4.1.1 Countering Terrorist Content Online: A Best Practice Guide for Hosting Service Providers

Grant Agreement ID	101080100	**Acronym**	FRISCO
Project Title	Fighting Terrorist Content Online		
Start Date	15/11/2022	**Duration**	24 Months
Project URL	TBA		
Contractual due date		**Actual submission date**	
Nature	R = Document, report DEM = Demonstrator, pilot, prototype DEC = Websites, patent fillings, videos, etc.	**Dissemination Level**	PU = Public CO = Confidential CI = Classified
Author(s)	Arwa Ben Ahmed, Rositsa Dzhekova		
Contributor(s)	Isabelle Arnson , Adeline Kugler		
Reviewer(s)	Pierre Sivignon		

This project has received funding from the European Union's Internal Security Fund (ISF) programme under Grant Agreement No 101080100. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Document Revision History *(including peer reviewing & quality control)*

Version	Date	Changes	Contributor(s)
1	19/07/2023	First structure	Arwa Ben Ahmed, Rositsa Dzhekova
2	29/02/2024	First draft	Arwa Ben Ahmed, Rositsa Dzhekova
3	24/03/2024	Second draft	Arwa Ben Ahmed
4	20/04/2024	Peer review and quality control	Pierre Sivignon, Pal Boza
5	10/05/2024	Final Draft	Arwa Ben Ahmed, Rositsa Dzhekova, Isabelle Arnson, Adeline Kugler
6	28/05/2024	Final version for submission after linguistic review and quality control	Arwa Ben Ahmed, Jonathan Russel, Rositsa Dzhekova

Countering Terrorist Content Online: A Best Practice Guide for HSPs

Executive Summary

The prevalence of terrorist content online poses a serious threat to public safety and security. Developments in communications technology and social media platforms continue to facilitate the spread of extremist ideologies, recruitment, and radicalisation. Yet technology also plays an increasing role in identifying and monitoring extremist activity, thereby strengthening counter-terrorism efforts.

The European legislation has responded to these developments by enacting laws designed to curb terrorist content online which result in new obligations for internet companies.

To raise awareness of the EU regulation on the dissemination of terrorist content online (TCO) and the obligations resulting from it, the *Fighting Terrorist Content Online* FRISCO project has developed this best practice manual for Hosting Service Providers (HSPs) and relevant stakeholders.

The manual provides a comprehensive overview of core obligations as well as proactive approaches to identify and remove terrorist content online so that HSPs can fulfil the requirements of the TCO.

In developing this manual, authors benefited from the valuable insights of several stakeholders, including European Union Authorities, HSPs (including large and small online platforms), and civil society representatives. Their contributions helped in shaping the content of this guide.

Table of Contents

List of Terms & Abbreviations

Abbreviation	Definition
TCO	Terrorist Content Online
FRISCO	Fighting Terrorist Content Online project
DSA	Digital Services Act
WP	Work Package
HSP	Hosting Service Provider
ISP	Internet service providers
LEA	Law Enforcement Agency
PPP	Public Private Partnership
SMEs	Small and Medium-sized enterprises
VLOPs	Very large online platforms
VLOSEs	Very large online search engines
GDPR	General Data Protection Regulation
PERCI	Plateforme Européenne de Retraits de Contenus Illégaux sur Internet
CA	Competent Authority
PoC	Point of Contact

1. Introduction

About the FRISCO project

FRISCO OVERVIEW

FRISCO ("*Fighting Terrorist Content Online*") is an EU-funded project for which the main objective is to raise awareness among small online platforms ("Hosting Service Providers") and help them comply with the EU Regulation on Terrorist Content Online ("TCO Regulation").
By supporting efforts to counter terrorist content in Europe, we are helping prevent and counter violent extremism online and create a safer online environment.

FRISCO is implemented by a consortium of 8 partners in the period between 2022 – 2024.

We aim to:

- **Support HSPs** in their compliance journey and content moderation efforts;
- **Alleviate their operational burdens** relating to the TCO Regulation; and
- **Foster multi-stakeholder** collaboration and partnerships.

