CENTER FOR DEMOCRACY & TECHNOLOGY

User Notifications

From CDT's *Making Transparency Meaningful: A Framework for Policymakers*

Technology companies may notify users about a variety of activities that affect their speech, access to information, and privacy. Three types of user notifications that most strongly impact – and can help protect – user privacy and speech are: (1) *government demands for user data*; (2) *legal demands for content removals or restrictions*; and (3) *content moderation decisions by companies*. Notice about government demands for user data gives the user the opportunity to challenge the release of their data to the government and helps shed light on the often opaque processes of government surveillance. Similarly, notice about legal demands for content removals or restrictions gives users the opportunity to challenge those demands and reveals how governments and civil litigants obtain content takedowns or other restrictions. Notice about content moderation can educate users about intermediaries' content policies and reveal how and why intermediaries moderate content. All of these forms of user notifications inform public opinion and policymaking, helping hold governments accountable for their online surveillance and censorship activity and intermediaries for their content moderation practices.

////

Current Approaches to User Notifications

Government demands for user data

Governments around the world may demand data about users from technology companies, including users' content and non-content data such as traffic data as well as subscriber and billing information. Many tech companies have a policy of informing users of government demands for their data before turning it over unless they are prohibited from doing so by law or by other limited exceptions to their policies, such as emergency circumstances that threaten serious injury or death.[1]

In the United States, certain laws or judicial orders can prohibit a company from notifying users about a government demand for their data or require that they delay providing such notice. For example, the federal wiretap statute, Title III, prohibits a

1 *See* Nate Cardozo et al., *Who Has Your Back? Government Data Requests 2017*, Elec. Frontier Found. (July 10, 2017) (evaluating twenty-six major technology companies on their policy and advocacy positions concerning "handing data to the government," including user notifications).

provider of wire or electronic communication service from disclosing the existence of a wiretap (an ongoing form of surveillance).[2] The Stored Communications Act (SCA) permits the government to obtain some forms of electronic communications data without itself providing notice to the targeted user if the government obtains a warrant, or with delayed notice if it obtains a subpoena or court order under 18 U.S.C. § 2703(d) and meets certain statutory criteria. The SCA further authorizes issuance of a gag order precluding the company that receives the warrant, subpoena, or order from providing notice to the targeted user in certain circumstances.[3] The SCA also authorizes the FBI to issue a gag with a National Security Letter (NSL), a type of administrative subpoena, precluding the recipient from disclosing the existence of the NSL, if the FBI certifies that certain statutory criteria are met.[4] Providers are not permitted to disclose the fact that they have received orders to produce data pursuant to the Foreign Intelligence Surveillance Act (FISA), and Section 604(a) of FISA[5] permits providers to report only statistical information on the number of demands they receive under particular authorities. While these legal provisions can prevent or delay a company from notifying users of government demands for their data, some tech companies have a policy of providing notice after a legal prohibition on notice is lifted or expires.[6]

Legal demands for content removals or restriction

Governments may also demand that companies that host user generated content remove or otherwise restrict content (such as by geoblocking it) because it is allegedly illegal. In addition, private parties may demand that hosts remove or restrict content based on claims that it violates civil law, such as for defamation. Both governments' and private parties' legal demands for content removals or restrictions are often made by serving a court order or other legal authority on the host. A few hosts have a policy of informing users of legal demands for removal or restriction of their content unless they are prohibited from doing so by law, certain narrow emergency circumstances apply, or notice would be futile or ineffective.[7]

---

2 18 U.S.C. § 2511(2)(a)(ii).

3 18 U.S.C. § 2703(b)(1); *id*. § 2705. The SCA also permits the government to obtain non-content records without notice and to obtain a gag order precluding the provider of electronic communication service or remote computing service from notifying the affected user. *See* 18 U.S.C. § 2703(c); *id.* § 2705(b). Department of Justice guidelines limit the circumstances under which it will seek a gag order pursuant to § 2705(b) and limit gag orders' duration to one year other than in exceptional circumstances. *See* Memorandum from Rod J. Rosenstein, Deputy Attorney General, to Heads of Dep't Law Enforcement Components, Dep't Litigating Components, Director, Exec. Office for U.S. Attorneys, All United States Attorneys (Oct. 19, 2017). However, the Department's policy is intended "only to improve the internal management of the Department of Justice," and the Department expressly contemplates that orders of a longer duration may be necessary. *Id*. at 1 n.1 & 2 n.3.

