



DSA Shadow Report 2024

**Compiled by
Adinde Schoorl &
Tamas Berecz**
2024

TABLE OF CONTENTS

INTERNATIONAL NETWORK AGAINST CYBER HATE – INACH	2
INTRODUCTION	3
MONITORING	4
TRUSTED FLAGGER SYSTEM	6
DIGITAL SERVICES COORDINATORS	8
CONCLUSION	10
BIBLIOGRAPHY	12

International Network Against Cyber Hate – INACH

INACH was founded in 2002 to use intervention and other preventive strategies against cyber hate. The member organisations are united in a systematic fight against cyber hate, for example as complaints offices, monitoring offices or online help desks. In their respective countries, they provide important contacts for politicians, internet providers, educational institutions, and users.

Funding for INACH is provided by its members, the European Commission, the BPB and other donors. The International Network Against Cyber Hate (INACH) unites multiple organizations from the EU, Israel, Russia, Africa, South America, and the United States. While starting as a network of online complaints offices, INACH today pursues a multi-dimensional approach of educational and preventive strategies.

This publication has been produced with the financial support of the Citizenship, Equality, Rights and Values (CERV) Programme of the European Union. The contents of this publication are the sole responsibility of the International Network Against Cyber Hate and can in no way be taken to reflect the views of the European Commission.



Supported by the Citizenship, Equality,
Rights and Values (CERV) Programme of
the European Union

Introduction

The Digital Services Act has been in force for the Very Large Platforms (VLOPs) since August 2023 and for all platforms since 17 February 2024. VLOPs are the online platforms that have more than 45 million users in Europe. The European Commission is in charge of supervising the compliance of the platforms. They are responsible for monitoring and enforcement of the obligations that apply to these platforms. The goals of the DSA entail better protection of fundamental rights, less exposure to illegal content, greater democratic control, oversight of systemic platforms and mitigation of systemic risks, such as manipulation or disinformation (European Commission 'The Digital Services Act').

The aim of this report is to focus on the effects of the DSA on our work in preventing and countering online hate speech. Even though the DSA has only been in force since a year - at least for the VLOPs - this report is a chance for us to present a baseline of experiences so far that can be revisited annually. It is necessary to add however, that although the DSA has been in force, not all processes are in place yet to oversee it. For example, in many countries the Digital Services Coordinators (DSCs) have not been officially appointed and the requirements and processes to apply for Trusted Flagger are not public yet. Of course, those facts are also part of our experience with the DSA so far for Civil Society Organisations: the lack of clarity around the processes and formalities. This is in general a fact that shines through all the new laws and acts that are currently being prepared and introduced in the EU: the role and position of CSOs in all these new processes are very unclear. For example, the new CoC on illegal hate speech, the Trusted Flagger system under the DSA and other new acts like the AI act do not involve the CSOs in the development and roll out. We would welcome the efforts that involve CSOs more in these processes, since the organisations in the end have to work with these new bodies of law and their views are essential.

In this first annual report we would like to share our experience, insights and recommendations regarding three different topics within the DSA: monitoring, the system of Trusted Flaggers and the contact with the DSCs. We would like to end this report with recommendations on how to improve this process of the rolling out process of the DSA in all EU countries.

1. Monitoring

Many of the INACH members and the INACH secretariat participate every year in the Monitoring Exercise organized by the European Commission, to check if the social media platforms adhere to the rules within the Code of Conduct on Illegal Hate Speech Online that they signed. Next to that, within the SafeNet project, INACH and other partners monitor online hate speech on a daily basis. Due to these monitoring exercises and projects, we have extensive experience in monitoring online hate speech. One would expect that the DSA is helping to streamline, harmonize and intensify the moderation efforts of platforms, but so far that has not been the case.