FRISCO OBJECTIVES

We strive to raise awareness among small HSPs, foster their compliance with the TCO Regulation, and help them protect their services against terrorist content. We support them through:

- **Awareness raising** - we inform HSPs about their new obligations.
- **Tools production** - we develop tools, frameworks, and mechanisms for them.
- **Knowledge sharing** - we share information, experiences, and best practices.

FRISCO ACTIVITIES

We provide HSPs with resources produced to support your compliance journey:

- **A Toolbox**: a self-assessment questionnaire, a process map and a content moderation tool.
- **A Training Programme**: seven training modules accessible via an online platform.
- **Best practices materials**: this manual, workshop insights, articles and further resources.

About this manual

This manual builds on the research and tools developed within the FRISCO project as well as existing guidelines, information, and best practices from the sector, to offer HSPs practical information and guidance.
As determined in the FRISCO Mapping report at the beginning of the project, awareness about the regulation and its implications amongst potentially affected platforms was relatively low.

The results of the **FRISCO Mapping Report** revealed that:

42.4% of respondents were not aware of the TCO Regulation at all.

48.5% of respondents could not define terrorist content.

69.7% of respondents lacked automated measures to identify and remove content.

57.6% of respondents do not moderate user-generated content.

This manual is divided into six chapters. Before delving into the detail of the different measures and obligations of HSPs to address the dissemination of terrorist content, the manual first introduces the FRISCO project (Chapter 1) and provides a brief overview of the most relevant aspects of the TCO regulation, clarifies what can be considered terrorist content, referring the reader to more detailed information in the FRISCO training material (Chapter 2).

Chapters 3 and 4 provide a brief contextual overview of key definitions and outlines the **core obligations** for HSPs to comply with the TCO Regulation namely:

- **Establishing clear Terms of Service (ToS**), explicitly prohibiting terrorist content, and establishing **user reporting mechanisms** (Chapter 3)
- **Appointing a Point of Contact or Legal Representative**, to be the contact point for competent authorities (Chapter 4.1)
- **Responding to removal orders upon request from competent authorities** and dealing with content posing immediate threat to life (Chapters 4.2 and 4.3)
- **Providing a Complaint Mechanism** for users to report potentially terrorist content and preserving content (Chapters 4.4 and 4.5)

Chapter 5 explains the **specific measures** outlined under the TCO as additional measures required for HSPs that have been exposed to terrorist content by use of:

- **Content Moderation** efforts to proactively identify and remove terrorist content and related tools and approaches (Chapters 5.1 and 5.2)
- Making use of **industry partnerships and collaboration frameworks** (Chapter 5.3)

- **Transparency Reporting** by publishing regular reports on the volume of terrorist content removed and the measures taken to address it (Chapter 5.4)

The final Chapter 6 provides conclusions.

2. The TCO Regulation: Overview and Scope of Application

2.1. The TCO regulation – a paradigm shift for HSPs

In April 2021, the EU adopted the regulation on addressing the dissemination of terrorist content online (TCO), which entered into force on 7 June 2022. The regulation supplements existing legal instruments that govern online content dissemination in the EU. It addresses online terrorist content and is aimed directly at detecting and removing terrorist content. This means that HSPs are expected to engage in reducing access to online terrorist content and to cooperate with the authorities in its dismantling.

Previously, courts had difficulty determining how platforms were linked to terrorist content. The TCO clarifies the situation by making platforms more responsible for removing such content.
(**For more details, see FRISCO Training Module 4**).

