CELE's Submission on the Draft Delegated Act on Transparency Reports (detailed rules and templates) under the DSA

June 24, 2024
Buenos Aires, Argentina

The Center for Studies on Freedom of Expression and Access to Information (CELE) is an academic research center affiliated with Universidad de Palermo in Argentina. The Center provides technical, legal analysis on issues affecting this fundamental right, and since 2012 has been studying freedom of expression on the Internet as a specific research area. The Center is a leading voice on the promotion and protection of freedom of expression nationally, regionally and internationally.

This submission was prepared in response to the consultation on the draft Delegated Act implementing the transparency reporting obligations of providers of intermediary services and of providers of online platforms under the Digital Services Act (hereinafter, DSA). It aims to contribute to the efforts of the European Commission in the implementation of these obligations, with an analysis that emphasizes freedom of expression. It will focus on three aspects of the templates and provide concrete suggestions on: the uniformity and consistency of reporting across different platforms; the challenges raised by the categories of content in the templates; and the balance between quantitative and qualitative information requested. A better understanding of the context in which decisions are made will be essential for adequate DSA implementation and for the development of the societal infrastructure needed for its success.

As stated by the European Court of Human Rights in Handyside v. United Kingdom, "freedom of expression constitutes one of the essential founda-

tions of a democratic society, one of the basic conditions for its progress and for the development of every person". In the digital age, a healthy democracy is not achievable where online freedom of expression is not guaranteed. Transparency and access to information, like the one fostered by this delegated act, is key to protect and promote freedom of expression. Mandated transparency needs to be necessary and proportionate and when mandating information or transparency, the EC should ensure that the information be requested objectively towards the legitimate ends and purposes described in the enabling law. Promoting biased, decontextualized or incomplete transparency reports could threaten the understanding of the impact of technology in society and result in unnecessary or disproportionate restrictions to freedom of expression for every platform user in ways incompatible with international standards of freedom of expression and access to information.

Uniformity and consistency

Key among the existing challenges with transparency reports that are currently published voluntarily by companies is that they are not machine readable, they are not comparable with one another and they don't necessarily cover the same timeframes. Reasons for this incompatibility are varied, including but not limited to the use of the same terms to refer to different things (i.e. how each company defines a single piece of content for quantity purposes or what they mean by hate speech), different terms of service, difference services offered, and different perceptions of their own obligations. The Delegated Act provides much needed guidance and common obligations for companies to abide by in the preparation of such reports -although as mentioned hereafter there is still much more room for improvement and challenges that so far are still not addressed. The Delegated Act should firstly aim to ensure that reports are machine-readable and structured in a way that makes reports issued by different companies comparable.

Furthermore, Annex II to the Delegated Act should include even more precise instructions concerning the methods that all platforms should follow for counting and reporting, and eventually narrow the scope of such reporting to what the EC deems relevant for DSA enforcement purposes. For

instance, there should be a uniform procedure, common to all platforms, to define "cases" for counting and reporting. As academics and civil society organizations have pointed out in the past, a "case" or a content can be defined differently by different companies (i.e. the same content reported across two platforms owned by the same company; or a single report including more than one individual piece of content). Similarly, cases will arise where some order or notice include different URLS where the exact same piece of content is replicated (i.e. an allegedly infringing image may be uploaded 100 times by 100 different accounts). Moreover, there should be a uniform way of counting and reporting cases where a single item is flagged multiple times (by different flaggers) as illegal or marked as incompatible with the Terms of Service. Some platforms could report such a notice or removal as a single one, while others could decide to report it as multiple.

Intertemporal consistency is also key: templates should be designed in such a way that enable comparison between reports issued in different years. These requirements are consistent with the DSA's goal of securing access to relevant data for independent researchers.

Problematic content categories for transparency purposes

Under Article 10 of the European Convention on Human Rights (hereinafter, ECHR), everyone has the right to freedom of expression, which encompasses the freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. These freedoms may only be legitimately restricted upon fulfilling the three prong test: legality, necessity and proportionality. Restrictions need to be prescribed by law and necessary in a democratic society. They must also be proportionate and pursuant to legitimate aims and purposes. Under Article 1 of the European Convention, states have the duty to respect and to guarantee the rights therein recognized to everyone under their jurisdiction. Article 19 of the International Covenant on Civil and Political Rights (hereinafter, ICCPR) provides similar protections.

International human rights law and European human rights law require that states differentiate between illegal and permissible content. Per the le-

gality requirement, restricted speech needs to be clearly and unambiguously identified in a law. So when states mandate that content be assessed, identified and acted upon, there are differences as to what a member state can expect companies to do vis a vis illegal speech than what they may expect and request companies to do regarding legally protected speech. Moreover, states are prohibited from restricting speech, directly or through third parties, in ways inconsistent with Article 10 of the ECHR and Article 19 of the ICCPR. Collapsing categories of illegal speech with categories of speech that are legally permissible but generally banned by online platforms may lead to confusion and potential misrepresentation, especially when the indicators are solely quantitative and aggregated rather than qualitative and contextualized. Furthermore, the conflation itself promotes ambiguity and generates incentives for over-removal of legal content.

