRANSPARENCY FOR ONLINE PLATFORMS

MICHAEL KARANICOLAS*

ABSTRACT

Transparency has become the watchword solution for a range of social challenges, including related to content moderation and platform power. Obtaining accurate information about how platforms operate is a gatekeeping problem, which is essential to meaningful accountability and engagement with these new power structures. However, different stakeholders have vastly different ideas of what robust transparency should look like, depending on their area of focus. The platforms, for their part, have their own understanding of transparency, which is influenced by a natural drive to manage public perceptions.

This paper argues for a model of platform transparency based on better practice standards from global freedom of information or right to information systems. The paper argues that moves by platforms to assume responsibility over the truth or falsity of the content they host and amplify justifies a shift in how we understand their obligations of transparency and accountability, away from traditional self-reporting structures and towards a quasi-governmental standard where data is “open by default.” This change in posture includes creating a mechanism to process information requests from the public, to accommodate the diverse needs of different stakeholders. The paper also suggests establishing a specialized quasi-independent entity (a “Facebook Transparency Board”) which could play a role analogous to an information commission, including overseeing disclosure decisions and acting as a broader champion of organizational transparency. Although these changes represent a significant conceptual shift, they are not entirely unprecedented among private sector entities whose role includes a significant public function, and the paper notes a number of examples, such as the Internet Corporation for Assigned

* Drafted by Michael Karanicolas, Executive Director, UCLA Institute for Technology, Law & Policy. Thanks to John Villasenor, Caroline Mala Corbin, Christina Koningisor, Israel Balderas, Eric Freedman, Thomas Streinz, Sarah Haan, Thomas Kadri, Niklas Eder, Jisu Kim, Francesca Procaccini, Nikolas Guggenberger, Przemyslaw Palka, Artur Pericles Lima Monteiro, Carlos Liguori, Maren Woebeking, Christoph Busch, Pauline Trouillard, Daniel Maggen, Adam Posluns, Rafael Bezerra Nunes, Alicia Solow-Niederman, and Leah Ferentinos for their helpful feedback and commentary.

Names and Numbers' Documentary Information Disclosure Policy, which could serve as a model for the platforms to follow.

INTRODUCTION

These days, everyone seems to be a fan of transparency. Politicians of every political stripe are keen to preach their belief in it,¹ regardless of whether the rhetoric matches their actual record.² From marketing,³ to healthcare,⁴ to journalism,⁵ transparency has become a buzzword for solving just about every problem. Calls for transparency are a ubiquitous feature of debates around content moderation.⁶ For their part, the major platforms at the center of these conversations have been keen to tout their own belief in the benefits of transparency, and the unprecedented investments they are making toward this goal.⁷ It's a curious thing: if everyone is so committed to transparency, why do people keep complaining about a lack of transparency?⁸

1. See, e.g., *Trump: I'm the most transparent president' in history*, POLITICO (May 22, 2019), <https://www.politico.com/video/2019/05/22/trump-im-the-most-transparent-president-in-history-068163>; Stephen Harper, Can. Prime Minister, Address to Caucus on Parliament Hill in Ottawa (May 21, 2013) (transcript available at <https://www.theglobeandmail.com/news/politics/full-text-of-harpers-speech-on-senate-expenses-scandal/article12035932/>) (“Canada now has one of the most accountable and transparent systems of governance in the entire world and this is something Canadians are rightly proud of.”).

2. Annie Karni, *Meet the guys who tape Trump's papers back together*, POLITICO (June 10, 2018), <https://www.politico.com/story/2018/06/10/trump-papers-filing-system-635164>; Canadian Press, *Info czar warns against government's new obstructive tactics*, CBC NEWS (Sept. 27, 2013), <https://www.cbc.ca/news/politics/info-czar-warns-against-government-s-new-obstructive-tactics-1.1870656>.

3. Blake Morgan, *Why Transparency In Marketing Is Key For the Customer Experience*, FORBES (Sept. 8, 2015), <https://www.forbes.com/sites/blakemorgan/2015/09/08/why-transparency-in-marketing-is-key-for-the-customer-experience/?sh=16fc49af61b9>.

4. Soroush Sagharian & Wallace J. Hopp, *The Role of Quality Transparency in Health Care: Challenges and Potential Solutions*, NAT'L ACAD. OF MED. (Nov. 18, 2019), <https://nam.edu/the-role-of-quality-transparency-in-health-care-challenges-and-potential-solutions/>.

5. Eddie Scarry, *'Transparency' in journalism doesn't exist, but it could*, WASH. EXAM'R (Sept. 10, 2019), <https://www.washingtonexaminer.com/opinion/transparency-in-journalism-doesnt-exist-but-it-could>.

6. See, e.g., John Bowden, *Biden hits social media firms over lack of transparency*, THE HILL (Dec. 5, 2017), <https://thehill.com/policy/international/363299-biden-hits-social-media-firms-over-lack-of-transparency>; Kate Connolly, *Angela Merkel: internet search engines are 'distorting perception'*, THE GUARDIAN (Oct. 27, 2016), <https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception>.

7. Mark Zuckerberg, *Starting today, we're officially creating a new standard of transparency for online political ads*, FACEBOOK (May 24, 2018), <https://www.facebook.com/zuck/posts/starting-today-were-officially-creating-a-new-standard-of-transparency-for-onlin/10104972903079161/>; Daniel Howley, *Twitter CEO Jack Dorsey: 'We can do more to provide algorithmic transparency'*, YAHOO! (Mar. 24, 2021), <https://www.yahoo.com/entertainment/twitter-ceo-jack-dorsey-we-can-do-more-to-provide-algorithmic-transparency-190416046.html>.

8. Chris Mills Rodrigo, *Facebook Oversight Board director knocks lack of transparency in Trump ban*, THE HILL (May 5, 2021), <https://thehill.com/policy/technology/551959-facebook-oversight-board-director-knocks-lack-of-transparency-in-trump-ban>.

“Transparency,” as it turns out, is a pretty flexible concept, which means different things to different people, and which can be twisted in any number of directions.⁹ For anti-corruption watchdogs, transparency means being able to trace flows of cash, particularly when they involve people in positions of power.¹⁰ For authoritarians, transparency can mean being able to monitor the activities of their political opponents, including those trying to expose corruption.¹¹ Both definitions have an internal logic to them, though they are rooted in diametrically opposed value structures.

Debates around transparency in the content moderation space have been ongoing for years.¹² However, they gained new prominence over the course of 2020, as the Covid-19 pandemic, along with conflict accompanying the 2020 U.S. election, finally broke a long-running impasse over the role of platforms in the public discourse, and, in particular, the companies’ responsibility to vet the truth or falsity of the content they host and amplify.¹³

This Article considers the implications flowing from platforms’ pivot towards stronger intervention against misinformation, and argues for a shift in how we understand their obligations of transparency and accountability, away from traditional self-reporting structures and towards a quasi-governmental standard where data is “open by default.”¹⁴ In particular, the Article calls for transparency structures that are modeled on freedom of information legislation (otherwise known as the right to information), such that the platforms would operate under a presumption that information should be open and accessible to the public, subject to narrowly constructed exceptions to protect legitimate interests.¹⁵

9. *Transparency in Historical Perspective*, in *TRANSPARENCY: THE KEY TO BETTER GOVERNANCE?* 3, 4 (Christopher Hood & David Heald eds., 2006); Tarleton Gillespie, *CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA* 212 (Yale University Press ed. 2018).

10. See, e.g., *Access Info Europe and OCCRP Call for Open Company Registers in Europe*, ORGANIZED CRIME & CORRUPTION REPORTING PROJECT (Apr. 7, 2016), <https://www.occrp.org/en/daily/5118-access-info-europe-and-occrp-call-for-open-company-registers-in-europe>.

11. Cassie Maas, *Top EU court rules against Hungary NGO transparency law*, JURIST (June 19, 2020), <https://www.jurist.org/news/2020/06/top-eu-court-rules-against-hungary-ngo-transparency-law/>.

12. See, e.g., Michael Karanicolas, *Stand Up for Digital Rights! Recommendations for Responsible Tech*, CTR. FOR LAW & DEMOCRACY, June 15, 2016, at 85–96.

13. Evelyn Douek, *The Year That Changed the Internet*, THE ATLANTIC (Dec. 28, 2020), <https://www.theatlantic.com/ideas/archive/2020/12/how-2020-forced-facebook-and-twitter-step/617493/>.

14. See, e.g., *OPEN BY DEFAULT AND MODERN, EASY TO USE FORMATS*, GOV’T OF CAN. (2016), <https://open.canada.ca/en/content/open-default-and-modern-easy-use-formats>.

15. Centre for Law & Democracy & Int’l Media Support [IMS], *Freedom of Expression Briefing Note Series*, at 30–46 (July 2014), <https://www.mediasupport.org/wp-content/uploads/2015/02/foe-briefingnotes-ims-cld.pdf>.

Part I of this Article proceeds by tracing the platforms' evolution from a relatively hands-off approach to the current posture as de facto stewards of the global public discourse. Part II considers the implications of this shift, and the failure of traditional transparency models, to argue that the platforms' role carrying out an essentially public function of managing the public discourse, combined with the collective interest in robust public oversight, justifies a quasi-governmental standard of transparency for their content moderation operations. Part III discusses freedom of information and right to information rules, and introduces a model for applying these standards to private sector platforms. Part IV concludes.