The below table provides a summary of categories of HSPs that fall under the TCO:

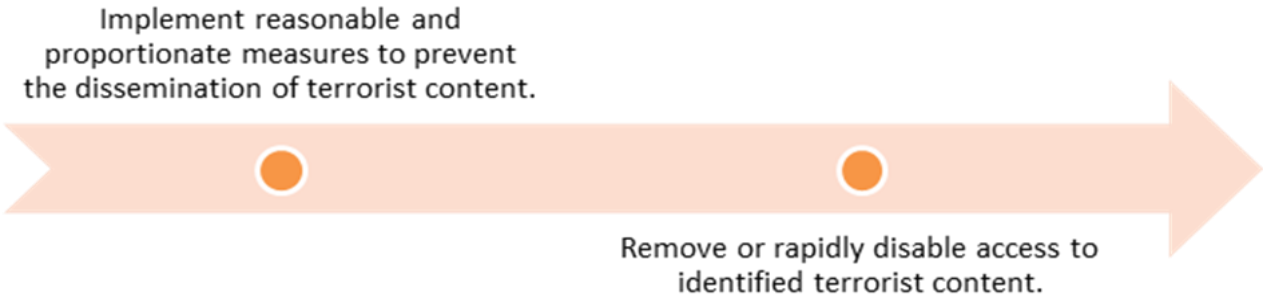
Who is concerned	**Target Audience simplified**
Article 1 of the TCO Regulation sets out its scope. This legal text applies to HSPs offering services in the EU, regardless of their principal place of establishment, provided that they disseminate information to the public. - Hosting service providers **operating in the EU or serving EU users**. - Platforms like **social media, video-sharing, image-sharing, and audio-sharing services**. - Websites and platforms that **store and share user-generated content**. - HSPs **offering services in the EU**, whether they are mainly based within the Member States or not. - TCO applies to **small and micro HSPs** **Exemptions:** - **Educational/Journalistic/Artistic Content:** Content for these purposes is not considered terrorist content, even if controversial.	**Hosting service provider:** Stores information provided by and at the request of a content provider (user) and disseminates it the public making available the information provided. • *e.g :Facebook or Tiktok.* • These are types of online platforms include: **Online storage and distribution platforms** • *e.g. webhosting, online media sharing platforms, file storage and sharing platforms.* **Networking, collaborative production and matchmaking platforms** • *e.g. providers of social media, collaborative production, online marketplaces, collaborative economy and online games* **Selection and referencing platforms** • *e.g. rating and reviewing platforms*

- Fundamental Rights: **Freedom of expression and information are protected.**

The TCO Regulation marks a paradigm shift for HSPs:

- It applies to:
  - small and micro-enterprises defined in Commission Recommendation 2003/361/EC24. These include companies with fewer than 50 employees and a turnover below EUR 10 million.
  - HSPs offering services in the EU, regardless of their location, with content accessible to the public.
- Passive content moderation is no longer enough, (once a HSP has been exposed to terrorist content), proactively preventing the dissemination of terrorist content on its services then becomes an obligation.
- Failing to comply can result in significant legal repercussions, including fines and business suspensions considering size of the company with circumstances listed under (Art. 18 (2) (f) TCO Regulation).

In summary, HSPs are called to:



Addressing Knowledge gaps:

The below case study stems from a direct interview conducted by the FRISCO team with a small HSP. The platform, based in France and Germany, boasts 200,000 monthly active users and primarily hosts user-generated photos and videos. They utilize a combination of automated and human content moderation practices.

Our conversation with the HSP revealed a critical knowledge gap regarding regulations surrounding terrorist content online. While they employ content moderation tools and address basic concerns, a deeper understanding of the TCO Regulation is necessary.

Are small HSPs prepared to address terrorist content?

A small online platform interviewed by Violence Prevention Network team, with 200,000 monthly active users in France and Germany. The platform hosts photos and videos (primarily user-generated) and employs a mix of automated and human content moderation.

The challenge: While the platform faces legal requests and police reports for harmful content, it lacks awareness of the specific requirements and procedures outlined in the TCO Regulation.

Observations:

- Limited knowledge of TCO: HSP has not received any removal orders under the TCO Regulation, despite having encountered content flagged by their automated system as potentially terrorist-related (Nazi symbols, ISIS flags).
- Reliance on reactive approach: HSP waits for issues to arise before taking action.
- Insufficient understanding of legal obligations: The platform owner expressed uncertainty about their legal obligation to share information with law enforcement when requested. This highlights the need for a clearer guidance for small platforms.
- Focus on basic content moderation: HSP prioritises content moderation tools that filter out nudity, weapons, gore, and terrorist symbols and did not include appeal mechanism. While this addresses basic concerns, it might not be sufficient to comply with the TCO's broader definition of terrorist content.

