Countering Harmful Content Online: Tools, Approaches, and Methods

Berlin 5-6 December 2023

Workshop summary and conclusions

Organised by Violence Prevention Network (VPN)
Authors: Arwa Ben Ahmed and Rositsa Dzhekova

Summary

The EU-funded FRISCO project (*Fighting Terrorist Content Online*) aims to raise awareness and encourage exchange of experience among Hosting Service Providers (HSPs) to support their compliance with the Regulation (EU) 2021/784 on Addressing the Dissemination of Terrorist Content Online (TCO). As part of this effort, on 5 and 6 December 2023, a workshop was held in Berlin, Germany, organised by Violence Prevention Network in collaboration with project partners. It gathered 30 participants in person and online from different countries across Europe and backgrounds. Participants included representatives of four (4) Hosting Service Providers (HSPs), two (2) industry associations, the German regulator Federal Network Agency, reporting platforms, trust & safety experts, representatives of large online platforms such as META, X and Tiktok.

Participants engaged in interactive discussions, identifying challenges and possible low-threshold compliance solutions for HSPs with the TCO Regulation.

The audience was introduced to the FRISCO toolbox, providing added value and support for HSPs to achieve overall compliance and multi-stakeholder cooperation.

Introduction

Day 1 (5 December 2023) started with a welcome statement by Violence Prevention Network's Director of the International Department Jonathan Russel and, Deputy Director Rositsa Dzhekova.

An introduction to the programme by Arwa Ben Ahmed, Senior Adviser to Violence Prevention Network International Department providing background on the framework of this Workshop as part of FRISCO. Violence Prevention Network is leading on activities focusing on increasing awareness of micro and small HSPs around their obligations to comply with the TCO regulation, strengthening preparedness and capabilities of micro and small HSPs to comply with the TCO regulation and promoting partnerships and networking between micro and medium size HSPs and other partners from multiple fields for effective implementation of the TCO regulation.

To achieve this, Violence Prevention Network is developing a manual on best practices targeting primarily micro and small scale HSPs who have limited resources, together with medium HSPs which operate at moderate scale with more infrastructure. In a series of workshops, Violence Prevention Network and its partners are collecting feedback from HSPs and various stakeholders on barriers for compliance with the TCO Regulation but also practical input on tools, approaches and solutions that can be used by these target groups.

The main objectives of the workshop were threefold:

- To introduce the FRISCO project and its toolkit to key stakeholders.
- To assess the status quo in the implementation of the DSA and TCO Regulations in Germany, including progress made, challenges and next steps for both industry and national authorities.
- To discuss the tools, approaches, and methods for compliance and exchange expertise between industry specialists on trust and safety measures for effective content moderation, prevention and monitoring of online terrorist content, while preserving the respect and balance of fundamental rights.

Setting the scene: current trends and challenges in implementation of the TCO Regulation

Current state of affairs of the TCO Regulation and the DSA: *Friederike Wegener, DG Home, EU Commission*

Ms. Wegener began by introducing the relevance and timely organisation of this workshop, as hate speech leading to terrorism is more prevalent online than ever.

She referred to the DSA as a horizontal regulation within the EU related to the online space. She stressed that the DSA does not define "illegal content" as a concept that should be dealt with on a national level for EU Member States, subject to jurisdiction. She then distinguished the DSA from the TCO as a *lex specialis* regulation that specifically addresses terrorist content. Ms. Wegener emphasised the preventive approach of the TCO as a tool to help curb the spread of terrorist content online and the obligation of online service providers to remove content upon request the competent authorities to ensure its effectiveness.

Ms. Wegener explained the background to the adoption of this regulation, which serves to strengthen public-private partnerships within the Union. She also highlighted the challenges related to a commonly accepted definition of terrorism and the clear delineation from violent extremism and called for joint efforts and cooperation between platforms, authorities, and civil society actors to ensure a seamless implementation of the regulation.

State of play of the TCO Regulation implementation in Germany and beyond: Tim-Jesko Gabriel, Subdivision Internet Digitization, Federal Network Agency, Germany

Mr. Tim-Jesko from the Federal Network Agency provided an overview of the responsibilities and tasks of the two competent authorities in the implementation of the Regulation in Germany – namely the *Bundesnetzagentur* – BNetzA (Federal Network Agency) and the *Bundeskriminalamt* – BKA (Federal Criminal Police Office).

He highlighted practical challenges facing both authorities and HSPs about implementing removal orders, concerns related to the authenticity of communication on removal orders and uncertainties related to the requirements of specific measures to be implemented by HSPs.

To address these challenges, several recommendations and initiatives are implemented:

I. Competent Authorities

- **Bundeskriminalamt**
  - issues referrals
  - issues removal orders via PERCI (Art. 3)
  - checks cross-border removal orders (Art. 4)

- **Bundesnetzagentur**
  - monitors the implementation of specific measures (Art. 5.4 et seq.)
  - conducts fine proceedings for certain violations (Art. 18)

- Authenticity of removal orders:

  o Removal orders will be sent only from authorized email addresses (Europol's PERCI platform and national competent authorities[1]).
  o There will be a more targeted approach to informing HSPs about the content that needs to be removed. Instead of generic alerts, they will receive more specific information (per email) about the content in question, such as the type of content or the reason for removal, as per the removal order template, accompanied by a file attachment with more information on the content to be removed.
  o HSPs will potentially have access to Europol's system PERCI: more transparency into the content removal process and related reporting and better understanding among HSPs of the reasoning behind removal orders.