4 18 U.S.C. § 2709(c); *see also* 18 U.S.C. § 3511.

5 50 U.S.C. § 1874.

6 *See* Cardozo et al., *supra* n.1.

7 *See* Andrew Crocker et al., Who Has Your Back? Censorship Edition 2019, Elec. Frontier Found. (June 12, 2019) (evaluating sixteen major technology companies on their content moderation policies, including user notifications regarding content takedowns and account suspensions in response to legal demands).

Some governments may also seek the removal or restriction of content that is not illegal but allegedly violates a host's content policies. In such cases, through Internet Referral Units or other government entities, the government notifies a host that particular content violates the host's content policy, and the host may remove or restrict it pursuant to its content policy.[8] These governmental efforts to leverage hosts' content policies to obtain removal of speech that is not illegal have been criticized both for allowing extra-legal government censorship and for their lack of transparency, since hosts and governments rarely notify users when their content has been removed pursuant to the host's content policy as a result of a governmental notification.[9] If users do not receive notifications about these government referrals, they may be unable to challenge their legality and may not even know that they are under government scrutiny.

Content moderation

User notifications concerning content moderation decisions can be divided into three categories or phases of notice: (1) Terms of service and content policies; (2) Notifications of enforcement actions; and (3) Appeals.

Intermediaries that host user-generated content usually notify users about what content is and is not allowed on their services. Intermediaries' terms of service may state what content is allowed or forbidden at a high level of generality,[10] and they often have additional, more detailed content policies, which are sometimes called "community standards."[11] The earliest content policies were relatively simplistic and lacking in detail. However, some – though not all[12] – now consist of a complicated and lengthy system of rules, with exceptions and caveats.[13] Content policies educate users about what they can say and how they should behave on a service, and while some users will intentionally break the rules, others will make a genuine attempt to understand and stay

8 Jason Pielemeier & Chris Sheehy, *Understanding The Human Rights Risks Associated With Internet Referral Units*, VOX-Pol (Mar. 26, 2020).

9 *See, e.g.*, Tomer Shadmy & Yuval Shany, *Protection Gaps in Public Law Governing Cyberspace: Israel's High Court's Decision on Government-Initiated Takedown Requests*, Lawfare (Apr. 23, 2021) (describing the "invisible handshake" between the Israeli IRU and hosts, through which "[a]ffected individuals are aware that content they posted was removed by an online platform because of incompatibility with the applicable community standards or terms of use; they are not aware of the fact that the platform acted in response to a government takedown request").

10 *See, e.g.*, *Terms of Service*, Facebook at Section 3 (last visited Nov. 29, 2021); *Twitter Terms of Service*, Twitter at Section 3, Twitter (last visited Nov. 29, 2021).

11 *See, e.g.*, *Facebook Community Standards*, Facebook (last visited Nov. 29, 2021); *The Twitter Rules*, Twitter (last visited Nov. 29, 2021).

12 Some content policies provide minimal information. For example, social cataloguing website Goodreads' Community Guidelines consist of eight bullet points with some introductory text and two disclaimers. *Community Guidelines*, Goodreads (last visited Nov. 29, 2021). Its Community Guidelines do not define terms used in it to describe prohibited content, such as "hate speech," "nudity" or "graphic violence."

13 For example, Facebook's content policy prohibiting nudity specifies that it allows images of female breasts if they are "depicting acts of protest, women actively engaged in breast-feeding and photos of post-mastectomy scarring." *Adult Nudity and Sexual Activity*, Facebook (last visited Nov. 29, 2021).

within them. Content policies are generally public and available to anyone, even if they do not have an account on the service.

