

Staff Memo

Content Moderation on Internet Platforms



To: Jeff Hales, Chair of the Standards Board

Project Lead: Greg Waters, Associate Director of Research

Date: April 14, 2022

Purpose of Memo

Staff is writing this memo to recommend to the Chair of the SASB Standards Board that the Content Moderation on Internet Platforms research project be closed. This recommendation is based on staff's view that the objectives of this research project have been achieved.

Project Objective & Current Status

This research project was approved by the Chair of the Board at the December 2019 Board meeting with the objectives of evaluating companies' management of user generated content across the Technology & Communications (TC) sector and determining whether standard setting would be appropriate in any industries.

Two prior outcomes of this research project are (1) the Board's approval of a standard-setting project, Content Governance in the Internet Media & Services Industry, in September of 2020, and (2) the publication of the [Content Moderation Taxonomy](#) document in November 2020.

This research project has remained open as staff continued to evaluate two aspects of content moderation uncovered during staff's research. Specifically, staff was interested in further evaluating (1) the mental health implications for workers who perform content moderation for social media platforms, and (2) whether themes such as harmful content and freedom of expression present financially material risks and opportunities for companies in industries in the TC sector other than Internet Media & Services (IM).

Overall, staff has reached two conclusions:

- At this time, staff does not believe that standard-setting should be pursued in any industries outside of Internet Media & Services, although staff will continue to monitor how the emerging field of "Trust & Safety" evolves across the sector in the future.
- The mental health of content moderators remains an area of interest for some investors and may be financially impactful for some internet platforms with a large number of users, but continued research would be better addressed through future workstreams of the Human Capital research project.

Research Summary

Since the launch of the Content Governance standard-setting project in September 2020, staff has dedicated nearly all its available resources in the TC sector to that project. As a result, the [Content Moderation Taxonomy](#) remains the principal evidence base for the vast majority of research performed during the Content Moderation research project. Staff encourages interested Board members to review the taxonomy should they wish to review the findings, which focused on social impacts and related business activities.

The below sections summarize incremental findings applicable to this research project during the past year that fell outside the scope of the Content Governance standard-setting project.

Mental health impacts on content moderators

Concerns regarding the mental health of content moderators were intentionally kept out of scope of the Content Governance standard-setting project in September of 2020. This was driven by (1) a relative lack of interest expressed by investors during consultations, the majority of whom were not aware of this as a potential sustainability issue, (2) a lack of clarity around the magnitude of financial impacts associated with this sustainability issue, and (3) a desire to keep the scope of the standard-setting project focused given the complexity and breadth of issues already under consideration.

During consultations held in connection with the Content Governance standard-setting project in the first part of 2021, three investors, without specific prompts from staff, noted their interest in information regarding the mental health of content moderators. This was a noticeable difference from consultations during the first half of 2020, when no investors proactively brought up this topic. As such, staff believes there may be an increasing level of attention paid to this element of content moderation in the investor community.

In its monitoring of the tech sector since the publication of the Content Moderation Taxonomy, staff has noted a steady, if not increasing, flow of news reports regarding the mental health of content moderators. These include:

- November 2020: an open letter written by content moderators at Facebook asking the company to do more to bolster moderator mental wellbeing given the ongoing pandemic, noting that moderators were forced to work in-office while most other full-time employees were not¹
- June 2021: current and former employees of Citizen, a crime reporting app, discussed challenges related to performing the work, including mental health challenges and limited resources²

¹ <https://www.engadget.com/facebook-content-moderators-coronavirus-protections-202420431.html>

² <https://www.theguardian.com/technology/2021/jun/02/citizen-app-employees-mental-health>

- July 2021: previously-announced settlement between Facebook and former moderators approved by a judge; settlement included a \$52 million fund for ongoing mental health treatment and other payments³
- December 2021: a former TikTok moderator filed suit against the company for failing to protect her mental health⁴
- August 2021: a *New York Times* feature story on Facebook's \$500m annual contract with Accenture, with significant coverage dedicated to the mental health implications for Accenture employees performing the work⁵

While not exhaustive, this is a sample of the types of stories that appear in the media fairly frequently. Staff does not yet have a strong view regarding the financial materiality of this issue in the IM industry, but notes that the biggest risks appear to be highly concentrated in the largest social media platforms, which employ thousands of moderators reviewing content posted by billions of users. As such, the question regarding the prevalence of this issue across the IM industry is particularly pronounced (even in the context of content moderation, which as an issue itself has come under scrutiny from the Board and staff members on this basis).

Other Industries in the Technology & Communications Sector

Staff has not dedicated significant resources towards other industries in the TC sector following the publication of the Content Moderation Taxonomy. In its routine sector coverage, staff has not encountered significant evidence of investor interest or financial impact regarding the topic of content moderation in any industries outside of the IM industry.

One notable event related to this topic was the mass “de-platforming” of former President Donald Trump in the wake of the riots at the US Capitol on January 6, 2021. In addition to social media platforms, several other types of platforms in the technology ecosystem decided to cut ties with Mr. Trump, including e-commerce platform Shopify, which had been providing services to his campaign's website.⁶ Relatedly, a startup social media platform named Parler found itself without key vendors after its perceived lax moderation policy was blamed for facilitating the January 6th riots: the app was pulled by the app stores of both Apple and Google (making it impossible for new users to join its service); Amazon's AWS hosting service also cut ties, citing a breach of terms of service (pulling the platform offline for months).⁷

The decisions made by non-social media platforms in the wake of the January 6th riots were consistent with the previously-established evidence cited in the Content Moderation Taxonomy: while generally reticent to police content or get involved in policy matters, there are occasionally one-off controversies or societal outcries that lead to a broader array of tech companies making business decisions akin to content moderation.

³ <https://www.reuters.com/legal/transactional/judge-oks-85-mln-settlement-facebook-moderators-ptsd-claims-2021-07-23/>

⁴ <https://www.bbc.com/news/technology-59819664>

⁵ <https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html>

⁶ <https://techcrunch.com/2021/01/07/shopify-pulls-donald-trump-stores-off-its-website>

⁷ <https://www.wsj.com/articles/twitter-facebook-and-others-silenced-trump-now-they-learn-whats-next-11610320064>

As seen in previous examples, the risks related to these one-off actions run both ways: companies such as AWS and Apple were both lauded and criticized for ceasing to do business with Parler. However, staff was unable to find any evidence that there were significant financial impacts on these companies associated with their decisions. Rather, Parler's predicament was a novel piece of evidence showing that social media platforms risk losing their ability to secure the services of key vendors and suppliers should they fail to establish and enforce effective content policies.

Staff Rationale for Closing the Research Project

Overall, staff believes that the objectives of this Research Project have largely been met, and that the remaining open research item would best be handled through a different workstream. These are based on two key conclusions:

1. Staff does not believe that standard-setting regarding content moderation themes should be pursued in any industries outside of the Internet Media & Services industry. While there are occasional controversies related to content policy that pose challenges for a broader universe of players in the infrastructure of the internet, staff does not believe there is significant evidence of financial impact or investor interest in this topic in other TC industries.
2. The mental health of content moderators should remain under consideration as a topic that could be included in the SASB Standards in the future; however, the work would best be carried forward through its integration into the Human Capital Research Project, which has a specific workstream dedicated to evaluating the topic of worker mental health.

As such, staff recommends that this research project be closed.