- Specific measures:

  o BnetzA is committed to stay in bilateral exchange with HSPs to ensure the appropriate specific measures are put in place by HSPs where needed, in accordance with Art. 5 of the TCO Regulation.
  o Mr. Tim-Jesko pointed to a study on content moderation commissioned by BNetzA, which outlines the status quo with content moderation and specific measures applied by HSPs. The study further provides examples of good practices, suggests technical measures and recommendations for minimum standards based on the size of HSPs needed to effectively moderate content.

To the date of the workshop no penalties from German authorities for non-compliance with the TCO Regulation have been imposed.

Overall, these initiatives aim at improving collaboration and communication between authorities and HSPs, with the goal of making the compliance process more efficient and effective.

The FRISCO project and HSPs' needs and barriers to compliance: Pal Boza, Co-founder, Tremau and Pierre Sivignon, Project Officer, Civipol

This session presented the key objectives of FRISCO and the findings of the FRISCO mapping report on needs and barriers for compliance with the TCO Regulation among small and micro HSPs, including:

- There is low level of awareness and understanding of the TCO Regulation among small and micro HSPs, compared to the DSA, which has garnered significantly more attention.
- This lack of awareness translates to limited preparedness for compliance. Many small and micro HSPs feel they are at low risk for terrorist content, leading to delayed investments in necessary processes and tools.
- The resource burden discourages proactive investments in compliance measures, often leading to a reactive approach triggered only by imminent threats.

---

[1] https://home-affairs.ec.europa.eu/policies/internal-security/counter-terrorism-and-radicalisation/prevention-radicalisation/terrorist-content-online/list-national-competent-authority-authorities-and-contact-points_en

Industry and Community Responses and Experiences: how can small HSPs benefit?

Content moderation tools for small HSPs: Hadelin Feront, Global Regulations Lead, META

Mr. Feront described the surge of terrorist content from ISIS on social media platforms in 2015/2016 as a turning point for Meta's evolving and multifaceted approach to combat this threat. Starting with clear and readily understandable community standards, they implemented a binary content moderation policy. Proactive detection and automatic tools evolved to take down glorification, support, or representation of terrorist or extremist groups. These include photo and video detection tools, text-based classifiers, behavioral and metadata classifiers, as well as a combination of human review and machine learning towards enforcement at group or network level to prevent fanning out from one violating account or group to others.

To ensure fairness and due process, a robust appeal system was established.

Meta constantly adapted its measures with the evolving regulatory architecture (incl. the NetDG, the Code of Conduct on Countering Online Hate Speech, the TCO Regulation, the DSA, and the EU Internet Forum) and is actively refining its tools and strategies to stay ahead of the curve. Aligning with the guiding principles of GIFCT, Meta's efforts extend beyond their own platforms, fostering multi-stakeholder collaboration for a safer online environment.

Mr. Feront presented the **Hasher Matcher Actioner** tool[2] to the audience. This tool is aimed to help smaller online platforms to identify and flag potentially harmful content through images or video matching, further strengthening the collective defense against hate and extremism.

META combining proactive measures, responsive tools, and collaborative initiatives show its aims to create a safer digital landscape.

---

[2] Hasher-Matcher-Actioner, developed by Meta, is designed to help online platforms identify and remove harmful content like terrorism-related content, child exploitation material, or anything else that violates their policies.

The role of reporting platforms: Yann Lescop, Legal Adviser Trust and Safety, Point de Contact

Mr. Lescop spoke about the role of the French NGO Point de Contact in the treatment of harmful/terrorist content. Through its reporting system, Internet users can report/flag potentially illicit online content encountered. Its team of legal experts analyses and processes these user reports in order to have the content removed if it is classified as illegal.

Given the lack of established case law in French and the need to uphold freedom of speech, harmonized tools for legal interpretation are crucial to support consistent outcomes.

Smaller HSPs often face resource limitations in accessing legal support to review and understand illicit content on their services. This was further reiterated by the representative of a small French HSP participating in the workshop.

On removal orders, responding swiftly is critical, yet relying solely on reactive measures falls short. Proactive measures are needed, but resource constraints and implementation difficulties hinder progress. The path forward should focus on adopting basic tools like reporting systems to protect users, establishing a content moderation board, and implementing automatic detection tools for visuals. By building capacity and fostering collaboration, Mr. Lescop argued for a balanced approach that can effectively address harmful content while safeguarding freedom of expression.

The FRISCO toolbox and solutions for HSPs

The FRISCO project took center stage at the workshop's morning session of Day 2 (6 December 2023). Pal Boza from Tremau presented the tools, frameworks and mechanisms developed under FRISCO to support small HSPs in the implementation of the TCO regulation. Many small and micro HSPs are missing the toolkit when it comes to dealing with terrorist content, whether it be identifying, monitoring, and removing such content.