Recommendations:

- HSPs should seek legal counsel to understand their obligations under the TCO and the DSA.
- They should proactively review the TCO regulations and update their content moderation policies accordingly.
- Investing in educational resources and training for staff on the TCO and user safety protocols is crucial.
- Establishing clear communication channels with law enforcement is essential for efficient information sharing.

Before diving into solutions, let's explore the definition of terrorist content and how it specifically impacts HSPs.

2.2. What is terrorist content and how are HSPs affected?

WHAT IS TERRORIST CONTENT?

The TCO Regulation (Art. 2(7)) defines **'terrorist content'** as any type of material (including text, audio, video) that incites terrorist offences[1] or glorifies terrorist acts, provides instructions for making weapons or using them for terrorist purposes, or promotes terrorist groups or recruitment/participation therein. However, the regulation excludes content used for education, journalism, artistic expression, research, or raising awareness against terrorism. This ensures a balance between protecting public safety and freedom of expression[2].

This definition plays an important part in determining the scope of application and guiding LEAs and HSPs. It is necessary for HSPs and LEAs to assess and identify content that falls under this definition.

To assess terrorist content, HSPs should consider:

- **The nature of the material:**
  - Content that threatens or instruct violence.
  - Content that glorifies terrorist acts or dehumanises the victims.
- **The context:**
  - Journalistic or media content, research, or creative expression?
  - Who is the target audience?
  - Can it be misunderstood or manipulated?
- **The source:**
  - Is it linked to known terrorist groups or individuals?
  - Is the content linked to entities on official terrorist designation lists3?
  - Is it distributed through known or suspected channels?
- **The effect:**
  - Can it inspire or facilitate violence?
  - Does it constitute a real threat to public safety?

Terrorist content is not the only source of concern in the digital environment and represents only a small proportion of the harmful content published online. It must be distinguished from other types of illegal content, such as online hate speech, hate crime or other violence, as well grey area of "borderline" content that may not be illegal but can be highly damaging. (**Please see FRISCO Training Module 2 for more details**).

[1] At the EU level, there is no definition of terrorism per se : only terrorist attacks and offences are defined by the EU Directive 2017/541, Art. 3
[2] Article 2 of the TCO regulation
[3] Council of the European Union website: This is the official source for the EU terrorist list. The website provides information on the sanctions regime, criteria for listing, and the procedures for listing and delisting groups. You can find it here: EU Terrorist List: https://www.consilium.europa.eu/en/policies/fight-against-terrorism/terrorist-list/

There is a need to balance **countering and dismantling terrorist content online, and freedom of expression**[4] in order to protect one of the EU's essential foundations for a democratic society.

Remember:

Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers[5].

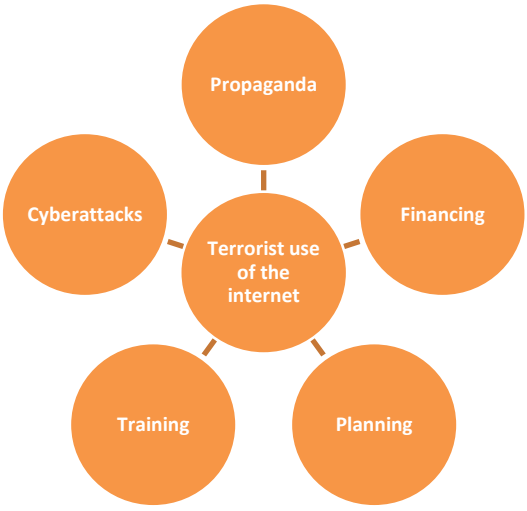
This also means the freedom and pluralism of the media shall be respected. This right is enshrined in Article 11 of the Charter of Fundamental Rights.

- Strike a balance between freedom of expression and public safety.
- Respect user privacy.
- Seek expert advice if in doubt.
- Document your assessment process.