I. GUN-SHY AUTOCRATS

Barely a decade ago, Google made news for refusing to remove obscene webpages from the results of search queries for Republican politician Rick Santorum, claiming that such a direct intervention would be antithetical to their role as a conduit for third-party content.¹⁶ The company insisted that the material they hosted was merely “a reflection of the content and information that is available on the web,” and that such direct action would be inappropriate, except where the content at issue was plainly illegal.¹⁷

Today, such a laissez-faire approach would be unthinkable, as tech platforms have shifted from “the free speech wing of the free speech party”¹⁸ to a much more nuanced understanding of their role and responsibilities with respect to the content they host.¹⁹ However, even as platforms have demonstrated a reasonable willingness to combat child-sexual abuse material

16. Tarleton Gillespie, *Algorithmically Recognizable: Santorum's Google Problem, and Google's Santorum Problem*, 20 INFO. COMM. & SOC. 63, 70 (2017).

17. Alexander Burns, *Santorum: Google spreads 'filth'*, POLITICO (Sept. 20, 2011, 2:49 PM), <https://www.politico.com/story/2011/09/santorum-google-spreads-filth-063952>.

18. Emma Barnett, *Twitter Chief: We Will Protect Our Users from Government*, THE TELEGRAPH (Oct. 18, 2011), www.telegraph.co.uk/technology/twitter/8833526/Twitter-chief-We-will-protect-our-users-from-Government.html.

19. *Facebook, Social Media Privacy, and the Use and Abuse of Data: Hearing Before the Subcomm. on the Judiciary and the Subcomm. on Commerce, Sci. and Transp.*, 115th Cong. 1 (2018) (testimony of Mark Zuckerberg, Chairman and Chief Executive Officer, Facebook).

(“CSAM”),²⁰ hate speech,²¹ and even “terrorist speech,”²² policing false speech, or misinformation, has traditionally been the third-rail of content moderation.²³ In part, the reticence to branch into this space may be attributable to the companies’ traditional connection to American constitutional values, which are comparatively protective of the right to speak falsely.²⁴ However, this stance may also be understood as a political calculation, in response to the potential blowback from taking positions on what is or is not misinformation.

Content moderation at scale is hard, and inevitably results in errors, both in terms of potentially rule-breaking content which goes unnoticed (false negatives), and in terms of acceptable content which is wrongly flagged (false positives).²⁵ Moderation systems can be calibrated so that they err more on one side or another, but a certain amount of “by-catch” is inevitable, particularly as much of the process is either carried out by automated systems, or by human reviewers who are often underpaid, and inadequately trained for the difficult and contextual policy-interpretation exercises they face.²⁶ Although the major platforms have instituted various appeals structures aimed at rectifying erroneous moderation decisions, most notably the Facebook Oversight Board, the massive scale on which these decisions must be made effectively precludes a meaningful standard of review for most decisions, meaning that all categories

20. See e.g., Nikola Todorovic & Abhi Chaudhuri, *Using AI to Help Organizations Detect and Report Child Sexual Abuse Material Online*, THE KEYWORD (Sept. 3, 2018), <https://www.blog.google/around-the-globe/google-europe/using-ai-help-organizations-detect-and-report-child-sexual-abuse-material-online>; Facebook Security, *Want to know how Facebook uses PhotoDNA? Read a recent blog post by the head of our Safety Team*, FACEBOOK (Aug. 10, 2011), <https://www.facebook.com/security/posts/want-to-know-how-facebook-uses-photodna-read-a-recent-blog-post-by-the-head-of-o/234737053237453/>.

21. Monika Bickert & Chris Sonderby, *Explaining Our Community Standards and Approach to Government Requests*, FACEBOOK: NEWSROOM (Mar. 15, 2015), <https://about.fb.com/news/2015/03/explaining-our-community-standards-and-approach-to-government-requests/>.

22. Shane Harris, *Social Media Companies Scramble to Block Terrorist Video of Journalist’s Murder*, FOREIGN POL’Y (Aug. 19, 2014), <https://foreignpolicy.com/2014/08/20/social-media-companies-scramble-to-block-terrorist-video-of-journalists-murder/>.

23. Tom McCarthy, *Zuckerberg says Facebook won’t be ‘arbiters of truth’ after Trump threat*, THE GUARDIAN (May 28, 2020), <https://www.theguardian.com/technology/2020/may/28/zuckerberg-facebook-police-online-speech-trump>.

24. See, e.g., *United States v. Alvarez*, 567 U.S. 709, 716 (2012). These protections are not absolute, of course, and there are a number of constitutionally valid restrictions which apply to false speech, from defamation, to false advertising, to financial misrepresentations.

25. Internet Society North America Bureau, *COMO Summit 3 - Under the Hood: UGC Moderation (Part 1)*, YOUTUBE (May 7, 2018), <https://www.youtube.com/watch?v=stB23tNBI2o>.

26. SARAH T. ROBERTS, *BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA* 6, 14 (Yale University Press ed. 2019).

of moderation necessarily lead to collateral damage.²⁷ However, there are practical differences in how the consequences of these decisions manifest.

In the context of child-sexual abuse material, for example, adverse moderation decisions that are based on false-positives either fly under the radar, or attract little mainstream sympathy.²⁸ A similar point might be made about the platforms' increasing hardline against "terrorist speech."²⁹ Although these restrictions have had a significant adverse impact against Arabic-speakers, with real and harmful consequences for journalists and civil society voices who operate in more repressive parts of the world,³⁰ their impacts are nonetheless relatively far-removed from the core political constituencies which are the main focus of platforms' attention.³¹

On the other hand, the question of addressing "misinformation" or "fake news" as a standalone harm drags the platforms into a morass of definitional and practical challenges.³² The problem will be apparent to anyone familiar with how the term "fake news" has morphed in U.S. politics, and is now commonly used by a segment of the population to dismiss any narrative which they find problematic.³³ The drift in usage to suit political convenience is typical of the

27. Evelyn Douek, *The Facebook Oversight Board's First Decisions: Ambitious, and Perhaps Impractical*, LAWFARE (Jan. 28, 2021), <https://www.lawfareblog.com/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical>.

28. See, e.g., Justin Paine & John Graham-Cumming, *Announcing the CSAM Scanning Tool, Free for All Cloudflare Customers*, CLOUDFLARE BLOG (Dec. 18, 2019), <https://blog.cloudflare.com/the-csam-scanning-tool/>. This blog from Cloudflare takes for granted that false positives are "the lesser evil." The authors frame the challenge of false positives from their own CSAM scanning system not in terms of any harm to legitimate speech, but rather as a technical requirement to avoid overtaxing the reporting resources.

29. Belkis Wille, "Video Unavailable" Social Media Platforms Remove Evidence of War Crimes, HUM. RTS. WATCH (Sept. 10, 2020), <https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes>.

30. Svea Windwehr & Jillian C. York, *One Database to Rule Them All: The Invisible Content Cartel that Undermines the Freedom of Expression Online*, ELEC. FRONTIER FOUND. (Aug. 27, 2020), <https://www.eff.org/deeplinks/2020/08/one-database-rule-them-all-invisible-content-cartel-undermines-freedom-1>.

31. "Facebook has allowed major abuses of its platform in poor, small and non-western countries in order to prioritize addressing abuses that attract media attention or affect the US and other wealthy countries." Julia Carrie Wong, *Revealed: the Facebook loophole that lets world leaders deceive and harass their citizens*, GUARDIAN (Apr. 12, 2021), <https://www.theguardian.com/technology/2021/apr/12/facebook-loophole-state-backed-manipulation>. See generally, Chinmayi Arun, *Facebook's Faces*, 135 HARV. L. REV. F. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3805210.

32. See, e.g., Nellie Bowles, *The Complex Debate Over Silicon Valley's Embrace of Content Moderation*, N.Y. TIMES (June 30, 2020), <https://www.nytimes.com/2020/06/05/technology/twitter-trump-facebook-moderation.html>.

33. Margaret Sullivan, *What it really means when Trump calls a story 'fake news'*, WASH. POST (Apr. 13, 2020, 1:34 PM), <https://www.washingtonpost.com/lifestyle/media/what-it-really->

history of “false news” and “misinformation” restrictions, which have traditionally been a weapon for authoritarians to target critical journalists or civil society opponents.³⁴ In the absence of a generally accepted set of human rights standards, platforms seeking to police “misinformation” face the conceptually difficult task of trying to define when garden-variety lies rise to a threshold where moderation actions are justified. One assumes that platforms have no interest in being in the business of scrutinizing every user who fibs about their height, or their CV, or whether they actually went to a particular restaurant.³⁵ So when does a lie, or a genuinely expressed false statement, become “misinformation”?

Applying these questions to the political dimension opens an additional can of worms, given that politicians, in general, are expected to be flexible in massaging or curating facts to seek their chosen narrative, or to hype their own accomplishments, or to diminish those of their opponents.³⁶ Determining when normal political spin crosses over into the territory of lying is a difficult and controversial task, and one which is open to debate and interpretation.³⁷ At times, the truth itself can shift under our feet, transforming a particular narrative from being “misinformation” to being a reasonably posited theory.³⁸ Given that imposing a broad rule against false speech, or even against misinformation, would necessitate the platforms inserting themselves into enormously controversial political debates on a daily basis, it is easy to understand why the companies took such pains to steer clear of acting as “arbiters of truth.”³⁹

The platforms’ relatively dismissive attitude towards misinformation even persisted past the two watershed events in 2016 that served to galvanize public consciousness around the threat of “fake news,” namely the U.S. election of that

means-when-trump-calls-a-story-fake-news/2020/04/13/56fbc2c0-7d8c-11ea-9040-68981f488eed_story.html.

34. Michael Karanicolas, *Even in a Pandemic, Sunlight Is the Best Disinfectant: COVID-19 and Global Freedom of Expression*, 22 OR. REV. INT’L L. 1, 3–4 (2021).