Mr. Boza demonstrated the **FRISCO Toolbox** consisting of three instruments for compliance developed to respond to the needs and barriers of HSPs to comply with the TCO:
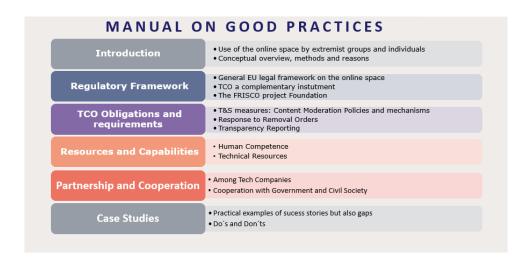
1. Compliance Gap Assessment Tool - self-assessment questionnaire
2. Compliance Workflow Builder - process map
3. Content moderation tools based on Tremau's in-house solution.

Arwa Ben Ahmed, Adviser to Violence Prevention Network presented the scope and structure of the **FRISCO Manual on Good Practices** targeted at HSPs. This comprehensive resource aims to raise awareness and equip companies with practical guidance on navigating the complexities of the TCO Regulation.

This manual will serve as a guide, providing HSPs with actionable information and lessons learned from the experience of their peers in implementing the TCO Regulation. It aims to strengthen trust and safety measures by establishing a solid framework for identifying and mitigating the risks associated with terrorist content and developing effective content moderation policies through the implementation of preventive, proactive and reactive mechanisms.

MANUAL ON GOOD PRACTICES

Introduction	• Use of the online space by extremist groups and individuals • Conceptual overview, methods and reasons
Regulatory Framework	• General EU legal framework on the online space • TCO a complementary instument • The FRISCO project Foundation
TCO Obligations and requirements	• T&S measures: Content Moderation Policies and mechanisms • Response to Removal Orders • Transparency Reporting
Resources and Capabilities	• Human Competence • Technical Resources
Partnership and Cooperation	• Among Tech Companies • Cooperation with Government and Civil Society
Case Studies	• Practical examples of sucess stories but also gaps • Do´s and Don´ts

Discussion highlights: tools, methods, and good practices for HSPs

The concluding session of the workshop buzzed with collaborative energy as participants tackled the crucial question: **how can we equip HSPs with the tools and policies needed to achieve compliance with the Terrorist Content Online (TCO) regulation?**

Through dynamic group discussions, potential solutions emerged, encompassing both policy changes and practical tools.

Basic needs for compliance:

The foundation of the discussion centred on ensuring that HSPs possess the basic building blocks for effective compliance with the TCO Regulation. Participants identified key requirements, including:

a- User blocking capabilities: The ability to swiftly block users suspected of disseminating terrorist content was deemed essential.
b- Clear information channels: Establishing dedicated points of contact for authorities and users alike was seen as crucial for efficient communication and collaboration.
c- Reporting systems: A standardised system for reporting identified terrorist content was emphasized as a critical element of transparency and accountability.
d- Content removal capabilities and processes: Empowering HSPs to swiftly remove illegal content was acknowledged as a cornerstone of TCO compliance.

A toolbox for success:

Moving beyond basic needs, the discussion delved into practical tools that could equip HSPs for proactive and reactive content moderation. The proposed set included:

a- Moderation boards: Establishing dedicated teams trained in identifying and flagging terrorist content was identified as a valuable proactive measure.
b- Automatic detection tools: Leveraging technology for automated detection of harmful content in various formats, such as videos, images, and text, was seen as a crucial time-saving measure.
c- Trusted flaggers: Harnessing the potential of reporting services and trusted flaggers, as envisaged in the DSA regulation, to identify and report terrorist content was recognized as a valuable resource for smaller HSPs.

Follow-up: What is next for FRISCO?

The workshop acknowledged that solutions must be multifaceted, encompassing both policy adjustments and readily available tools. Participants emphasized the need for continued collaboration between stakeholders, including government agencies, tech companies, and civil society organizations, to refine and implement these solutions effectively.

Participants left equipped with insights, solutions, and a renewed commitment to tackling the challenges of TCO compliance. But the journey doesn't end here. Recognizing the crucial role of continued dialogue and knowledge sharing, the FRISCO project announced several upcoming activities for engaging HSPs and other relevant stakeholders:

- **Piloting of e-Training Modules** on the TCO Regulation among HSPs in several EU languages;
- **Testing and validation of the FRISCO toolbox** among HSPs in Europe;
- **Two upcoming workshops** organised by Violence Prevention Network and FRISCO partners on:
    - o *multi-stakeholder collaboration*: how to foster effective partnerships and collaborative approaches in addressing TCO between authorities, HSPs, trust and safety industry and civil society.
    - o *lessons learned*: enabling learning from practical experience on TCO implementation and good practices.

By committing to these upcoming activities, FRISCO partners are taking a proactive step towards building a more collaborative and informed ecosystem for combating online terrorism. These events hold the potential to strengthen existing partnerships, amplify best practices and drive continuous improvement.