TERRORIST USE OF THE INTERNET

More than two decades into the 21st century, we have seen numerous examples of terrorists turning to new and emerging technologies such as drones, virtual currencies, social media, and encrypted technologies. With AI becoming increasingly accessible, it is crucial to stay ahead of the curve and be prepared for any eventuality involving its misuse[6]. In this context, it is important to have clear conceptual definitions and understanding of emerging threats including terrorism, and how the internet is used for terrorist purposes.

To spread mass fear, radicalise, and recruit members, terrorists engage in increasingly varied propaganda on a variety of platform types to reach a wider audience. Social media, messaging platforms, podcasts and the dark web have been used as recruitment tools to incite violence and facilitate terrorist attacks, and these public-facing surfaces are crucial for recruitment and spread of ideology.



Smaller platforms, which are less well-resourced and less closely monitored, make it easier for terrorist

[4] Article 6 of the TEU "The Union recognises the rights, freedoms and principles set out in the Charter of Fundamental Rights of the European Union of 7 December 2000, as adapted at Strasbourg, on 12 December 2007, which shall have the same legal value as the Treaties"
[5] EU Commission, Freedom of expression and information, accessible here: Freedom of expression and information - European Commission (europa.eu)

entities to disseminate extremist content without strict moderation. Below we explain ways of misuse.

HOW ARE HSPs AFFECTED BY TERRORIST USE OF THE INTERNET?

- **User-generated content:** all platforms enabling the sharing of user-generated content are at risk to be exploited by terrorist actors, to become a cog in relatively complex dissemination strategies.
- **Multi-platform approach of terrorist (mis)use**: terrorist actors are utilizing multi-platform tactics to disseminating terrorist content, including platforms of all sizes (including small and micro HSPs). GIFCT identifies the following main types of platforms used for terrorist content dissemination[6]:
    - Attracting attention and redirecting the target audiences from **beacon platforms** (e.g. large social media, messenger, and video sharing apps), where content is quickly removed.
    - Storing relevant content in **content stores** (content storage, hosting, sharing and pasting sites, archive services).
    - Centralising and facilitation content diffusion via **aggregators**: pasting sites (e.g. JustPaste.it, 1fichier.com), social media (e.g. Vkontakte)
    - Avoiding detection through **circumventors** (VPNs, archiving and file mirroring services, Dweb).
- **Uneven exposure of HSPs:** Services such as **file sharing and file storage, archiving and content pasting are the most exposed**, but social media, link-shortening, emailing, messaging, video sharing and hosting, web hosting, forum, audio streaming or search engines are also at risk.[7]
- **Small platforms with weak content moderation** (due to lack of capacities or robust policies) and attractivity features **are at higher risk**! Based on data by the Terrorist Content Analytics Platform (TCAP), these features are security, stability, audience reach and usability.[8]

To learn more on how terrorist actors use the internet and how HSPs can be exploited, please see **FRISCO Training Module 3**, **FRISCO Mapping report** and the **FRISCO Insights**.

The Christchurch Mosque attack, where terrorist content spread rapidly online, serve as a stark reminder of the immense harm that can occur if such content isn't swiftly addressed.

[6] GIFCT Technical Approaches Working Group: Gap Analysis and Recommendations for deploying technical solutions to tackle the terrorist use of the internet.
[7] Terrorist Content Analytics Platform (2023), TCAP Insights: Understanding Terrorist Exploitation Online by Tech Platform Type. https://www.terrorismanalytics.org/research-news/TCAP-Insights-Terrorist-Exploitation-by-Type
[8] Tech Against Terrorism (2021), GIFCT Technical Approaches Working Group: Gap Analysis and Recommendations for deploying technical solutions to tackle the terrorist use of the internet, p. 17.

Manifesto of the Christchurch Mosque attacker analysis and the terrorist use of the internet - Tech Against Terrorism

The global tech sector responded quickly in taking down the terrorist's manifesto and videos of the attack: Facebook prevented 1.5m attempts by the public to re-upload the attack video within the first 24 hours.