35. Jordan Liles, *Did Scott Walker Tweet a Year-Old Pizza Picture Twice?*, SNOPES (Dec. 28, 2020), <https://www.snopes.com/fact-check/scott-walker-pizza/>.

36. Conor McCann, *As a matter of fact: Fact-checker extraordinaire Daniel Dale reflects on the Trump era*, CBC NEWS (Jan. 23, 2021, 8:00 AM), <https://www.cbc.ca/news/canada/newfoundland-labrador/daniel-dale-trump-era-1.5884246>.

37. *Lies? False Claims? When Trump’s Statements Aren’t True*, N.Y. TIMES (June 25, 2018), <https://www.nytimes.com/2018/06/25/reader-center/donald-trump-lies-falsehoods.html>.

38. Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, FACEBOOK: NEWSROOM (May 26, 2020), <https://about.fb.com/news/2020/04/covid-19-misinfo-update/#removing-more-false-claims>.

39. Callum Borchers, *Twitter Executive on Fake News: ‘We Are Not the Arbiters of Truth’*, WASH. POST (Feb. 8, 2018), <https://www.washingtonpost.com/news/the-fix/wp/2018/02/08/twitter-executive-on-fake-news-we-are-not-the-arbiters-of-truth/>; Supraja Srinivasan, *We Don’t Want to be Arbiters of Truth: YouTube CBO Robert Kyncl*, ECON. TIMES (Mar. 24, 2018), <https://economictimes.indiatimes.com/internet/we-dont-want-to-be-arbiters-of-truth-youtube-cbo-robert-kyncl/articleshow/63438805.cms>.

year, and the Brexit vote.⁴⁰ Evidence of a change in tone began to manifest in 2018, at least with regard to misinformation related to foreign election interference.⁴¹ However, the real shift came with the Covid-19 crisis, as high profile complaints about an “infodemic” around both the virus and the vaccine helped to crystalize the harms that flow from online misinformation.⁴² YouTube, which historically exercised a strong preference towards keeping false material up, announced in April 2020 that its policy for Covid-19 would be to remove anything which was “medically unsubstantiated” according to World Health Organization recommendations.⁴³ Facebook similarly strengthened enforcement of misinformation policies in relation to the pandemic.⁴⁴ Twitter introduced a number of measures including an expanded definition of harmful content in its Terms of Service.⁴⁵ The disputes around the 2020 U.S. election, and the aggressive response of platforms to purge accounts promoting the “big lie” of widespread fraud, further solidified this stance.⁴⁶

Although many of the formal policy changes at platforms were specifically directed against these two crises,⁴⁷ there is a common mission creep associated with stronger moderation postures.⁴⁸ Once platforms demonstrate a willingness, and an ability, to moderate aggressively against a particular type of content, it becomes more difficult to hold the line against pressure to target other parallel examples of problematic speech.

40. Casey Newton, *Zuckerberg: The Idea that Fake News on Facebook Influenced the Election is 'Crazy'*, VERGE (Nov. 10, 2016), <https://www.theverge.com/2016/11/10/13594558/mark-zuckerberg-election-fake-news-trump>.

41. Sheera Frenkel & Mike Isaac, *Inside Facebook's Election 'War Room'*, N.Y. TIMES (Sept. 19, 2018), <https://www.nytimes.com/2018/09/19/technology/facebook-election-war-room.html>.

42. UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis, UN DEP'T OF GLOB. COMM. (Mar. 31, 2020), <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-'infodemic'-misinformation-and-cybercrime-covid-19> [<https://perma.cc/GE69-4LCR>].

43. *Coronavirus: YouTube bans 'medically unsubstantiated' content*, BBC NEWS (Apr. 22, 2020), <https://www.bbc.com/news/technology-52388586>.

44. Rosen, *supra* note 38.

45. Vijaya Gadde (@Vijaya) & Matt Derella (@Derella), *An update on our continuity strategy during COVID-19*, TWITTER: BLOG (updated Apr. 1, 2020), https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html.

46. Makena Kelly, *Twitter expands misinformation rules to cover premature election results*, VERGE (Sept. 10, 2020), <https://www.theverge.com/2020/9/10/21431027/twitter-misinformation-election-president-donald-trump-policy-rules>.

47. *Spam, deceptive practices, & scams policies*, GOOGLE, <https://support.google.com/youtube/answer/2801973> (last visited Aug. 19, 2021); *Misinformation Policies*, GOOGLE, <https://support.google.com/youtube/answer/10834785> (last visited Aug. 19, 2021) (providing YouTube's Community Guidelines which specifically prohibits false claims of impropriety with regards to the 2020 U.S. presidential election); *see also Elections misinformation policies*, GOOGLE, <https://support.google.com/youtube/answer/10835034> (last visited Aug. 19, 2021) (same).

48. Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1051 (2018).

While there have been arguments for years that the platforms should be subjected to greater scrutiny and accountability, the recent shift into taking a more active stand against misinformation is notable for two reasons. First, as noted earlier in this section, determinations around what constitutes “misinformation” are far more challenging to implement than, say, a ban against nudity. Although the latter can be the subject of controversy and debate,⁴⁹ “misinformation” is a far more inherently contextual category, which therefore creates a greater impetus for consultation and public dialogue in determining how to apply the rules against particular types of content. This includes not only assessing the veracity of a claim, but also considering its dissemination and likely impact. As a result, the change in posture means that platforms face a heightened impetus to seek robust and continuous stakeholder feedback into their moderation decisions, in order to ensure that they are making the right calls in addressing such controversial and heavily localized questions.

Second, there is an argument that moderation functions targeting “misinformation” are categorically different than moves to remove pornographic content, or harassment, insofar as these push the platforms into the center of political and public policy debates. While content moderation has, to a greater or lesser degree, been a feature of the platforms’ operations virtually throughout their history, they have traditionally been relatively gun shy about targeting political speech, or being seen to insert themselves directly into the political discourse.⁵⁰ Now that ship has sailed, and there is a significantly heightened public interest in tracking and monitoring their operations, with a concomitant need to reconsider their obligations to be accountable to their users and, more generally, to the public at large, for the type of moderation structure they are enforcing, and the impact of these decisions on the political discourse.

II. “WHO ELECTED YOU?”

The enormous expansion of platforms’ power and influence has had a profound impact on the regulatory dynamic surrounding freedom of expression.⁵¹ In response, there have been a range of proposals for resolving the inherent accountability deficit underlying the platforms’ role in managing and curating the online discourse, from models based around traditional public

49. Mark Scott & Mike Isaac, *Facebook Restores Iconic Vietnam War Photo It Censored for Nudity*, N.Y. TIMES (Sept. 9, 2016), <https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html>; Jenna Wortham, *Facebook Won’t Budge on Breastfeeding Photos*, N.Y. TIMES: BITS (Jan. 2, 2009), <https://bits.blogs.nytimes.com/2009/01/02/breastfeeding-facebook-photos/>.

50. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1618–21 (2018).

51. Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011 *passim* (2018).

service media,⁵² to modifying common carriage rules,⁵³ to calls for the largest platforms to be broken up⁵⁴ or even nationalized.⁵⁵ However, while a stronger role for governments in this space may seem like an intuitive solution to the existing accountability deficit, this would not be a practical option.

Many progressive democratic governments, including the United States, have constitutional rules which would preclude them from moderating content the way that platforms do.⁵⁶ Debates around content moderation generally do not revolve around material which is actually illegal, and where platform takedowns are relatively uncontroversial, but rather focus on “awful but lawful” speech.⁵⁷ “Misinformation” is a perfect example here, given that much of this category of speech would be constitutionally protected in the United States.⁵⁸ A number of other courts around the world have rejected government efforts to prohibit content on similar grounds.⁵⁹

There are good reasons underlying the reticence against allowing governments to aggressively police false speech, insofar as these laws, where they exist, are routinely used to jail journalists, opposition critics, or anyone else whose narrative runs counter to the official government line.⁶⁰ The platforms’

52. Ethan Zuckerman, *The Case for Digital Public Infrastructure*, KNIGHT FIRST AMEND. INST. AT COLUM. UNIV. (Jan. 17, 2020), <https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure>.

53. Eugene Volokh, *Justice Thomas Suggests Rethinking Legal Status of Digital Platforms*, REASON: THE VOLOKH CONSPIRACY (Apr. 5, 2021), <https://reason.com/volokh/2021/04/05/justice-thomas-suggests-rethinking-of-legal-status-of-digital-platforms/>.

54. Chris Hughes, *It’s Time to Break Up Facebook*, N.Y. TIMES (May 9, 2019), <https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>.

55. Blayne Haggart, *Why Not Nationalize Facebook?*, NAT’L POST (Mar. 31, 2018), <https://nationalpost.com/pmn/news-pmn/why-not-nationalize-facebook>.

56. See, e.g., *United States v. Alvarez*, 567 U.S. 709, 729–30 (2012); *R. v. Zundel*, [1992] 2 S.C.R. 731, 733 (Can.).

57. Daphne Keller, *Six Constitutional Hurdles For Platform Speech Regulation*, CTR. FOR INTERNET & SOC’Y: BLOG (Jan. 22, 2021, 6:49 AM), <https://cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation>.

58. *Alvarez*, 567 U.S. at 718.

59. *Zundel*, 2 S.C.R. at 733; *Onyango-Obbo & Mwenda v. Att’y Gen. of Uganda*, Constitutional Appeal No. 2 of 2002, [2004] UGSC 81, Judgement (Uganda); *Chavunduka & Others v. Minister of Home Affs. & Another*, SC36/2000, [2000] JOL 6540 (ZS) (Zim.); *Chipenzi & Others v. The People*, (2014) 112 ZMHC J1, J23–J25 [High Ct.] (Zam.); *Hector v. Att’y Gen. of Ant. & Barb.*, [1990] Privy Council Appeal No. 32/1988, Judgement, 5–6, (E. Caribbean Sup. Ct.).