The evaluation of hate speech comments on social media platforms is becoming more and more inconsistent. Reports on hate speech comments are being evaluated in prolonged periods. Also, the reporting user is not properly informed about the outcome of the report. Reporting with the DSA in force has also actually become more complicated on some platforms. For a while, all platforms had an extra reporting button: 'reporting under EU law.' Right now, only X still has that option. When one reports with this option the platform asks for legal knowledge that the average user does not possess and the ability to verbalize why content is illegal which is an unfair requirement for regular citizens. Also, the platform demands private data from users which might jeopardize their privacy and safety, and lowers the likelihood of users deciding to report content. By the way, it is unclear why other platforms do not have the option to report under EU law any more. In general, the reporting forms change constantly, which leads to confusion and demotivates reporting. Finally, communication by platforms is conducted in incorrect or bad translations into national languages. All together, reporting under the DSA as a regular user means more obstacles instead of less and demotivation instead of more incentives to report illegal content.

When it comes to removal, platforms vary in how they handle the reported content, but there has been a general decline in removal rates. There is an increase of content that is clearly illegal and is not being removed. Also, there is a new measurement in place that X uses, called geoblocking. This means that the reported content is only restricted in the country it has been reported from, but not in the rest of the EU. However, often when we checked, the reported content was still available in the restricted country or had actually been restricted in the incorrect country.

One thing that the DSA has led to, is more information. Under the DSA, platforms are required to publish their transparency reports twice a year. The first transparency reports were published in November 2023 and especially the numbers on moderation efforts per EU language by different platforms are very insightful. Most platforms are not investing much human moderation into languages other than English. For example, X does not have moderators in Estonian, Greek, Hungarian, Lithuanian, Maltese, Slovak, Slovenian, Czech, Danish, Finnish, Romanian, and Swedish. LinkedIn in fact only has moderators in seven out of all EU languages. One can wonder how they will moderate hate speech and disinformation during this European election year?

In general, all platforms have significantly more moderators for the English language than any of the other languages. For example, YouTube has 15,142 moderators in English while the second-largest group of moderators is 507 for Spanish. Moreover, it is not clear whether these numbers only apply to EU content or also globally. Because would that mean 20 moderators of X monitor all Spanish content globally or 30 moderators on LinkedIn for all French content globally? (Global Witness 2023). These numbers on moderators do explain part of the struggles that we and our network experience while we monitor and report illegal hate speech online.

At the same time, the DSA has also already led to real steps by the European Commission to hold platforms accountable for their policies regarding illegal online hate speech. The European Commission has formally opened an investigation into X to see whether they have breached the DSA rules 'in areas linked to risk management, content moderation, dark patterns, advertising transparency and data access for researchers'. It follows a request for information sent by the Commission to X in October after it received indications the platform was spreading illegal content and disinformation, particularly terrorist and violent content in the context of the Israel-Hamas war (Jones 2023). The European Commission also started an investigation of TikTok regarding illegal content, protection of minors, advertising transparency, data access for researchers, as well as the risk management of addictive design and harmful content (European Commission 2024).

2. Trusted Flagger system

INACH and most of the members within its network were Trusted Flaggers (TF) before the DSA was in force and therefore have extensive experience in what factors contribute to a well functioning Trusted Flagger system. Currently there is still much unclear on what the Trusted Flagger system under the DSA should look like. Article 22 of the DSA permits every EU country's government to appoint the Digital Services Coordinator (DSC), who will decide on what organizations will become Trusted Flaggers.

These are the requirements for becoming Trusted Flagger under the DSA. 'It' refers to the applying organization:

1) It has particular expertise and competence for the purposes of detecting, identifying and notifying illegal content; (b) It is independent from any provider of online platforms; (c) It carries out its activities for the purposes of submitting notices diligently, accurately and objectively ([20240216 Article22 GuidanceForm Branded vF KW.pdf \(cnam.ie\)](#)).

NGOs will have to apply for this position if they are of the opinion that they can carry the responsibility that the system requires. For example, there is only a limited amount of organizations that will be allowed into the system, there is a minimum amount of content that needs to be reported by the TF annually, the TF needs to produce annual reports on their monitoring and the TF needs to ensure its independence as an organization. In return, the DSC will have the power to revoke the position of TF if an organization does not adhere to all the requirements. Finally, it seems TF's will not receive any remuneration for their work.