An intermediary may also provide a user with notice when it takes an enforcement action against the user's content or account. Notice may be detailed – including information identifying the content removed, the specific part of the content policy that was violated, how the content was detected and removed, and an explanation of how the user can appeal the decision[14] – or it may be perfunctory. Some intermediaries warn users before taking certain enforcement actions,[15] while others provide notice only after the fact. In addition, whether an intermediary provides a user with notice may depend on the type of enforcement action taken.[16] For example, an intermediary that enforces its content policy using purposefully opaque content moderation practices, such as keeping an account active but allowing only the account holder to view the content they post,[17] may intentionally not notify a user of the enforcement action it takes.

Finally, some intermediaries give users the ability to appeal enforcement decisions, providing a further opportunity to communicate with users about content moderation practices and decisions. The appeals process may allow a user to present new information to the intermediary and ideally results in the intermediary notifying the user of the results of its review with information that is sufficient to allow the user to understand the decision.[18]

////

14 *See* *The Santa Clara Principles on Transparency and Accountability in Content Moderation* (last visited Nov. 21, 2021) [hereinafter "Santa Clara Principles"]. A 2019 report by the Open Technology Institute found that YouTube, Facebook, and Twitter met some though not all of the "notice" recommendations in the Santa Clara Principles. Spandana Singh, *Assessing YouTube, Facebook and Twitter's Content Takedown Policies*, New America (May 7, 2019).

15 For example, Instagram sends a warning to an account at risk of deletion for repeated violations of its Community Standards Enforcement that includes a timeline documenting the account's previous violations. *Account Disable Policy Changes on Instagram*, Instagram (July 18, 2019).

16 Content moderation is not just a binary decision to either take down content or accounts or allow them to remain on a service; depending on how they have designed their service, intermediaries can take a wide variety of actions against violative content, some of which may not be immediately obvious to the user who posted the content. For example, intermediaries may decrease the availability of a post by removing or downgrading its visibility in search results. They may stop recommending certain content or display it less prominently in users' feeds. They may also restrict forwarding or sharing of content. *See* Eric Goldman, *Content Moderation Remedies*, Mich. Tech. L. Rev. (Forthcoming 2021).

17 These opaque content moderation practices are often referred to as "shadowbanning." Gabriel Nicholas, *Spotlight on Shadowbanning*, Ctr. for Democracy & Tech. (Oct. 4, 2021).

18 *See Santa Clara Principles, supra* n.14; A 2019 report by the Open Technology Institute found that YouTube, Facebook, and Twitter met many of the "appeals" recommendations in the Santa Clara Principles. *See* Singh, *supra* n.14.

Improving User Notifications: Considering Tradeoffs

What are the costs and benefits of giving technology companies greater legal authority to disclose government demands for user data?

As explained above, in some cases, tech companies are precluded by law from notifying users about government demands for their data, or they must delay in providing such notice. Laws permitting these gag orders help protect against the risk of undermining an investigation by notifying the target. At the same time, gag orders increase the likelihood that illegitimate and unconstitutional surveillance will go unnoticed and unchallenged, since a target of an unlawful government surveillance order cannot challenge it unless they know it exists. Because broad authority to gag companies from notifying users of government demands for user data creates the potential for abuse, policymakers should consider whether existing legal authority permitting these gag orders is appropriately narrow. In particular, policymakers should consider whether the legal basis on which a gag order may be sought should be further limited, the duration of a gag order further restricted, or the ability to seek a gag order at all removed in certain circumstances. Policymakers should also consider whether companies should be permitted to make certain or additional aggregate information about government demands for user data publicly available, even if individual orders must be kept secret.[19]

Are gag orders on technology companies that receive government demands for user data constitutional?