The majority of smaller tech platforms hosting the video and manifesto quickly removed the video and the manifesto from their platforms.

The use of internet technologies in this attack resembled the methodology of ISIS and al-Qaeda. Smaller file-sharing platforms were used with large platforms as "beacons" guiding users to outbound URLs. "Supporter networks" amplified terrorist propaganda by re-sharing and re-uploading material across an increasingly broad and fragmented range of smaller platforms.

The Christchurch attacks demonstrate that terrorist use of technology is a threat that affects the entire tech eco-system and that tech platforms are often exploited in combination.

3. Aligning your Trust & Safety efforts with the TCO Regulation

By addressing both policy and process aspects, this chapter aims to provide HSPs with actionable guidance on integrating Trust and Safety efforts into their compliance strategies, ultimately contributing to a safer online environment.

The TCO regulation imposed a new set of liability and transparency obligations on HSPs and requests them to deploy additional vigilance and risk prevention strategies when exposed with terrorist content. These requirements depend on the type of HSPs but have both a policy and process dimension. Different obligations apply to HSPs depending on whether they have been exposed to terrorist content. The table below outlines what the TCO requires from platforms at a minimum, and what they need to implement upon being exposed to terrorist content. At a minimum, HSPs in scope of the TCO must establish points of contact and legal representation in the EU, and clearly prohibit terrorist use of their services in their Terms of Service. Once an HSP has received two or more official removal orders, they are "exposed" to terrorist content. At this point, a broader suite of obligations kick in.

Navigating the TCO complexities requires a multi-faceted approach. While understanding the legal specifics is essential, the real difference lies in building a strong culture of trust and safety within a platform through defining appropriate terms of services and content guidelines for users. Aligning

your Trust and Safety efforts with the TCO can help create effective and compliant internal policies and processes.

3.1. Policy Development: Terms of Service and Content Guidelines

Defining the core measures or sets of rules and guidelines defining the allowed conduct on a platform, is crucial to frame the HSPs' duty of care obligations. This requires a proactive approach to prevent terrorist exploitation of HSPs and the first step towards that is **defining content guidelines and putting in place a robust Terms of Service/Terms and Conditions that clearly and decisively prohibit the dissemination of terrorist content** (legally mandated by the TCO Regulation under Article 7) and other harmful content on their platforms.

Prevention is Key: By clearly defining what is allowed and what is not, these terms and community guidelines aim to proactively create a safer and more trusted online environment for all users while minimizing risks associated with TCO regulations.

Content guidelines

The requirements for content guidelines vary depending on the focus of the platform, the type of content hosted and the targeted user. Platforms that host user-generated content, have multiplayer online games, or aimed at children will generally require, different, but usually more elaborate content guidelines[9].

What to consider when defining your policies in your Terms of Service:

- **Platform purpose and target**: HSPs targeting children versus HSPs targeting adult users will define their policies differently.
    - Consequently, it is crucial to understand who may (mis)use your platform and create your policies accordingly.
- **Structure and architecture:** A marketplace comment section will require different policies than a live multiplayer online game.
    - Keep this in mind when defining your policies.
- **Core principles:** Your core principles need to be the centre of your policies. Think about what you want your platform or services to embody and translate that into guiding principles for your policies.
    - For example, Wikipedia's core principles – 'neutral point of view', 'verifiability', and 'no original research' - drive their policies for what content is acceptable or not.

[9] Goldmedia GmbH Strategy Consulting (2023), Status Quo of Specific Measures of Hosting Services for Content Moderation. Study commissioned by the Federal Network Agency.

- **Legal requirements:** Make sure that your principles are in line with the legal requirement of the TCO such as defining the procedures to prevent and take down harmful content.
    - These procedures/measures cover different content formats - video, multimedia, text and images.
- **Flexibility**: Your policies should not be set in stone. Make sure to account for the changing technological and regulatory environment and update your policies accordingly
    - Don't forget to notify your users when changes have been made.