60. See, e.g., *Egypt: Activists Arrested in Dawn Raids*, HUM. RTS. WATCH (May 31, 2018), <https://www.hrw.org/news/2018/05/31/egypt-activists-arrested-dawn-raids> [https://perma.cc/5VE C-FUPZ]; *Sesupo Rantsimako, BPF Spokesman Arrested for Fake News*, BOTS. GAZETTE (Apr. 16, 2020), <https://www.thegazette.news/news/bpf-spokesman-arrested-for-fake-news/30891/#.XtgXsy2Q1p8> [https://perma.cc/AXQ2-6ACB]; *Algeria Blocks 3 News Websites and Criminalizes False News*, COMM. TO PROTECT JOURNALISTS (Apr. 22, 2020), <https://cpj.org/2020/04/algeria-blocks-3-news-websites-and-criminalizes-fa/> [https://perma.cc/7DTN-UYV9]; *Cambodia: Reporter Jailed for Quoting Hun Sen on COVID-19*, HUM. RTS. WATCH (Apr. 10, 2020),

status as private entities, operating at arm's length from governments, is what allows them greater freedom to act against harmful content which is constitutionally protected. A stronger direct role for governments in enforcing content restrictions would necessarily result in less aggressive moves to curate the discourse, particularly to promote integrity and veracity, which are challenging areas for governments to wade in to.

In the absence of a credible alternative to having the platforms carry out this moderation function, attention has focused on pushing for improved governance at the companies themselves and, in particular, on expanding the transparency and accountability of platforms' decision-making. The most ambitious proposals have emanated from Europe. In particular, the proposed EU Digital Services Act ("DSA") includes requirements for platforms to publish more information about content moderation structures as part of their Terms and Conditions, as well as additional reporting requirements, including with regard to their risk mitigation efforts around misinformation, and an obligation to provide data access to vetted researchers.⁶¹ These would build on the existing European Commission Code of Practice on Disinformation, which also includes commitments on transparency in political advertising,⁶² and on national frameworks, particularly Germany's *Netzwerkdurchsetzungsgesetz* ("NetzDG"), which requires reporting on a range of parameters related to the takedown of illegal material.⁶³ All of these regulatory measures are in addition to the proactive transparency reporting which has, by now, become an industry standard, and which is continuously expanding.⁶⁴

The default towards increasing transparency to improve the legitimacy and accountability of content moderations systems is logical on a number of levels. First, content moderation systems are heavily dependent on third party oversight, both to flag problematic content and, more broadly, in order to drive structural improvements by noting where the system is failing to perform as advertised. This last point has been a consistent feature of private sector content moderation efforts, whereby civil society watchdogs have traditionally had to

<https://www.hrw.org/news/2020/04/10/cambodia-reporter-jailed-quoting-hun-sen-covid-19> [<https://perma.cc/T8UD-E7FS>]; *Newspaper Fined for "False Publication" on COVID-19 Status of Inmates*, MEDIA FOUND. FOR W. AFR. (Apr. 27, 2020), <https://www.mfwa.org/issues-in-focus/newspaper-fined-for-false-publication-about-covid-19/> [<https://perma.cc/FJ2H-C6UC>].

61. *European Commission Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, at arts. 12, 23, 31, COM (2020) 825 final (Dec. 15, 2020).

62. EUR. COMM'N, *EU CODE OF PRAC. ON DISINFORMATION* (2018), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454 [<https://perma.cc/W3T2-ZRKF>].

63. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [NetzDG]* [Network Enforcement Act], Oct. 1, 2017, at § 2 (Ger.), https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf.

64. *See, e.g., Google Transparency Report*, GOOGLE, <https://transparencyreport.google.com/about?hl=en> (last visited Dec. 20, 2021).

rely on media attention and public relations threats in order to push for substantive change.⁶⁵ However, the ability of third parties to play this oversight role is necessarily dependent on their ability to access accurate and comprehensive information about how the systems are functioning. While this dynamic has been in place for years, the pivot towards policing more locally contextual and political forms of speech, particularly misinformation, leads to a much greater need for substantive outside oversight over platforms' operations, particularly in markets which are further removed from their main geographic areas of focus (i.e. anywhere outside of North America and Western Europe).⁶⁶ At the very least, an accurate understanding of how moderation systems operate is a necessary precondition to offering meaningful feedback on their efficacy.

Transparency is also a typical avenue for generating public trust, and for fostering perceptions of legitimacy. Jonathan Zittrain has described the current era of digital governance as being defined by a need to develop robust procedural standards for settling controversial questions.⁶⁷ Transparency is at the center of this challenge, since even an exceptionally robust and thoughtful decision-making process would be meaningless towards building legitimacy if it were not accompanied by an effective avenue for allowing the public to see and understand it. The inverse to this is also true, insofar as the absence of information about how decisions are made can generate theories of bias, discrimination, or conspiracy as to the reasons for an adverse decision.⁶⁸ A recent study noted persistent confusion among the subjects of moderation decisions, which naturally degrades trust and belief in the integrity of these systems.⁶⁹

65. See, e.g., *Myanmar groups accuse Facebook of failing to curb hate speech*, CBS NEWS (Apr. 6, 2018), <https://www.cbsnews.com/news/myanmar-groups-accuse-facebook-of-failing-to-curb-hate-speech/>; Jillian C. York & David Greene, *Amid Systemic Censorship of Palestinian Voices, Facebook Owes Users Transparency*, ELEC. FRONTIER FOUND. (May 25, 2021), <https://www.eff.org/deeplinks/2021/05/amid-systemic-censorship-palestinian-voices-facebook-owes-users-transparency>.

66. Julia Carrie Wong, *Revealed: the Facebook loophole that lets world leaders deceive and harass their citizens*, GUARDIAN (Apr. 12, 2021), <https://www.theguardian.com/technology/2021/apr/12/facebook-loophole-state-backed-manipulation>.

67. Jonathan Zittrain, *Three eras of digital governance*, VÖLKERRECHTSBLOG (Nov. 27, 2019), <https://voelkerrechtsblog.org/three-eras-of-digital-governance/>.

68. Of course, suspicions of algorithmic bias can also be well founded. This is not to suggest that bias does not exist in content moderation structures, but rather that, in the absence of accurate and comprehensive information, it is impossible to discern where instances of genuine discrimination are or are not manifesting.

69. Nicolas Suzor et al., *What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation*, 13 INT'L J. OF COMM'N 1526, 1532 (2019).

However, there is a key structural problem which underlies all of the recent efforts to promote transparency in moderation structures.⁷⁰ At the most fundamental level, transparency reporting relies on companies' messaging arms to provide a window into how they are operating.⁷¹ For outsiders, there will always be a challenge connected to this, insofar as the public will never be certain as to whether the results they receive present a complete, accurate, and unvarnished picture of what is actually going on.⁷² One particularly notorious example in this regard is from the Global Network Initiative ("GNI"), which faced a significant challenge following leaks regarding the extent of platforms' collaboration with government surveillance systems.⁷³ Though the GNI continues to operate and, by most accounts, provides a valuable and useful service in bringing civil society, academic, and industry voices to discuss issues of common concern, there remains significant controversy about what level of collaboration actually took place, despite the GNI requirements for regular and independent audits on precisely this kind of question.⁷⁴

A related challenge with transparency reporting is that it often fails to adequately deliver the content or formats of information which are useful to researchers, or other stakeholders who are seeking to use the material to carry out independent oversight over the platforms' activities. Ivar Hartmann, in a recent paper published by the Wikimedia/Yale Law School Initiative on Intermediaries and Information, noted the challenges in obtaining meaningful data around online advertising, due to the incomplete and inconsistent nature of disclosures across the different platforms' ad libraries.⁷⁵ There have been similar criticisms against the transparency reports published in under Germany's

70. The possible exception to this is the DSA's new requirement to provide outside researchers with direct access to the platforms' data, though the devil will be in the details.

71. Christopher Parsons, *The (in) effectiveness of voluntarily produced transparency reports*, 58 BUS. & SOC'Y 103, 106 (2019).

72. Mike Ananny & Kate Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 NEW MEDIA & SOC'Y 973, 985 (2018).

73. Letter from Danny O'Brien, Int'l Dir., & Jillian C. York, Dir. for Int'l Freedom of Expression, to Jermyn, Susan, & David, Global Network Initiative (Oct. 10, 2013) (on file with the Electronic Frontier Foundation).

74. Declan McCullagh, *NSA surveillance retrospective: AT&T, Verizon never denied it*, CNET (June 12, 2013), <https://www.cnet.com/news/nsa-surveillance-retrospective-at-t-verizon-never-denied-it/>. In 2016, the author also received direct feedback to a piece he had published from a representative of one of the companies, claiming the NSA slides depicting industry collaboration were being taken out of context, and denying that there had been significant collaboration. The author has no means of assessing the veracity of these claims, and this interaction is only mentioned in the context of establishing the platform's position vis-à-vis the Snowden disclosures.

75. Ivar Hartmann, *Combining Ad Libraries with Fact Checking to Increase Transparency of Misinformation*, in TACKLING THE "FAKE" WITHOUT HARMING THE "NEWS", A PAPER SERIES ON REGULATORY RESPONSES TO MISINFORMATION 67 (Michael Karanicolas ed., 2021).