Some tech companies have challenged the constitutionality of the gag order provisions for SCA orders and NSLs under the First and Fourth Amendments. Both the Third and Ninth Circuits have applied strict scrutiny to gag orders precluding providers from engaging in speech regarding requests for their customer's data and upheld the constitutionality of Section 2705(b) gag orders and NSL gag orders, respectively.[20] However, the Supreme Court has not addressed the constitutionality of these gag orders, and some advocates and commentators argue that they are prior restraints subject to an even higher level of scrutiny or that they do not satisfy strict scrutiny.[21] In addition, although Congress enacted some limits on the duration of NSL gag orders as

19 For example, policymakers should consider amending Section 604(a) of FISA to allow providers to report more granular statistical information about the number of demands they receive.

20 *Matter of Subpoena 2018R00776*, 947 F.3d 148, 155 (3d Cir. 2020); *In re National Sec. Letter*, 863 F.3d 1110, 1123 (9th Cir. 2017). The Second Circuit dismissed as moot constitutional challenges by Microsoft and Google to Section 2705(b) gag orders after disclosure was made to the affected customers. *Microsoft v. United States*, No. 20-1653 (2d Cir. May 14, 2021); *Google v. United States*, No. 19-1891 (2d Cir. May 14, 2021).

21 *See, e.g.*, Al-Amyn Sumar, *Prior Restraints and Digital Surveillance: The Constitutionality of Gag Orders Issued Under the Stored Communications Act*, 20 Yale J.L. & Tech. 74 (2018); *Br. for Amici Curiae the Chamber of Commerce of the United States of America et al. in Support of Appellant*, *Microsoft Corp. v. United States*, No. 20-1653(L) (2d Cir. Dec. 21, 2020), ECF No. 125. Whether a particular gag order survives strict scrutiny may depend on the statutory authority under which it is authorized; for example, it may be easier for the government to meet strict scrutiny for nondisclosure under FISA than other laws.

part of the USA FREEDOM Act in 2015, these limits are insufficient and do not cover all types of gag orders. In considering amendments to gag order provisions or new gag order provisions, policymakers should require gag orders to meet at least a strict scrutiny standard, *i.e.*, the gag order must be justified by facts showing that the order is narrowly tailored to promote a compelling state interest, and that there is no less restrictive alternative that furthers those aims. Moreover, to avoid Fourth Amendment concerns, authorization for gag orders should provide binding limits on their duration.[22]

When should companies notify users about legal demands for content removals or restrictions, and what information should be included in these notifications?

Notice from hosts of user-generated content that inform users when their content is removed or restricted based on a legal demand such as a court order gives users the information they need to legally challenge legal demands for content removals or restrictions or alert the public about the demands. In rare circumstances, it may be appropriate for hosts not to provide such notice: when they are prohibited from doing so by law, certain narrow emergency exceptions apply, or providing notice would be futile or ineffective.[23] When notice is provided, at minimum it should "identify the specific content that allegedly violates the law" and "inform the user that it was a legal takedown request."[24] Ideally, the notice should also include a copy of the legal order or other written demand, the identity of the government official, agency, or other entity who has made the legal demand and the legal basis for the demand. However, providing such detailed notice may be more expensive and time consuming for hosts, and may not be feasible for the smallest services.

There are additional considerations for hosts to weigh when government officials flag or refer content to the company, but the host removes the content under its own content policies. Clear notifications to users that the government was involved in flagging their content for review would allow users to bring legal challenges and draw public attention to this form of government action against their speech. However, such notifications may impose new costs on hosts, who may have to develop a process for tracking government referrals separately from other reports of violations of their content policies, so they can notify users of the government referrals. In addition, hosts may also object to providing user notices about government referrals because they fear it will give the false impression that a government referral *required* them to remove content pursuant to the host's content policy or improperly influenced their decision to remove

22 *See* Br. for Amici Curiae the Chamber of Commerce of the United States et al., *supra* n.21. (arguing that the SCA's allowance for indefinite gag orders itself may give rise to a Fourth Amendment violation, and citing *Wilson v. Arkansas*, 514 U.S. 927, 930 (1995); *United States v. Freitas*, 800 F.2d 1451, 1456 (9th Cir. 1986); *United States v. Villegas*, 899 F.2d 1324, 1336-37 (2d Cir. 1990)).

