

GNI Submission on the Draft Commission Guidelines on Trusted Flaggers under the Digital Services Act

Introduction

The Global Network Initiative (GNI) welcomes the opportunity to comment on the European Commission's [targeted consultation](#) on its Draft [Guidelines on the trusted flagger mechanism](#) under Article 22 of the [Digital Services Act \(DSA\)](#). GNI has engaged extensively with the Commission and others around the DSA and its implementation through [prior submissions on the regulation](#), its [delegated acts](#), and [related guidelines](#), as well as through our own multistakeholder [convenings](#) focused on risk assessment and mitigation.

GNI has [consistently](#) raised concerns about the potential for government flagging of content under company terms of service, which combines the private discretion of company content moderation decisions with the coercive power of the State in ways that create serious risks to freedom of expression, while avoiding transparency and accountability. While trusted flaggers have played an important role in content moderation, priority treatment of notices is a form of privileged access to content moderation systems. Where that privilege is extended to government entities, as contemplated in the DSA and the Draft Guidelines, it creates serious risks for freedom of expression.

The Draft Guidelines appropriately acknowledge this risk, emphasizing the inappropriateness of giving trusted flagger status to national authorities who are empowered to issue orders to act against illegal content (para 31). Like the DSA, the Draft Guidelines, also emphasize the importance of transparency on the part of trusted flaggers (para 92), providers (para 12), and the Commission (para 62), as a mechanism to help avoid the potential for misuse of this status. In addition, the DSA and the Draft Guidelines recognise the potential for trusted flagger status to be misused and create mechanisms to address such misuse (para 81 and 93). Most importantly, the guidelines recognize and reinforce that trusted flagger notices should not necessarily lead to content moderation action (para 11) and that priority treatment must not pre-empt diligent, independent assessment by providers (para 83).

These are important provisions that should be retained. GNI has provided recommendations below that would support further transparency and accountability, in the interests of safeguarding freedom of expression. The Draft Guidelines are ultimately voluntary. The Commission, Digital Service Coordinators, and the European Board for Digital Services, with support from providers, will therefore need to monitor and encourage uptake through diligent, coordinated effort, and to act decisively where there is evidence of undue or overbroad impacts on freedom of expression and other rights.

Detailed Analysis

The Draft Guidelines contain several welcome transparency measures. These include the publicly available Trusted Flaggers Database (para 62), the recommendation that statements of reasons indicate when a restriction was taken pursuant to a trusted flagger notice (para 12), and the Annex II reporting template, which requires notice-level detail including the legal provision invoked, whether the provider disputed the notice, and the action taken.

These measures do not, however, distinguish between government and non-governmental flaggers. The Draft Guidelines should be clear that statements of reasons and annual reports should indicate when the notifying entity was a government or government-affiliated body. Providers should also be encouraged to publish the proportion of trusted flagger notices they act on and reject, disaggregated between government-affiliated and non-governmental flaggers. In addition, paragraph 96 and Part II of Annex II requires trusted flaggers to disclose procedures ensuring independence from providers only. They should equally require disclosure of the measures in place to ensure independence from law enforcement and other state functions.

This distinction is important because reporting under Article 22(3) is annual and retrospective. Where a government seeks removal of content through an order under Article 9 of the DSA, the request requires judicial oversight, rule of law safeguards, and per-request transparency reporting under established company practice. A government entity acting as a trusted flagger can pursue similar outcomes without any of these safeguards.

Additionally, GNI welcomes the position set forth in paragraph 31 that authorities empowered to issue orders against illegal content within the meaning of Article 9 should not commingle those powers with trusted flagger status. However, these eligibility criteria do not address the structural conflict of interest that arises when a law enforcement body flags content. A law enforcement unit may lack formal order-making powers while holding an active investigative or prosecutorial interest in the very content it notifies. Paragraph 42(a) requires DSCs to verify impartiality and freedom from political or other influence, but the guidelines do not elaborate on what this means in practice, and the framing does not capture this conflict.

Further, allowing the commingling of these powers as long as the entity in question is “empowered ... with sufficient management and operational autonomy ... so as to allow that department to carry out its activities in an objective manner,” (para 31) without providing any criteria or details as to what “sufficient management and operational autonomy” or what it means to carry out these roles in “an objective manner,” is plainly insufficient to address the concern that the Commission has rightly identified.

These concerns are exacerbated by the fact that trusted flagger status is awarded for an indefinite period. Without periodic review, there is no way to ensure that changes in government, funding, training, or other relevant factors are not undermining the “sufficient management” and/or “operational autonomy” criteria. Furthermore, dormant or repurposed entities could retain privileged access long after the basis for their designation has lapsed.

As such, the Draft Guidelines should be amended to make clear that only “national authorities not empowered for the adoption of orders to act against illegal content” should be considered for trusted flagger access. In addition, more details should be provided to guide DSCs in carefully evaluating and avoiding the possible commingling of law enforcement powers with trusted flagger status. Finally, a provision should be added to require government-affiliated trusted flaggers to reapply *de novo* for reauthorization every two years against minimum performance criteria, for example the proportion of an entity's notices meeting the precision and substantiation standards set out in paragraph 77.

Conclusion

In sum, GNI urges the Commission to amend the Draft Guidance as follows:

- **Require disclosure of government affiliation** of trusted flaggers in provider statements of reasons and annual reports, and encourage providers to publish action and rejection rates disaggregated between government-affiliated and non-governmental flaggers.
- **Hold government-affiliated trusted flaggers to transparency standards equivalent to formal government removal requests**, consistent with the safeguards attached to Article 9 orders.
- **Exclude law enforcement bodies from the mechanism** and provide additional details on how DSCs can scrutinize and ensure appropriate oversight and autonomy of government entities that are seeking trusted flagger status.
- **Replace indefinite status with periodic re-application** against minimum performance criteria.

GNI remains committed to constructive engagement with the Commission, DSCs, and other stakeholders as the guidelines are finalized.