So far, NGOs do not know the interpretation of the rules and the advantages or disadvantages to apply for Trusted Flagger. We don't even know whether organizations can maintain their original TF status if they do not apply to it within the DSA system. There is increasing concern that social media platforms and Member States will treat only official Trusted Flaggers as their partners and not sufficiently take into account the perspectives and knowledge of other key stakeholders, including previous Trusted Flaggers. While there is some anecdotal evidence that social media platforms will continue to engage with TF's appointed under the Code of Conduct without a formal role, a loss of knowledge and networks developed under the previous regime is a risk to the system. The relationship between TF's appointed under the DSA, TF's from the previous monitoring regime, social media platforms and the Commission should be further clarified (Facing Facts 2024). Second, the requirement of independence is

unclear and leaves room for national DSC's interpreting to their liking. Many NGOs worry there is the chance that national governments will abuse their position when appointing Trusted Flaggers. There are many organizations that work to protect minority rights in a hostile environment where even national governments are aggressively going against them. Organizations that are connected to the government or organizations that work for minority rights that fall within the government's liking will be more likely to be awarded the TF status over the aforementioned NGOs.

Another worry is whether the national DSC's will be taking inclusivity into account. For example, #StopFisha in France worries that no organization has been appointed that is involved in gender based hate. This means there is a higher chance that this type of hate speech will be overlooked. From our Croatian member, Centre for Peace Studies, it was reported that the government is silent on the process of appointing the DSC's and Trusted Flaggers. Their worry is that, once appointed, these TFs might be exposed to huge pressure by the public - even if their names remain anonymous, the name of the organization will be publicly known and they doubt if their organization would be okay under this amount of pressure.

There seems to be a consensus among DSCs that they do not expect a lot of organizations to apply for becoming Trusted Flagger due to the non existing remuneration for the work and the bureaucratic workload that comes with the position. Also, so far it seems that the original Trusted Flagger system will remain to exist even though that has not been confirmed or denied by platforms. However, currently those systems are still functioning while the DSA is already in force. A worry is that the DSCs will require Trusted Flaggers to prove their legal knowledgeability while many NGOs do not have the funds to hire lawyers. Besides, NGOs have extensive experience in monitoring illegal hate speech which is much more important than a legal degree.

Finally, in most countries, it is unclear when and how the process of application for Trusted Flagger will start. In many EU countries the DSC has not officially been appointed yet and therefore the application process for TFs has not started yet either. This means that NGOs feel insecure and unsure about their position now and in the future, especially because of the aforementioned doubts. It is of essential importance for NGOs to feel included in the roll-out process of the DSA.

3. Digital Services Coordinator

As explained before, every national EU government appoints or creates the institution that will become the Digital Services Coordinator (DSC). The bodies that are or will become DSC can be divided in three main groups: telecommunication offices (e.g. Austria), broadcasting regulators (such as in Slovakia, Council for Media Services will hopefully become a DSC starting 1 September) and consumer protection bodies. The first meeting of the Board where all DSCs meet took place already, with many national bodies being proto-DSCs, i.e. a law to appoint them has been drafted but not passed. The EU Board of DSCs will convene twice a year in person and twice a year online. Within the Board there are a few national DSCs that are clearly leading the process, like Italy, France, Slovakia and the Czech Republic. As these institutions are mostly placed within the ministries of economy, they are in general less experienced with the issue of illegal online hate speech. At the same time, from first conversations with some of them we gather that they are definitely open to cooperation.

However, we also see reason to worry about the appointment of DSCs and the room that is left within the DSA for interpretation by national governments. In the case of Hungary for example, the DSC is the National Media and InfoCommunications Authority. They are not an independent organization and are under the influence of the government. Therefore, the fear is that this DSC will not be able to set up a completely independent process of appointing Trusted Flaggers. The standards that the DSC put out in Hungary states that neither the person monitoring nor the organization that employs them can receive any money from social media platforms. On top of that, Trusted Flagger organizations and their employees have to sign a NDA that lasts even after the employment. This experience is very different from, for instance, the experience with the Dutch DSC, the 'Authority on Consumer and Market' (ACM). The ACM explained that as long as the Trusted Flagger has set up some type of Code of Conduct or guardrails to make sure the organization can do its work independently, it should not be a problem to receive funds from social media platforms. From other organizations we heard that as long as social media platforms are not their sole funders, it should not be an obstacle for being granted the position of Trusted Flagger. In sum, it seems unfair that these DSCs are setting out the rules differently and therefore organizations will be more restricted in some countries than in others.