Terms of Service

Establishing clear Terms of Service and/or Community Guidelines outline acceptable behaviour and content on an HSP. By establishing these rules, HSPs demonstrate a good faith effort to regulate content. ToS shifts responsibility for uploaded content to the user. Although it does not eliminate liability completely under the TCO Regulation, when users agree not to violate these terms, it helps limit the HSP liability for user-generated content.

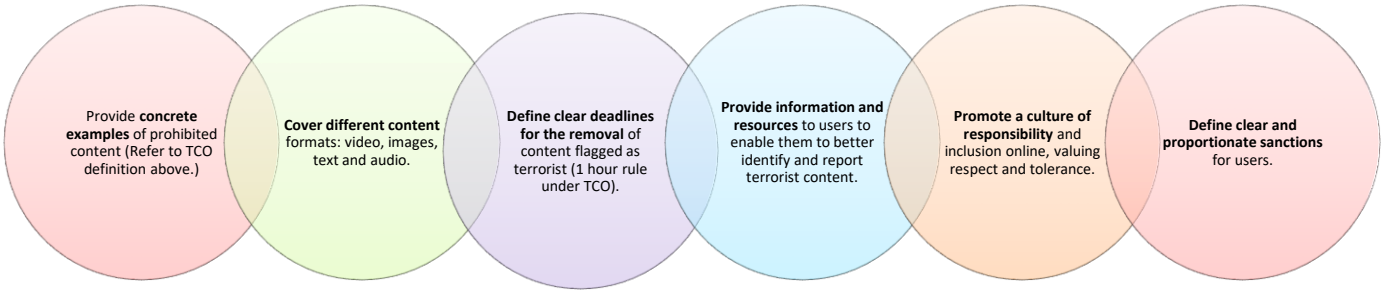
These Terms of Service (also known as Terms of Use, or Terms and Conditions)[10] are an essential component to address harmful content, including terrorist content, and ensure compliance with the TCO. According to Article 7. 1 of the TCO "***Hosting service providers shall set out clearly in their terms and conditions their policy for addressing the dissemination of terrorist content****, including, where appropriate, a meaningful explanation of the functioning of specific measures, including, where applicable, the use of automated tools*".

In short, the TCO requires HSPs to prohibit terrorist activity and content on their platform and outline how they use "specific measures" to address terrorist content. You will learn more on specific measures in Chapter 5.

The ToS is a type of **binding agreement** between a HSP and its user, defining the rules for use of the service (a website, an app, etc.) and defines allowed and prohibited content and conduct by users. Having clear ToS will help HSPs prioritise clear communication with users and maintaining transparency about content moderation practices is key to build trust and foster a responsible online environment that effectively minimizes the dissemination of harmful terrorist content.

When drafting Terms of Service, HSPs should consider:

[10] TCO Article 2. 8 'terms and conditions' means all terms, conditions and clauses, irrespective of their name or form, which govern the contractual relationship between a hosting service provider and its users.



For relevant examples of TCO-compliant Terms of Service or policies prohibiting terrorist content, see the following from HSPs:

- Dailymotion's "Rules related to terrorist content" provide a clear definition of how Dailymotion defines terrorist content.
- Kofi's "Content Guidelines" provide a clear prohibition of terrorist content, distinguishing between hate speech, religious and political radicalism, and terrorism and violent extremism.
- Bitly's "Acceptable Use Policy" provides a clear prohibition of terrorist use of Bitly.

For more insights on how to build effective Trust & Safety policies and processes, consult **FRISCO Training Module 1**.

3.2. User reporting System for Illegal and ToS-Violating Content

Reporting Systems and Appeal Mechanisms

After establishing your terms of service and content guidelines, it's crucial to align your processes with these policies. One essential step is to set up user reporting systems that enable users to flag and report illegal or ToS violative content.

Note that while establishing a user reporting mechanism is part of the specific measures required by the TCO Regulation (Art. 5) only after an HSP has been exposed to terrorist content (see Chapter 5

for further details), is a proactive and preventive approach to ensure preparedness and demonstrate compliance recommended. Equally important is to ensure that HSPs establish means for users to appeal any removal or blocking decisions in advance, as the reaction time after receiving a removal order is limited.