NetzDG law.⁷⁶ Researchers have complained that the mandatory public audits published under the auspices of the Federal Trade Commission (“FTC”) are “so vague or duplicative as to be meaningless.”⁷⁷ Internationally, the situation is even more challenging, as researchers, particularly from the developing world, face a constant struggle to find accurate data regarding how policies are being implemented, or even which policies are operative in a particular region.⁷⁸

The challenges in obtaining meaningful transparency through existing disclosure regimes are particularly relevant in the context of parallel moves, by the platforms, to insulate themselves from other forms of internal and external accountability. This includes implementing aggressive counter-measures against tools designed to study their operations.⁷⁹ Even the platforms’ internal governance structures are designed, in many cases, with an eye to curtailing meaningful oversight. While traditional understandings of corporate governance rely on shareholders to exercise some level of oversight over the company’s direction, two of the biggest online platforms operate under a dual-class share structure which effectively insulates senior management from any meaningful accountability to shareholders.⁸⁰ Similarly, a number of platforms, particularly Facebook, have the same person acting as CEO and as Chair of the Board, effectively making that person accountable to themselves.⁸¹ These governance choices are ill-advised in any corporate context, but are particularly problematic given the enormous public importance of the role that platforms play.⁸² The

76. Amélie Heldt, *Reading between the Lines and the Numbers: An Analysis of the First NetzDG Reports*, 8 INTERNET POL’Y REV., no. 2, 2019, <https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports>.

77. Megan Gray, *Understanding and Improving Privacy “Audits” under FTC Orders*, STAN. CTR. FOR INTERNET & SOC’Y, Apr. 2018, at 4, <https://cyberlaw.stanford.edu/files/blogs/white%20paper%204.18.18.pdf>.

78. Sergei Hovyadinov, *Toward a More Meaningful Transparency: Examining Twitter, Google, and Facebook’s Transparency Reporting and Removal Practices in Russia*, SSRN (Nov. 30, 2019), <https://ssrn.com/abstract=3535671>; Agustina Del Campo, *Social media in Latin America: Caught between a rock and a hard place*, GLOB. VOICES (Sept. 17, 2020), <https://globalvoices.org/2020/09/17/social-media-in-latin-america-caught-between-a-rock-and-a-hard-place/>.

79. Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools — Including Ours*, PROPUBLICA (Jan. 28, 2019), <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools>; Issie Lapowsky, *Platforms vs. PhDs: How tech giants court and crush the people who study them*, PROTOCOL (Mar. 19, 2021), <https://www.protocol.com/nyu-face-book-researchers-scraping>.

80. Nathalie Maréchal et al., *Better Processes Lead to Better Outcomes: Corporate Governance as a Tool to Address Misinformation*, in TACKLING THE “FAKE” WITHOUT HARMING THE “NEWS”, A PAPER SERIES ON REGULATORY RESPONSES TO MISINFORMATION 10, 24–25 (Michael Karanicolas ed., 2021).

81. *Id.* at 18.

82. Marc Goergen et al., *On the Choice of CEO Duality: Evidence from a Mandatory Disclosure Rule* (CFR Working Paper No. 18-06, 2018), <https://www.econstor.eu/bitstream/10419/191061/1/1045535931.pdf>; Kosmas Papadopoulos, *Dual-Class Shares: Governance Risks and*

reflexive use of non-disclosure agreements adds yet another layer of opacity to the platforms' operations,⁸³ as do moves to target internal whistleblowers.⁸⁴

The end result of all of this is that, for all their talk of transparency, the platforms have managed to insulate themselves from meaningful independent oversight “[t]hrough code and through contract.”⁸⁵ What is needed is a major reconceptualization of what transparency means in the context of these entities. Rather than relying on the platforms' largesse in delivering scraps of information, this paper argues for an approach to transparency based on best practices from the public sector and, in particular, that information about moderation structures should be “open by default.”⁸⁶ At its core, this requires substantial design changes in how information is managed at the platforms and a mechanism to connect information queries from the public to responses from the platforms: a FOIA for Facebook.

III. A FOIA FOR FACEBOOK

A. *The Right to Information*

The Internet age has provided enormous opportunities for advocates of open and transparent government.⁸⁷ Digital technologies allow information to be stored, sorted, reproduced, and delivered at a level of speed and efficiency that would be unthinkable a generation ago.⁸⁸ These transformations have, in turn, redefined citizens' relationship with data about the entities which govern their lives, both in terms of their capacity to use and process the information, and in terms of their expectations regarding its availability.⁸⁹ In democracies around the world, governments are implementing frameworks where information that they hold is “open by default,” meaning that all government information should

Company Performance, HARV. L. SCH. F. ON CORP. GOVERNANCE (June 28, 2019), <https://corpgov.law.harvard.edu/2019/06/28/dual-class-shares-governance-risks-and-company-performance/>.

83. Katie Canales, *Facebook Moderators, Tasked with Watching Horrific Content, are Demanding an End to NDAs that Promote a 'Culture of Fear and Excessive Secrecy'*, INSIDER (July 23, 2021), <https://www.businessinsider.com/facebook-moderators-letter-zuckerberg-culture-of-fear-nda-2021-7>.

84. Issie Lapowsky, *For Big Tech whistleblowers, there's no such thing as 'moving on'*, PROTOCOL (Apr. 15, 2021), <https://www.protocol.com/big-tech-whistleblowers>.

85. Paddy Leerssen, *The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems*, 11 EUR. J. OF L. & TECH. 1, 15 (2021).

86. See, for example, the first principle of the International Open Data Charter. PRINCIPLES, OPEN DATA CHARTER (2015), <https://opendatacharter.net/principles/>.

87. DAVID BANISAR, *The Right to Information in the Age of Information*, in HUMAN RIGHTS IN THE GLOBAL INFORMATION SOCIETY 73 (Rikke Frank Jørgensen ed., 2006).

88. Cary Coglianese, *The Transparency President? The Obama Administration and Open Government*, 22 GOVERNANCE 529, 535 (2009).

89. Michael Karanicolas, *Understanding the Internet as a Human Right*, 10 CAN. J.L. & TECH. 264, 264–65 (2012).

be treated as accessible to the public, either through proactive disclosure or via information requesting mechanisms.⁹⁰ This idea is at the core of most modern right to information legislation, and a foundational principle underlying the recognition of the right to information as a human right.⁹¹

In practice, freedom of information or right to information legislation grants the public with a broad ability to either request specific documents that are in the hands of an agency to which the law applies (for instance, a copy of the budget for the previous fiscal year), or to formulate a question to which the agency must respond (for instance, “what proportion of police stops target visible minorities?”).⁹² This right of access is generally interpreted broadly, at least in the context of progressive laws.⁹³ However, the right of access is not absolute, and right to information or freedom of information laws will generally include a list of categories of material which may be legitimately withheld from the public, or which should be released in redacted form, such as information whose disclosure would be harmful to law enforcement, national security, or the privacy of a natural person.⁹⁴

Due to the political sensitivity that often accompanies information requests, another core principle of the right to information is that these systems should be administered with as much independence as possible, including through creating separate, specialized offices to handle requests and, ideally, an independent administrative oversight body to hear appeals against cases where information may have been wrongly withheld.⁹⁵

Right to information or freedom of information laws were recognized as one of the main precepts of good administration as early as 1996, and provide a number of key benefits to the institutions which enact them.⁹⁶ These include promoting trust in institutions and improving their relations with the public.⁹⁷ There is also evidence tying the implementation of freedom of information and

90. As of June 2021, twenty-four national governments and sixty-one cities and local governments have endorsed the International Open Data Charter, *supra* note 86, which includes a core commitment to making information open by default.

91. Claude Reyes et al. v. Chile, Merits, Reparations, and Costs, Judgement, Inter-Am. Ct. H.R. (ser. C) No. 151 (Sept. 19, 2006); Társaság A Szabadságjogokért v. Hungary, App. No. 37374/05, 53 Eur. H.R. Rep. 3 (2009).

92. TOBY MENDEL, FREEDOM OF INFORMATION: A COMPARATIVE LEGAL SURVEY 31 (2d ed. 2003).

93. *Id.*

94. *Id.* at 35–37; MICHAEL KARANICOLAS & TOBY MENDEL, CTR. FOR L. & DEMOCRACY, ENTRENCHING RTI: CONSTITUTIONAL PROTECTIONS OF THE RIGHT TO INFORMATION 6–8 (2012), www.law-democracy.org/wp-content/uploads/2012/04/Const-Report-with-Annex.pdf.

95. MENDEL, *supra* note 92, at 38.

96. Alasdair Roberts, *Structural Pluralism and the Right to Information*, 51 U. TORONTO L. J. 243, 244 (2001).

97. Maria Cucciniello & Greta Nasi, *Transparency for Trust in Government: How Effective is Formal Transparency?* 37 INT’L J. OF PUB. ADMIN. 911, 912 (2014).

right to information legislation to better administrative processes, by imposing greater discipline and rigor, and focusing decision-makers' minds on the need to be thorough and consider all relevant factors.⁹⁸ Right to information and freedom of information legislation, where it exists, has also been employed as a key tool to rectify past wrongs or harms, such as human rights abuses.⁹⁹ Similarly, it has become a critical avenue for exposing (and thereby providing an opportunity to rectify) mismanagement and corruption.¹⁰⁰

A key element underlying all of these benefits, which sets the requesting process apart from other forms of disclosure, is that it fosters a dialogue between the agency and the public, as opposed to the one-way flow of curated information that typifies most other transparency processes. Allowing external observers to request information on their terms and ask questions pulls the narrative in a direction which is more difficult for the agency to curate and control, providing a clearer avenue to illuminate emerging problems, including those which may not be evident to the agency itself.¹⁰¹ This supports institutional trust because, although the exposure of problems may be harmful to public perceptions in the short term, a widespread confidence that structural problems will be brought to light and addressed leads to improved attitudes regarding the integrity of the organization.¹⁰²

B. *The Right to Information and the Private Sector*

Considering the scope and nature of benefits associated with a robust right to information or freedom of information system, it is surprising that there has been little academic attention paid to the possibility of applying such a framework to the context of online platforms, especially given the pressing need that these companies face to generate institutional trust (not to mention remedy

98. Australian Law Reform Commission, *Open Government: A Review of the Federal Freedom of Information Act 1982*, (Report No 77, December 1995) 8, <https://www.ag.gov.au/sites/default/files/2020-03/report-40.pdf>.

