

Response to the Targeted public consultation on the Guidelines on the trusted flagger mechanism under Article 22 of Regulation (EU) 2022/2065 (Digital Services Act)¹

The following statement summarizes the main areas in which it would be appropriate to supplement or clarify the European Commission's draft Guidelines on the Trusted Flagger mechanism under Article 22 of the DSA.

1. On the criteria of independence of Trusted Flaggers

There is room for further development with regard to the criteria for assessing the independence of trusted flaggers. The text of the guidelines indicates that particular concern is attached to links between a legal entity acting as a trusted flagger and providers of online platforms. Given the responsibilities associated with the role of these entities, it would be appropriate for the text to place greater emphasis on the fact that such entities should not only refrain from having connections with political parties, representatives of commercial entities that may have an interest in excluding competitors, or other actors in situations where conflicts of interest may arise, but should also provide further guidance on the methodology for verifying compliance with this independence requirement (point 42(a) and (b)). We therefore recommend introducing a broader range of conflict-of-interest indicators and requiring trusted flaggers, both as entities and through their individual members, to adopt and implement a conflict-of-interest policy.

2. On the granting of the status of “Trusted Flaggers” under Article 22(2) DSA

Although the guidelines state that the process of granting the status should be carried out electronically (point 52), the document does not further specify deadlines, the right to remedy deficiencies in the application, the obligation to provide reasons for decisions, or the right to review a decision. An alternative reference in this regard could be either a reference to time limits applicable under national administrative procedures, or at least the introduction of a general indicative timeframe in the guidelines. In addition, when regulating the process of granting status, we recommend establishing minimum procedural safeguards in line with the principle of good administration and Article 47 of the Charter of Fundamental Rights of the EU.

3. On the financing of Trusted Flaggers

This area can be seen as a kind of “grey zone”, as the guidelines indicate that, on the one hand, Article 22(2)(b) DSA provides that trusted flaggers must not be dependent on any online platform provider, while on the other hand they do not specify what level of financing can be considered acceptable (since dependence in this context is mainly focused on funding issues). If the guidelines allow financing of trusted flaggers by platforms, even though the DSA requires independence from any provider of online platforms, the question arises whether the Commission is effectively extending the content of Article 22(2)(b) DSA in a way that is not clearly supported by the text of the regulation. In this respect, we also propose clarifying acceptable forms of financing – e.g. public funding, project-based funding, revenue from NGO activities, etc. – in more concrete terms, taking into account proportionality.

¹ This work was carried out as a partial output of the projects APVV-24-0171 Digital balance – moderating illegal content and resolving disputes on digital platforms and VV-MVP-24-0038 Analysis of liability for Internet torts with machine learning methods.

4. On “significant number of inaccurate notices” and “accurate and inaccurate notices”

Although it is clear that the aim was to introduce a vaguer term that does not need to be specified in every case, the guidelines could at least generally clarify the acceptable level of error rate or the method for assessing a significant number of inaccurate notices, for example by introducing a percentage relative to the total number of notices. The text of the guidelines suggests that the assessment of the adequacy or inadequacy of a notice is carried out by the relevant platform, which evaluates individual notices, keeps statistics on them, and informs Digital Services Coordinators about their error rate. Here, a certain conflict of interest may be seen, as the online platform is the recipient of the notification (which may indicate that its internal systems for detecting illegal content have failed) and has its own economic interests; thus, the entity whose “decisions” are effectively monitored by the trusted flagger would simultaneously assess the quality of its own work. It should therefore be emphasized that the national Digital Services Coordinator is not bound by the platform’s assessment of the adequacy of a notice of illegal content.

5. On issues related to illegal content

Pursuant to Article 22(3) of the Digital Services Act, trusted flaggers shall publish, at least once a year easily comprehensible and detailed reports on notices submitted in accordance with the notice and action mechanism put in place under Article 16 of DSA during the relevant period. One of the categories of information to be provided in such a report is at least the number of notices categorised by „*the type of allegedly illegal content notified*”. According to point 95 of the Guidelines, trusted flaggers should report the illegal content notified using the categories of illegal content set out in the Commission Implementing Regulation 2024/2835. In this regard we would like to reiterate the importance of the provision of a detailed information on the nature of illegal content examined and reported by trusted flaggers using the mechanisms described in DSA. Specifically, the need to provide not only a simple reference to the applicable European Union or national law or the terms and conditions policies of individual providers, but also a detailed contextual information about their interpretation of illegal content under examination. Whereas in some instances the definition of a specific category of illegal content is clear (especially if the illegality of content is based on the European Union legislation, e. g. child sexual abuse material, terrorist content etc.), this may not always be the case in relation to other content categories, e. g. hate speech, content infringing privacy of an individual etc. In this regard, the publication of a detailed analysis provided by trusted flaggers with expertise in this area and, specifically, concrete anonymised examples of content evaluated as illegal, may prove significantly helpful in shaping the interpretation of certain categories of illegal content. Reflecting the experiences with first transparency reports published by providers, the lack of granularity in reporting the individual categories of illegal content and categorizing most reported violations under the category of “*Other*” should be avoided. Naturally, the fact that trusted flaggers focus on specific illegal content type will ensure certain degree of specificity. Nonetheless, it would be helpful if a more detailed reporting would be encouraged. In summary, we propose that the Guidelines should recommend that trusted flaggers provide more detailed contextual information regarding the illegal content under review, specify the factors taken into account in the assessment of content illegality, provide the main arguments that led to the interpretation of content as illegal, and include concrete anonymised examples of content classified as illegal.

6. On cross-border aspects of illegal content

The guidelines do not further regulate rules for determining applicable law, the territorial scope of illegality, and the cross-border effects of notifications. In this respect, it would be appropriate at least to introduce an obligation to specify the affected Member State and the legal basis of illegality, not just a type of illegal content (in general terms). Although it is undisputed that the legal systems of individual Member States are to some extent harmonized (and in some matters unified), this could help platforms decide whether to remove content only within the territory of the affected Member State or across the entire EU.

Summary

The most significant aspects of the draft guidelines that would require further clarification can be identified as the criteria of independence, the regulation of the process for granting trusted flagger status, issues relating to the permissibility and clarity of funding arrangements, the final assessment of the accuracy of submitted notices of illegal content, on issues related to illegal content in general and the cross-border aspects of illegal content. Although it is undisputed that the guidelines do not constitute a legally binding instrument and are purely recommendatory in nature, given the novelty of this mechanism, we consider it appropriate to further clarify certain selected aspects. Overall, however, the guidelines constitute a highly useful tool for stakeholders in implementing this mechanism in practice.

Department of Commercial and Business Law
Faculty of Law
University of Pavol Jozef Šafárik in Košice

Prepared by:

Simona Rudohradská
Laura Bachňáková Rózenfeldová
Regina Hučková
Diana Treščáková