23 Crocker et al., *supra* n.7 (explaining that emergency circumstances "should not be broader than the emergency exceptions provided in the Electronic Communications Privacy Act, 18 U.S.C. § 2702 (b)(8)" and that "[a]n example of a futile scenario would be if a user's account has been compromised or their mobile device stolen, and informing the 'user' would concurrently – or only – inform the attacker").

24 *Id.*

content pursuant to their content policy. Such concerns can be mitigated by notices that clearly explain that a government official referred content for removal to the host under its content policy, and not under law, and that the host made the independent determination that the content at issue violated its content policy.

Notices about content removals and restrictions under a host's content policy but as a result of a government referral should include at least the same information as in user notifications about content moderation.[25] Ideally, the notice should also include a copy of the government referral and the identity of the government official or agency that made the referral.

What information should be included in user notifications about content moderation?

Notifications can enhance the legitimacy of content moderation by helping users understand why certain content is moderated. They can educate users about what content is allowed and forbidden on an intermediary's service, inculcating community values in users and helping users correct violative behavior. They can also shed light on content moderation decisions that are erroneous or with which users may disagree.

To meet these goals, user notifications must contain enough information, communicated in a clear and understandable way, to actually inform users. More information is not always better; providing user notifications can be time and resource-intensive, and intermediaries must make decisions about the level of detail to include and how to design them to make them most effective. The information available may also depend on the type of service an intermediary offers and the content moderation methods it uses. The Santa Clara Principles, a set of principles for transparency and accountability in content moderation, recommend information that intermediaries' content policies and user notifications about content moderation decisions should include.[26] (While these recommendations provide a useful overview for policymakers of key considerations in the area of user notice, they are not model legislation and should not be incorporated wholesale into proposals that would mandate user notifications.)

Intermediaries and policymakers should also consider whether, in some instances, user notifications about content moderation may be counterproductive. For example, informing spammers about how and why their content has been moderated may enable them to evade moderation in the future. Similarly, users who intentionally violate an intermediary's content policies may respond to a notice that their content has been moderated or account has been banned or suspended by creating a new account through which they can continue to break the rules. While secret content moderation decisions may help prevent evasion of content policies, they can also undermine

25 *See infra* infra User Notifications at 7 ("What information should be included in user notifications about content moderation?").

26 *Santa Clara Principles, supra* n.14.

legitimacy, user education, and the ability to hold intermediaries accountable for their content moderation decisions.

Can user notifications about content policies and content moderation decisions be mandated in the United States, consistent with the Constitution?

As with transparency reports, American lawmakers considering mandates that require intermediaries to publish content policies and notify users of content moderation decisions should consider whether doing so is consistent with the First Amendment. In general, strict scrutiny applies to statutes that compel speech by private speakers. In addition, content policies and information about content moderation decisions go to the heart of intermediaries' exercise of editorial decisions about what content to allow on their services and how to display it, which is protected by the First Amendment. While requiring publication of content policies and user notifications of content moderation decisions may not be direct regulation of the editorial decisions intermediaries make, lawmakers should consider whether these requirements would exercise indirect governmental influence or control over intermediaries' editorial discretion and thereby violate the First Amendment.[27]

27 *Herbert v. Lando,* 441 U.S. 153 (1979); *Miami Herald v. Tornillo,* 418 U.S. 241 (1974).

This brief is a part of the December 2021 CDT report, *Making Transparency Meaningful: A Framework for Policymakers*.

Additional CDT work on this topic: https://cdt.org/insights/report-making-transparency-meaningful-a-framework-for-policymakers

For more info, please contact **Emma Llansó**, Director of the CDT Free Expression Project or **Caitlin Vogus**, Deputy Director of the CDT Free Expression project.

ellanso@cdt.org

@ellanso

cvogus@cdt.org

@CaitlinVogus

The **Center for Democracy & Technology** (CDT) is a 25-year-old 501(c)3 nonpartisan nonprofit organization working to promote democratic values by shaping technology policy and architecture. The organisation is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

@CenDemTech

CENTER FOR DEMOCRACY & TECHNOLOGY