99. BANISAR, *supra* note 87, at 74.

100. See e.g. Chetan Agrawal, *Right to Information: A Tool for Combating Corruption in India*, 3 J. OF MGMT. & PUB. POL'Y 26, 34 (2012); Anna Clark, *How an Investigative Journalist Helped Prove a City was Being Poisoned With its Own Water*, COLUM. JOURNALISM REV. (Nov. 3, 2015), https://www.cjr.org/united_states_project/flint_water_lead_curt_guyette_aclu_michigan.php; Phlis McGregor & Angela MacIvor, *Black People 3 Times More Likely to be Street Checked in Halifax, Police Say*, CBC NEWS (Jan. 9, 2017 6:00 AM), <https://www.cbc.ca/news/canada/nova-scotia/halifax-black-street-checks-police-race-profiling-1.3925251>.

101. Chetan Agrawal, *Right to Information: A Tool for Combating Corruption in India*, 3 J. OF MGMT. & PUB. POL'Y 26, 34 (2012).

102. A. J. Brown et al., *The Relationship between Transparency, Whistleblowing, and Public Trust*, in RESEARCH HANDBOOK ON TRANSPARENCY 30 (Padideh Ala'i & Robert G. Vaughn eds., 2014).

the harms they have previously been connected to).¹⁰³ Part of the reason for this may be that right to information or freedom of information systems are generally associated with governments, rather than private sector actors.¹⁰⁴ However, while the main applicability of right to information or freedom of information laws are in the context of public sector institutions, there are a growing number of commercial and non-governmental entities which are subject to similar requirements.

The most obvious are state-owned enterprises, which are frequently subject to the same transparency standards as governments, including a requirement to respond to right to information requests.¹⁰⁵ In at least ninety-seven countries, right to information or freedom of information laws extend even further and apply to some purely private sector entities, either because these entities receive or manage significant public resources or because they carry out a “public function.”¹⁰⁶ The definition of this latter category varies among the different laws, and can include, for example, the delivery of education, health, water, social services, or any other function that is contracted out from the government, or whose operations relate to the exercise of constitutional rights and freedoms, or which is otherwise undertaken as a natural monopoly.¹⁰⁷ This is relevant to note not only insofar as it illustrates the broad compatibility between right to information rules and private sector entities, but also because, across much of the world, there is at least a colorable argument that existing right to information legislation already applies to the major online platforms in the context of their local operations, though it is unclear whether this has ever actually been tested.

In addition to instances where private sector entities have been brought under the umbrella of government right to information rules, there are a number of examples of independent entities which have established their own quasi-right to information structures, either because their activities intersect closely with core public interests, or in order to fulfil a perceived need to cultivate public legitimacy in their operations.¹⁰⁸ The Internet Corporation for Assigned Names and Numbers (“ICANN”), a California-based not-for-profit public-benefit

103. David Kaye, *The Republic of Facebook*, JUST SEC. (May 6, 2020), <https://www.justsecurity.org/70035/the-republic-of-facebook/>.

104. Alasdair Roberts, *Structural Pluralism and the Right to Information*, 51 U. TORONTO L. J. 243, 244 (2001).

105. See, e.g., The Freedom of Information Act, 5 U.S.C. § 552(f); Access to Information Act, R.S.C. 1985, c A-1 (Can.); Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental, Código Civil [CC], art. 23, Diario Oficial de la Federación [DOF] 20-04-2015 (Mex.).

106. *Global Right to Information Rating: Indicator 12*, CTR. FOR L. & DEMOCRACY, <https://www.rti-rating.org/country-data/by-indicator/12/> (last visited Sept. 1, 2021).

107. *Id.* (with particular reference to the entries for Armenia, South Africa, Rwanda, Sri Lanka, Ukraine, Kazakhstan, Kenya, and Kyrgyzstan).

108. See, e.g., *Bylaws for Internet Corporation for Assigned Names and Numbers*, ICANN § 1.2 (Nov. 28, 2019), <https://www.icann.org/resources/pages/governance/bylaws-en>.

corporation which coordinates the global domain name system, has its own requesting mechanism which is roughly analogous to a governmental model.¹⁰⁹ The Documentary Information Disclosure Policy (“DIDP”) allows members of the public to request information concerning ICANN’s operational activities “unless there is a compelling reason for confidentiality,” as enumerated in the organization’s “Defined Conditions for Nondisclosure.”¹¹⁰ These include a number of exceptions which run parallel to what one might find in a typical right to information law, such as for attorney-client privilege, information “likely to endanger the life, health, or safety of any individual,” or information from vexatious requesters.¹¹¹ However, there are also some exceptions which are specific to ICANN’s unique position, such as information regarding “changes, modifications, or additions to the [Internet’s] root zone.”¹¹²

A number of international financial institutions, particularly development banks, also have their own mechanisms for facilitating public requests.¹¹³ For the most part, these too tack roughly to what one might expect from a governmental right to information system, including the ability to receive requests from any member of the public, a presumption of openness, and a limited and specific list of exceptions where information may be refused based on enumerated harms that are likely to flow from disclosure.¹¹⁴ As with ICANN, there are a number of exceptions within these access to information policies that are tied to the banks’ unique context, particularly concerning financial or other commercially sensitive information.¹¹⁵

To sum up, although most right to information or freedom of information systems apply primarily to the public sector, there are a number of instances where private sector commercial entities incorporate a similar presumption of openness into their functions, alongside a broad public right of access. This can either be because they have been compelled by law to adopt this posture, or

109. ICANN DOCUMENTARY INFORMATION DISCLOSURE POLICY (Feb. 25, 2012), <https://www.icann.org/resources/pages/didp-2012-02-25-en>.

110. *Id.*

111. *Id.*

112. *Id.*

113. MICHAEL KARANICOLAS & TOBY MENDEL, CTR. FOR L. & DEMOCRACY, OPENNESS POLICIES OF THE INTERNATIONAL FINANCIAL INSTITUTIONS: FAILING TO MAKE THE GRADE WITH EXCEPTIONS 1, 19–22 (2012), <https://www.law-democracy.org/wp-content/uploads/2012/01/IFI-Research-Online-HQ.pdf>.

114. WORLD BANK, THE WORLD BANK POLICY ON ACCESS TO INFORMATION 2–3 (July 1, 2015), <https://documents1.worldbank.org/curated/en/391361468161959342/pdf/548730Access011y0Statement01Final1.pdf>; ASIAN DEV. BANK, ACCESS TO INFORMATION POLICY (Sept. 28, 2018), <https://www.adb.org/sites/default/files/institutional-document/450636/access-information-policy.pdf>; INTER-AM. DEV. BANK, ACCESS TO INFORMATION POLICY 3 (Apr. 26, 2010), idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35167427.

115. *See e.g.*, WORLD BANK, *supra* note 114, at § 17; INTER-AMERICAN DEV. BANK, *supra* note 114, at § 4.1(e), (f), (h).

because they have independently developed their own, bespoke mechanism for receiving access requests. Although there are, at present, no online platforms which incorporate a requesting mechanism into their operations, the structural and public benefits of the right to information are closely related to the current challenges associated with major platforms. These include gaps in public trust and accountability,¹¹⁶ a significant need to account for harms that have taken place,¹¹⁷ and more than anything else, a strong institutional imperative to boost the engagement of external stakeholders in their decision-making processes in order to prevent such harms from happening again, and to improve the targeting of moderation systems at the local level.¹¹⁸

None of this is to suggest that the implementation of a presumption of openness, or of a requesting mechanism, would be an easy transition for the platforms. Real, meaningful transparency is always a difficult sell for the people being subjected to it.¹¹⁹ In the political realm, “transparency for thee but not for me” is a common aphorism.¹²⁰ Opposition leaders often change course dramatically on the importance of transparency after they join the government or majority.¹²¹ Private sector corporations, which are naturally risk averse, may

116. See e.g. *American views: Trust, Media and Democracy*, KNIGHT FOUND. (Jan. 16, 2018), <https://knightfoundation.org/reports/american-views-trust-media-and-democracy/>; Herb Weisbaum, *Trust in Facebook Has Dropped by 66 Percent Since The Cambridge Analytica Scandal*, NBC NEWS (updated Apr. 18, 2018), <https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011>.

117. See e.g. Jenny Domino, *Gambia v. Facebook: What the Discovery Request Reveals about Facebook's Content Moderation*, JUST SEC. (July 6, 2020), <https://www.justsecurity.org/71157/gambia-v-facebook-what-the-discovery-request-reveals-about-facebooks-content-moderation/>.

118. Alison Taylor et al., *Is Stakeholder Engagement the Key to Successful Community Standards?*, BUS. FOR SOCIAL RESP. (June 19, 2019), <https://www.bsr.org/en/our-insights/blog-view/stakeholder-engagement-key-to-successful-community-standards-social-media>.

119. See e.g. Kira Goldenberg, *Obama's Broken Promises on Transparency*, COLUM. JOURNALISM REV. (Oct. 10, 2013), https://archives.cjr.org/behind_the_news/cjp_report_on_us_press_freedom.php; *Justin Trudeau's Promise of Transparency is Starting to Look Empty: Editorial*, TORONTO STAR (Mar. 23, 2017), <https://www.thestar.com/opinion/editorials/2017/03/23/justin-trudeaus-promise-of-transparency-is-starting-to-look-empty-editorial.html>.