The distinction between legal and illegal content also contributes to distinguishing who and how reports should be made to companies and how companies should handle such requests for action. State agents may only act upon illegal content and following due process. The current templates could benefit from further nuances regarding the origins of reports, particularly following this important distinction between legal and illegal speech. This distinction and transparency over this information would generate the right incentives for companies to comply with Human Rights and fight back against potential state abuses where they see them, in compliance with the UN Guiding Principles on Business and Human Rights.

In cases where the law indicates that a certain content is illegal in some countries and legal in others, it would be useful to know whether the content is removed altogether and where more proportionate measures are taken, such as restrictions based on the geolocalization of the users. Further nuances may contribute to better understanding and to the contextualization of the information presented in the reports.

As for tab 5, which entails content moderation in which companies engage at their "own initiative", such activity may be carried out both pursuant to legal norms and to platform's terms of service. These two should be disaggregated, and qualitative information could be required to complement these data points to generate incentives for companies to adopt

international freedom of expression standards in their content moderation practices.

Determining how companies manage content and risks associated with content is particularly important. Such management is guided by different company interests, including but not limited to the kind of service they want to offer, the user activity they are seeking, business potential, but also civil society pressure and state official and unofficial pressure. In order to foster better compliance and enjoyment of international human rights online, including freedom of expression and privacy, transparency should contribute to a better understanding of the interactions between the public and private forces at play. Furthermore, in a predominantly quantitative analysis of information, it is even more important to contextualize numbers to easily ascertain how much is company and market action and how much is State-led influence in content management. This will also be key to understanding whether the DSA is serving the purposes for which it was created and whether it would allow for correction and revision.

Additionally, in tab 3 ("member state orders"), when providing information under Article 15(1)(a), it would be important to further disaggregate the requests issued by judicial authorities and those originated elsewhere (such as administrative authorities) and to inform, in each of these categories, the percentage of compliance by companies, expanding the reporting possibilities from "complied with" or "non complied with" to other categories of partial or qualified compliance. In tab 4 ("notices"), further disaggregation should be made based on two criteria: whether the subject issuing the notice is a state actor or not, and whether it is a trusted flagger or not. Therefore, the template could use four categories: ordinary notices received under article 16 from non-state actors, ordinary notices received under article 16 from state actors, notices received from trusted flaggers who are state actors, and notices received by non-state trusted flaggers.

Disaggregation per member state will be essential throughout all the report for oversight and enforcement of the DSA at the national level. Illegal content and content incompatible with the terms of services of platforms should be disaggregated.

Qualitative and quantitative indicators

While acknowledging the value of quantitative indicators, it is important to strike a right balance between those and qualitative ones. The proposed templates are much more inclined to quantity than they are to quality. Aggregated and bulk figures often fail to deliver a clear understanding of how company action affects freedom of expression or whether they proportionally contribute to guaranteeing other rights.

Need for further nuance within quantitative data and indicators

Quantitative data should be gathered following qualitative standards and categories should be more nuanced to provide for more accurate representations of user conduct, company conduct and state conduct in platform content management. For instance, aggregated information on the amount of content reported on a platform can be helpful to assess the size of the content flow and size of the sample that is being analyzed, however it says little about what concrete content issues the platform is facing. Similarly, asking for overall content reports or takedowns per content category without any additional qualifier may serve to define the sample but fails to assess the specific company action vis a vis such content.

Providing further nuance to the categories may help identify and distinguish wrongful or accurate company practices and standards. In tabs 3 and 4, for instance, templates could require platforms to further qualify the data establishing the percentage of reported or moderated content that dealt with public figures, especially politicians and public officials. This entails content uploaded by them and content uploaded by third parties about them. Smaller platforms should at least be able to disaggregate the data linked to verified accounts related to public officials or state institutions when they have such mechanisms. The EC could also encourage the inclusion of an additional category of "content reported or taken down within special circumstances" like protest or elections. And they should be provided with an opportunity to explain.

Further nuance is also necessary to understand content upholding or in-

stances where the company decides against the reporting party. Knowledge of the reasons behind content moderation decisions is necessary to understand how platforms interpret and apply the law. While statements of reasons are published for all restrictions applied to illegal or incompatible content, no information is made public in connection with those situations where companies receive requests or notices about alleged illegalities and decide to leave the content online. The information about requests or notices that are rejected, dismissed by the company, or appealed by the user and reinstated by the company could be further classified based on the grounds of the decision or provide some space in the template for companies to explain and contextualize their aggregated numbers. Finally, the number of content moderation decisions in tabs 4 and 5 that have been challenged by end users is to be made available pursuant to article 24(1)(a) of the DSA. This information is duly required in tab 6 of the template. However, for a better understanding of how content moderation systems are working and how the law and the platform's own rules are being enforced, this information should be shown disaggregated by category of content also in this tab.

Even if differences are not stark, nuances will also arise in the scope of content prohibitions, legal definitions and court interpretations of categories of unprotected expression and conduct. The law is not uniform across all EU states and, as a result, different sets of conducts and expressions are illegal in different countries. Consequently, companies should inform whether these differences are taken into account when receiving requests for action from national authorities, and whether the personnel in charge of processing these requests are being instructed in domestic law to overcome this difficulty.