A further concern is that the DSCs do not have the knowledge, background or network of relationships to effectively carry out their role. The different national DSCs are mostly

media services with little obvious experience in hate speech monitoring (Facing Facts 2024).

However, the same here applies as mentioned before regarding the Trusted Flagger system: most countries do not have the DSC yet in place and therefore a lot remains unclear on what the cooperation between DSC's and NGOs will look like.

This does give us space to indicate as NGOs how we would like to work together with our national DSC's. Ideally, each DSC serves as a connecting institution between the NGOs working in the hate speech- and other related fields, social media platforms and the governmental institutions. The DSC will be able to gather the information and experience of NGOs and connect them with social media platforms and empower the organizations to demand improvement of communication, removal rates and transparency from the platforms. In return, the DSC is always informed on all the trends, patterns and phenomena surrounding illegal hate speech online by the CSOs.

Conclusion

Reviewing our experiences with monitoring, the Trusted Flagger system and Digital Services Coordinators under the DSA, we would like to end this report with recommendations on what can be improved.

Regarding monitoring, we would like to recommend more consistency and less demanding reporting requirements for users. There needs to be more consistency in the reporting format and no constant changing in reporting formats all the time. There also needs to be more consistency in communication with users after reporting content and in removal decisions. Especially, the logic behind removal orders or the refusal of removal orders needs to be more clear and consistent and followed through with. It is impossible to understand why certain content that is clearly illegal will only be restricted in the country where it is reported and in practice is often not even removed at all. Finally, the lack of attention for other languages than English remains a concern that has been verbalized by us many times before and is reiterated in this report once more.

Regarding the Trusted Flagger system, we recommend generating more transparency on the process of appointing Trusted Flaggers. Inclusivity needs to be ensured so that no type of hate speeches are overlooked. So far, it seems that being a Trusted Flagger involves a huge bureaucratic workload which will be complicated for small NGOs that are involved in monitoring. Funds are needed to be able to hire people for the extra workload. The interpretation of the DSA requires explanation. For example, about independence of organizations to become Trusted Flagger, what does that mean? And the most essential factors need to be guaranteed by DSCs in appointing TF's; neutrality and the guarantee that political factors are not involved in choosing the TF's.

Regarding the DSCs, we would like to underline how the position of the DSC is the perfect opportunity to connect the CSOs, social media platforms and governmental institutions to foster more cooperation and exchange of information and knowledge.

Concluding, although the DSA applies to platforms inside the EU, its impact will be global, which is why we must ensure the most successful outcome possible and set an example and global standard in the regulation of this technology that can affect so many around the world. Democracies should enforce principles of transparency and oversight which are needed to ensure that the impact of online platforms do not go

against what they stand for. Think of the example it could be for the rest of the world to create similar laws that lead to digital dignity.

Bibliography

- [Digital Services Coordinators | Shaping Europe's digital future \(europa.eu\)](#)
- Euronews, 'Brussels launches legal action against Elon Musk's X over illegal content, disinformation', Mared Gwyn Jones, 18-12-2023, [Brussels launches legal action against Elon Musk's X over illegal content, disinformation | Euronews](#), accessed on 27-03-2024
- European Commission, 'The Digital Services Act', [The EU's Digital Services Act \(europa.eu\)](#), accessed on 27-03-2024
- European Commission, 'Commission opens formal proceedings against TikTok under the Digital Services Act', [Commission opens formal proceedings against TikTok under the Digital Services Act | Shaping Europe's digital future](#), 19 February 2024, accessed on 27-03-2024
- [Policy Briefing | Facing Facts](#)
- Global Witness 'How big tech platforms are neglecting their non-English language users', 30 November 2023, [How big tech platforms are neglecting their non-English language users | Global Witness](#), accessed on 27-03-2024