120. Josh Gerstein, *Transparency for Thee, But Not For Me?*, POLITICO (Oct. 31, 2014), <https://www.politico.com/blogs/under-the-radar/2014/10/transparency-for-thee-but-not-for-me-197939>; Doug Bandow, *Transparency for Thee But Not for Me*, CATE INST. (Apr. 27, 2009), <https://www.cato.org/blog/transparency-thee-not-me>; Bryan Berky, *Chuck Schumer's View of Transparency: For Thee, But Not for Me*, WASH. EXAM'R (May 4, 2018), <https://www.washingtonexaminer.com/opinion/op-eds/chuck-schumers-view-of-transparency-for-thee-but-not-for-me>.

121. Alexander Quon, *N.S. Premier Calls 2013 Election Promise a 'Mistake,' Experts Say it Shows Lack of Commitment to Transparency*, GLOB. NEWS (Oct. 18, 2018), <https://globalnews.ca/news/4441947/government-transparency-ns-mcneil/>; Elizabeth Thompson, *Access to Information Law Changes Won't Open up PMO, Cabinet Offices*, CBC NEWS (June 19, 2017), <https://www.cbc.ca/news/politics/access-information-government-secrecy-1.4167681>.

balk at measures that open up politically sensitive processes to greater public scrutiny, or even to legal liability.¹²²

As a consequence, getting a system which dramatically reconceptualizes the scope of transparency at major online platforms would likely require some intervention from governments, either through indirect forms of pressure,¹²³ or through a direct legislative mandate, as the European Union appears to be considering through their proposed Digital Services Act.¹²⁴ In the American context, there are a number of existing legislative proposals which mandate greater transparency from platforms by tying these changes to intermediary liability protections, though as of yet, none contemplate a requesting mechanism as part of these changes.¹²⁵ Moreover, such a regulatory move would likely attract a challenge on First Amendment grounds, alleging that it represented a form of compelled speech, whose outcome would be difficult to predict.¹²⁶

Ultimately, the focus of this paper is on exploring the policy-changes which necessitate a reconceptualization of how platform transparency is implemented, and on making the case for the benefits of a requesting mechanism as part of that recalibration. The specific political or advocacy process by which platforms might be cajoled into accepting such a significant operational reform is beyond the scope of this analysis. However, the next section considers how a requesting mechanism might be developed which adapts to the specific context in which the platforms operate, and which, as far as possible, mitigates potential objections to its implementation.

C. *Developing a Framework for Implementation*

Among the private sector entities which have implemented an information requesting mechanism, the procedures for access usually cleave closely to what one might find in a government-side freedom of information or right to information system. Typically, a member of the public merely needs to send a written query, usually via email or an automated electronic form, which specifies

122. Facebook's pushback against even limited requests for data regarding its operations in Myanmar are a good illustration of this challenge. See Poppy McPherson, *Facebook rejects request to release Myanmar officials' data for genocide case*, REUTERS (Aug. 6, 2020), <https://www.reuters.com/article/us-myanmar-facebook-idUSKCN2521PI>.

123. Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 50, 57 (2015).

124. *European Commission Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, at arts. 12, 23, 31, COM (2020) 825 final (Dec. 15, 2020).

125. See, e.g., Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020).

126. Robert Post, *NIFLA and the Construction of Compelled Speech Doctrine*, IND. L. J. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3798562; see also Meese v. Keene, 481 U.S. 465, 480 (1987).

the document or information under request.¹²⁷ Following receipt of the request, the entity commits to respond within a particular timeframe, often thirty calendar days or twenty business days, by either disclosing the information under request or informing the requester that they are withholding the information due to a particular exception to disclosure.¹²⁸ The basic structure of the mechanism is relatively simple, though behind the scenes there is significant organizational and records' management work that needs to take place in order for it to function effectively.¹²⁹

Given the massive public interest in the platforms' operations, which likely far outstrips the level of engagement that entities like ICANN or the World Bank attract, there are likely to be additional concerns as to the scale of requests that these companies might need to deal with, particularly in the immediate aftermath of adopting a requesting policy. One potential avenue to mitigate this would be to impose user fees on the system, which is an approach that governments themselves have employed.¹³⁰ While there are concerns with this strategy in a public sector context, insofar as it suggests that public bodies are failing to properly resource a core democratic accountability function, the unique size and scale in which platforms operate, as well as their commercial context, may make it more palatable as a means of controlling the flow of requests.¹³¹ Two other potential strategies to mitigate this challenge, both of which appear in various other models, are to allow for the dismissal of "frivolous" or "vexatious" requests, and to provide certain limits as to the scope and scale of information under request, in order to keep the queries manageable.¹³² The information management demands associated with moving to an "open by default" framework may also mean that it would be a good idea to bar retroactive requests, and only allow requests for material which was generated after the new policy came into effect, though there are potential tradeoffs to this strategy, insofar as it limits value of information requests to address harms that platforms may have previously been culpable for.

127. See, e.g., *Information Request*, INTER-AMERICAN DEV. BANK, <https://www.iadb.org/en/access-information/information-request> (last visited Dec. 21, 2021); ICANN DOCUMENTARY INFORMATION DISCLOSURE POLICY, *supra* note 109; WORLD BANK, *supra* note 114.

128. ICANN DOCUMENTARY INFORMATION DISCLOSURE POLICY, *supra* note 109; WORLD BANK, *supra* note 114; INTER-AM. DEV. BANK, *supra* note 114. Both government and non-government requesting policies frequently contain provisions for this time period to be extended if necessary.

129. See Julie Brooks, *Perspectives on the Relationship between Records Management and Information Governance*, 29 RECORDS MGMT. J. 5, 14 (2019).

130. Kristy Kirkup, *Suzanne Legault warns MPs against raising access-to-information user fees*, CBC NEWS (Dec. 5, 2014), <https://www.cbc.ca/news/politics/suzanne-legault-warns-mps-against-raising-access-to-information-user-fees-1.2862491>.

131. MENDEL, *supra* note 92, at 38–39.

132. ICANN DOCUMENTARY INFORMATION DISCLOSURE POLICY, *supra* note 109; Access to Information Act, R.S.C. 1985, c. A-1, § 6.1 (Can.).

Probably the most conceptually difficult task related to developing an information requesting mechanism for platforms revolves around the exceptions to disclosure. Among governments, there is a fairly well-established body of international standards for defining these conditions.¹³³ Where non-governmental entities have adopted these systems, however, they often include exceptions which are materially different from those that the governments rely on, such as ICANN's specific exception for information related to "changes, modifications, or additions to the [Internet's] root zone."¹³⁴ In the case of the platforms, there may need to be special care related to the disclosure of information about moderation decision-making which would allow bad actors to "game" the system, particularly with regard to the design of automated moderation processes.¹³⁵ While this is a significant concern, it is worth noting that it roughly parallels a similar calculus which takes place in public sector contexts every day, with regard to law enforcement information, and the risk that disclosing operational information will make it easier for criminals to escape detection or prosecution.¹³⁶ In order to address this problem, right to information laws typically include some kind of exception for information whose disclosure would cause meaningful harm to the efficacy of law enforcement efforts.¹³⁷ However, while there may be information under request which needs to be withheld or redacted under this exception, the calculus, in a law enforcement context, has generally been that the broader public interest is better served by subjecting these agencies to requests, rather than carving them off from the scope of freedom of information or right to information legislation entirely.¹³⁸ While making this case in a private sector context is naturally more difficult, the same basic reasoning should hold true. This is a tension which will always exist in enforcement efforts, which are easier for authorities to operate if the system is flexible and vague enough to suit their whims, though with obvious tradeoffs in terms of perceived legitimacy and procedural fairness.¹³⁹ For exactly this reason, clarity, around both the letter of the rule and how it will be implemented,

133. KARANICOLAS & MENDEL, *supra* note 94, at 6–8.

134. ICANN DOCUMENTARY INFORMATION DISCLOSURE POLICY, *supra* note 109.

135. The veracity or legitimacy of this claim is discussed in FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* 3 (2015); see also Nicholas Diakopoulos, *Accountability in Algorithmic Decision Making*, 59 COMM. OF THE ACM 56, 62 (2016).

136. A. Jay Wagner, *A Secret Police: The Lasting Impact of the 1986 FOIA Amendments*, 23 COMM. L. & POL'Y 387, 388 (2018).

137. See, e.g., 5 U.S.C. § 552(b)(7) (2000).

138. Louise Cooke & Paul Sturges, *Police and media relations in an era of freedom of information*, 19 POLICING & SOC'Y 410 (2009); Wagner, *supra* note 136; David Cuillier, *The People's Right to Know: Comparing Harold L. Cross' PreFOIA World to Post-FOIA Today*, 21 COMM. L. & POL'Y 433, 461 (2016).