While a description of the automated systems of content moderation and the statement of their error rate is of great relevance, more information could be provided in connection with these systems, such as whether they are trained to apply the three-part test, to recognize and give special treatment to public interest information -from or about public officials - and others where different standards should be applied, to distinguish satire from truthful speech, true threats from hyperbole, or incitement from sit-

uations from which no imminent lawless action is likely to stem and their accuracy.

Finally, even though it is not mandated by law, in the cases of tabs 3 and 4, CELE would suggest that member states and the EC be encouraged to gather and publish their own transparency reports, reporting the number of orders and requests they have made under articles 9 and 10 of the DSA, so they can be contrasted with those included in the transparency reports by the companies, similarly to what trusted flaggers are required to do under article 22(3) of the DSA.

Case studies

Encouraging qualitative data and case-studies is of utmost importance to make sense of the figures provided in the quantitative analysis. A "case studies" approach could also be included, following the GNI's assessment toolkit. Companies could be asked to produce after-the-fact assessments of their processes, not in general and broad terms but as they worked in specific contexts, "as applied" to specific problems. Companies should be relatively free to choose the cases they decide to highlight, but these should be relevant to DSA implementation and address a wide sample of issues the platform routinely addresses in their moderation processes.

Transparency from the EC should complement any private transparency obligation

The DSA mandates that transparency requirements be proportionate to the societal impact that companies have and to their type and size. VLOPs and VLOSEs' due diligence obligations are subjected to the highest standards as a result of their societal impact. While the most-often cited criteria to assess impact have been market size and user base, these criteria are set to be informed and potentially changed precisely as information and company transparency practices start flowing.

Mandated transparency, especially State-mandated transparency, needs to comply with legality, necessity and proportionality standards. As trans-

parency obligations are imposed to platforms based on their social impact, proportionality should serve as a tool to aggravate certain obligations of VLOPs and VLOSEs where needed, rather than as an attenuating factor for smaller companies, thereafter focusing on the necessity aspects of the test.

One of the biggest challenges in reigning in Internet platforms is the lack of adequate understanding of their societal impact. Technology evolves rapidly and societal impacts are not yet fully known. The access to data for researchers that the DSA aims to secure is a fundamental step in that direction. It would feed the peer-reviewed, independently-led process of professional research that is necessary to fill the knowledge gaps produced by the pace of technological evolution, adoption, and change; the challenges of empirical research at a global scale; and the unequal distribution of resources available for resource-intensive research between the global North and the global South. Transparency efforts will surely contribute with information and data points to inform such understanding. European Commission officials have voiced the intention for a regulatory dialogue among the many actors involved and benefited by the DSA, including trusted flaggers, civil society, and researchers. How the assessment of company and technology impact varies over time and how the different elements of the DSA implementation contribute to such change should also be transparent.

Conclusions

We celebrate the initiative of the European Commission to open up this Draft Act for feedback and consider this a unique opportunity to provide more depth to the analysis of transparency obligations to better protect and promote freedom of expression. This is particularly relevant when dealing with transparency mandates over the implementation of rules governing expression.

While we believe transparency is of utmost relevance for successful enforcement and oversight of the DSA, we would emphasize that mandated transparency also needs to comply with legality, necessity and proportionality standards, as set out in international human rights law, and highlight the need for further nuance and distinctions in certain aspects of the pro-

posed templates that raise especially important challenges for freedom of expression. These principles should inform the Commission in making the templates and in enforcing the provisions of the DSA.

The transparency reporting process would benefit from adopting these recommendations:

1. Foster machine readable, comparable reports, contributing to provide uniformity and consistency where needed, particularly procedural, like the use of a unified methodology for counting and reporting are essential for any meaningful analysis of transparency reports.
2. Distinguish and separate the categories of illegal and legal but harmful content for reporting purposes.
3. Distinguish and separate state led action from company and/or user led action. To this end, the template could benefit from adopting further nuance in the reporting obligations pertaining to who reports content and how different reports are treated based on how they originated (private or state led).
4. Quantitative data should be further contextualized. Companies should be provided with opportunities, within the same templates, to qualify and explain data where needed.
5. Qualitative indicators should inform quantitative categories. Companies should be expected to report -when possible- whether the content reported or acted upon involved public officials or public interest. Whether there were special circumstances surrounding reports or takedowns (elections, social unrest, etc.)
6. A "cases studies" approach could be a significant addition to the qualitative data points already requested by the report, and could enable the public to gain first-hand understanding of how abstract criteria of content moderation operate when applied to concrete situations.
7. The information stemming from transparency reports can provide valuable insights for the assessment of the societal impact of platforms, which will result in the adoption of better, more tailored policies. Proactive transparency from the EC in how their understanding and interpretations of the DSA vary over time is key for the effective-

ness of the attempted regulatory dialogue.

Thank you,

Agustina Del Campo
Director
Center for Studies on Freedom of Expression
and Access to Information (CELE)