139. *Musser v. Utah*, 333 U.S. 95, 97 (1948).

is a cardinal rule of international human rights law with regard to freedom of expression.¹⁴⁰

A similar challenge exists around protecting user privacy. This is a tension which platforms are constantly navigating, particularly since the Cambridge Analytica firestorm, which gave rise to parallel demands for better safeguards for personal information and for more transparency.¹⁴¹ Daphne Keller has also convincingly noted that there can be a particular tension here, insofar as pressure for platforms to provide more details into how they moderate content can lead them to track more, creating additional privacy risks.¹⁴² However, while there are unique aspects to the platforms' context, the broader challenge of protecting privacy in the context of an information structure which is "open by default" is nothing new. Public sector entities are often responsible for collecting and processing vast amounts of sensitive personal data, and in most developed democracies, these entities must do so while also adhering to applicable freedom of information or right to information rules. Moreover, while there is a real tension at the core of this dynamic which must be carefully mitigated, platforms have a tendency to conflate real privacy concerns with their own business interests, with the former used as a shield to deflect against disclosures which are counter to a company's public relations or commercial interests.¹⁴³

One additional aspect of the exceptions to disclosure which deserves special consideration is around trade secrets. Again, this is not unique to the platforms, as government freedom of information systems have to evaluate the sensitivity of private sector commercial information on a daily basis.¹⁴⁴ However, an important element of this dynamic, as it exists among more progressive right to information legislation, is that while commercial entities may object to the disclosure of particular materials in government hands, the final decision of whether or not to release them rests with the public body (and, ultimately, with the court, or the relevant specialized oversight body, should it come to that).¹⁴⁵ The challenge that a platform may face in fairly applying a test regarding the

140. U.N. Hum. Rts. Comm., General Comment 34, at art. 19, Freedoms of Opinion and Expression, U.N. Doc. CCPR/C/GC/34 (2011), <https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>.

141. Harry Davies, *Ted Cruz Using Firm that Harvested Data on Millions of Unwitting Facebook Users*, GUARDIAN (Dec. 11, 2015), <https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>.

142. Daphne Keller, *Some Humility About Transparency*, CTR. FOR INTERNET & SOC'Y BLOG (Mar. 19, 2021, 3:09 AM), <https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency>.

143. Wojciech Wiewiorowski, European Data Prot. Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research*, at 1, 5, 6 (Jan. 6, 2020), https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf.

144. Deepa Varadarajan, *Business Secrecy Expansion and FOIA*, 68 UCLA L. REV. (forthcoming).

145. *Id.*

commercial sensitivity of its own materials reinforces the need to build a measure of independence into disclosure structures, as is addressed in the next section.

D. The Facebook Transparency Board

As any journalist or researcher who has spent time grappling with America's FOIA system can tell you, simply having a law on the books does not mean it will be implemented in the spirit of promoting strong transparency.¹⁴⁶ The intent of the law, and the way that it is implemented, can be miles apart.¹⁴⁷ Even carefully drafted transparency structures may provide sufficient wiggle room for recalcitrant officials to avoid disclosing material that they want to keep secret.¹⁴⁸ Good right to information laws are often designed in an almost adversarial manner, on the understanding that some offices will use any possible legal avenue to skirt their obligations.¹⁴⁹

Better practice in the public sector is to delegate disclosure decisions to a specialized administrative body (an information office), and to allow for appeals against adverse disclosure decisions to an independent oversight body, typically an information commission or commissioner, though other regulators are sometimes tasked with this duty.¹⁵⁰ At ICANN, a requester who is unsatisfied with the organization's response to an information request can pursue an independent third-party review of the decision through the Independent Review Process spelled out in that organization's bylaws.¹⁵¹ Likewise, the Inter-American Development Bank has constituted an Access to Information Policy External Review Panel to perform a similar function around information requests they receive.¹⁵²

In a public sector context, a number of key values inform how this body should be constituted, with the most important being the availability of adequate and independently allocated resources, protection against interference through security of tenure and an independent appointments process, and the power to

146. Margaret B. Kwoka, *First-Person FOIA*, 127 YALE L. J. 2204, 2247 (2018).

147. *Id.*

148. David Cuillier, *supra* note 138.

149. One particularly extreme example is Bangladesh's law, which allows non-compliant officials to be personally fined if they fail to comply with their obligations. Right to Information Act (2009), § 27 (Bang.).

150. Michael Karanicolas & Margaret B. Kwoka, *Overseeing Oversight*, 54 CONN. L. REV. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3927852.

151. *Bylaws for Internet Corporation for Assigned Names and Numbers*, ICANN § 4.3 (Nov. 28, 2019), <https://www.icann.org/resources/pages/governance/bylaws-en>.

152. AMERICAN DEV. BANK, ACCESS TO INFORMATION POLICY § 9 (Apr. 26, 2010), <https://idb.docs.iadb.org/wsdocs/getdocument.aspx?docnum=35167427>.

enforce their decisions.¹⁵³ In the context of the platforms, the Facebook Oversight Board already incorporates a number of these values.¹⁵⁴ Although there are legitimate areas of criticism with regard to the Board, particularly in terms of its limited powers and remit, the measures which Facebook took to guarantee its structural and financial independence are well in line with international better practice for constituting strong oversight structures, such as through endowing its funding through an irrevocable trust and granting it a reasonable level of independence over its appointments and dismissal process.¹⁵⁵

This is not to suggest that the Facebook Oversight Board be tasked with making transparency decisions as well. For one thing, the skill set required to assess transparency questions is distinct from the specific issues that the Oversight Board was constituted to focus on.¹⁵⁶ Likewise, there are already well-documented capacity concerns with the Oversight Board.¹⁵⁷ The last thing needed is to add to their workload further, by pushing them into an entirely new thematic space. However, the Facebook Oversight Board presents a promising model for a potential “Facebook Transparency Board” (or, for that matter, a Twitter Transparency Board or a TikTok Transparency Board), which could hear appeals against refusals to disclose information and otherwise resolve disputes around transparency. Such a Board could also play a broader role as a champion of transparency in the organization through, for example, recommending changes to how data is collected or administered. This is analogous to the role of a good information commission in the public sector.¹⁵⁸ In constituting such a body, it could be useful to not only learn from the Oversight Board’s strengths, but also its weaknesses. In particular, broad access

153. Centre for Law & Democracy & Int’l Media Support [IMS], *Freedom of Expression Briefing Note Series*, at 39 (July 2014), <https://www.mediasupport.org/wp-content/uploads/2015/02/foe-briefingnotes-ims-cld.pdf>.

154. OVERSIGHT BOARD TRUST, FACEBOOK OVERSIGHT BD. § 1.4, (Oct. 16, 2019), <https://about.fb.com/wp-content/uploads/2019/12/Trust-Agreement.pdf>; OVERSIGHT BOARD BYLAWS, FACEBOOK OVERSIGHT BD. § 2.1 (Jan. 2021), <https://oversightboard.com/sr/governance/bylaws/>.

155. OVERSIGHT BOARD TRUST, *supra* note 154, at § 1.4; OVERSIGHT BOARD BYLAWS, *supra* note 154, at § 2.1.

156. *Frequently Asked Questions*, FACEBOOK OVERSIGHT BD. (Mar. 2020), <https://www.oversightboard.com/faq/>.

157. Evelyn Douek, “*What Kind of Oversight Board Have You Given Us?*”, UNIV. OF CHI. L. REV. ONLINE (May 11, 2020), <https://lawreviewblog.uchicago.edu/2020/05/11/fb-oversight-board-edouek/>.

158. *See, e.g.*, FED. INST. FOR ACCESS TO INFO. & DATA PROT., INTRODUCTION TO FEDERAL INSTITUTE FOR ACCESS TO INFORMATION AND DATA PROTECTION, <http://inicio.inai.org.mx/English/ABC%20IFAI%20English.pdf>.

to internal data, and the ability to make binding decisions regarding disclosure, would be important to insuring the body's efficacy.¹⁵⁹

The Facebook Oversight Board, along with other quasi-independent structures such as the Global Internet Forum to Counter Terrorism, could also serve as useful institutions to pilot how a requesting mechanism might be applied in the context of online platforms. The changes described in this section include significant conceptual and practical challenges which would mean, at the very least, that an "open by default" model would need to be developed over time, and implemented gradually. In the case of these standalone structures, however, there is less of a challenge with regard to sensitive personal and commercial information, since their mission is narrower and more specialized. Likewise, their status as multi-stakeholder entities (to a certain degree), and as a primary public interface to moderation efforts, should place a greater impetus for them to be on the front lines of transparency efforts. Developing an "open by default" model in these structures, complete with a requesting mechanism and an oversight or appeals structure, would be a welcome first step in assessing the viability of applying such a model to the platforms as a whole.

CONCLUSION

A persistent critique of platforms' content moderation policies is that they are too reactive, forever working to douse the fires caused by the platforms' policies, rather than looking ahead to mitigate challenges before they arise.¹⁶⁰ This dynamic means that platforms effectively outsource their institutional risk management to the journalists, civil society organizations, and members of the public who monitor their operations. However, while these independent accountability structures may be able to see potential consequences that are as yet unknown to the platforms, they are nonetheless limited by a lack of access to information from the platforms' side.¹⁶¹ This paper provides a formula for enhancing public oversight, both for the sake of basic democratic principles, and in order to equip these watchdogs with the tools to improve their engagement and oversight work.

It is easy for observers, and even the platforms themselves, to express their belief in the importance of transparency. Putting meaningful transparency structures in place is a much heavier lift. However, it is a necessary challenge to grapple with, given the platforms' entrenched position at the center of our political discourse. Rather than simply pushing for one or other piece of the puzzle, which might suit a particular research paper or project, stakeholders

159. See *Case Decision 2021-001-FB-FBR*, FACEBOOK OVERSIGHT BD. (May 5, 2021), <https://www.oversightboard.com/decision/FB-691QAMHJ>; Douek, *supra* note 157.

160. Maréchal, *supra* note 80, at 29.

161. Nicolas Suzor et al., *Evaluating the Legitimacy of Platform Governance: A Review of Research and a Shared Research Agenda*, 80 INT'L COMM'N GAZETTE 385, 395-96 (2018).

should consolidate around a structural change in how data is managed and delivered, which treats transparency and public accountability as the core value underlying the operation, rather than as an afterthought.