User reporting mechanisms can significantly support your moderation efforts. These systems allow users to flag content and behaviors that violate your terms of service and content guidelines. To streamline moderation practices, user reporting systems can also be integrated with content moderation software. Additionally, **appeal mechanisms** are vital to preserve freedom of expression and enable users to contest moderation decisions they believe were incorrect.

It is also important to outline your response processes for handling reported content. Your policies should guide the development of these processes. Detailing your response procedures is important to ensure consistency and transparency in your moderation efforts. Use the following checklist:

1. Set up user reporting systems.

- Specify a contact point (e.g., email address) and implement a dedicated report button.
- Integrate user reporting systems with content moderation software to streamline moderation practices and improve response time.
- Provide updates to users about the status of their reports.

2. Establish appeal mechanisms.

- Provide a contact point or easily accessible form for users to appeal moderation decisions.
- Set clear guidelines for the appeal process, including timelines, and required information.
- Assign a dedicated team/individual to review appeals and make decisions.
- Communicate the outcome of appeals to users, including the reasoning behind the decision.

3. Outline response processes for handling flagged content.

- Develop a clear escalation process for handling diverse types of content violations, with varying levels of severity.
- Assign a dedicated team or individual to review flagged content and make decisions based on established guidelines.
- Set response time targets for reviewing and addressing flagged content. Do not forget about the 24h for removal orders coming from authorities.
- Maintain records of moderation decisions and regularly review them to identify trends, improve processes, and ensure consistency.
- Provide training and support to moderators to ensure they are up to date with guidelines and best practices.

4. Removal Orders

4.1. Establishing Points of Contact and Legal Representative

The TCO Regulation establishes a clear system for swift removal of terrorist content online. This is realized through removal orders sent by the competent authorities to the HSP hosting the relevant content. A crucial element of this system is **the obligation to designate a contact point**[11] to which the removal order is to be sent.

If the HSP is not based in the EU, HSPs should also designate a person as their **legal representative**, who is responsible for receiving, complying with, and enforcing removal orders and decisions issued by competent authorities.

Learn from best practices to ensure your Point of Contact (PoC) is effective and compliant with the TCO Regulation:

Accessibility

Availability: Ensure the PoC information is communicated to your CA and available publicly. Establish a dedicated page for "Content Reporting."

24/7 Accessibility: Aim for round-the-clock accessibility. If full-time staffing isn't feasible, consider solutions like auto-responders acknowledging receipt and outlining next steps.

Multilingual Support: Support at least one official language of the EU Member State where your HSP is established. Ideally, offer options in multiple EU languages for wider accessibility.

Good Practice: When communicating with the authorities , provide the email address of your PoC

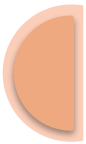
Clarity and Transparency

Clear Instructions: Provide clear instructions on how to submit removal orders to CA. Specify accepted formats (secure email channels for exp).

Confirmation: Consider sending an automated confirmation email upon receiving a removal order, letting the sender know their request has been received.

Language Transparency: Clearly state the languages in which your contact point can receive and process requests.

Good Practice: Consider including information about your removal order process in your terms of service and in your transparency report when applicable, demonstrating your commitment to addressing terrorist content.



Efficiency

Dedicated Team: Designate a team or individual within your company to handle removal order requests. This ensures proper training and efficient processing.

Response Time: Define a clear internal response time for processing removal orders. Ideally, aim for a swift response within a specified timeframe.

Clear Process: Establish a clear, documented process for reviewing and responding to removal orders. This ensures consistency and compliance with TCO templates.

Good Practice: Train the designated team on the TCO regulation and best practices for handling removal orders.

4.2. Receiving and responding to removal orders

This section outlines the procedures and obligations for HSPs regarding removal orders for terrorist content issued by competent authorities in the EU.

Once your PoC is set up, handling removal orders of terrorist content issued by competent authorities in EU Member States requires familiarity with the process.

UNDERSTANDING THE PROCESS

[11] Article 15 of regulation (EU) 2021/784 on addressing the dissemination of terrorist content online.

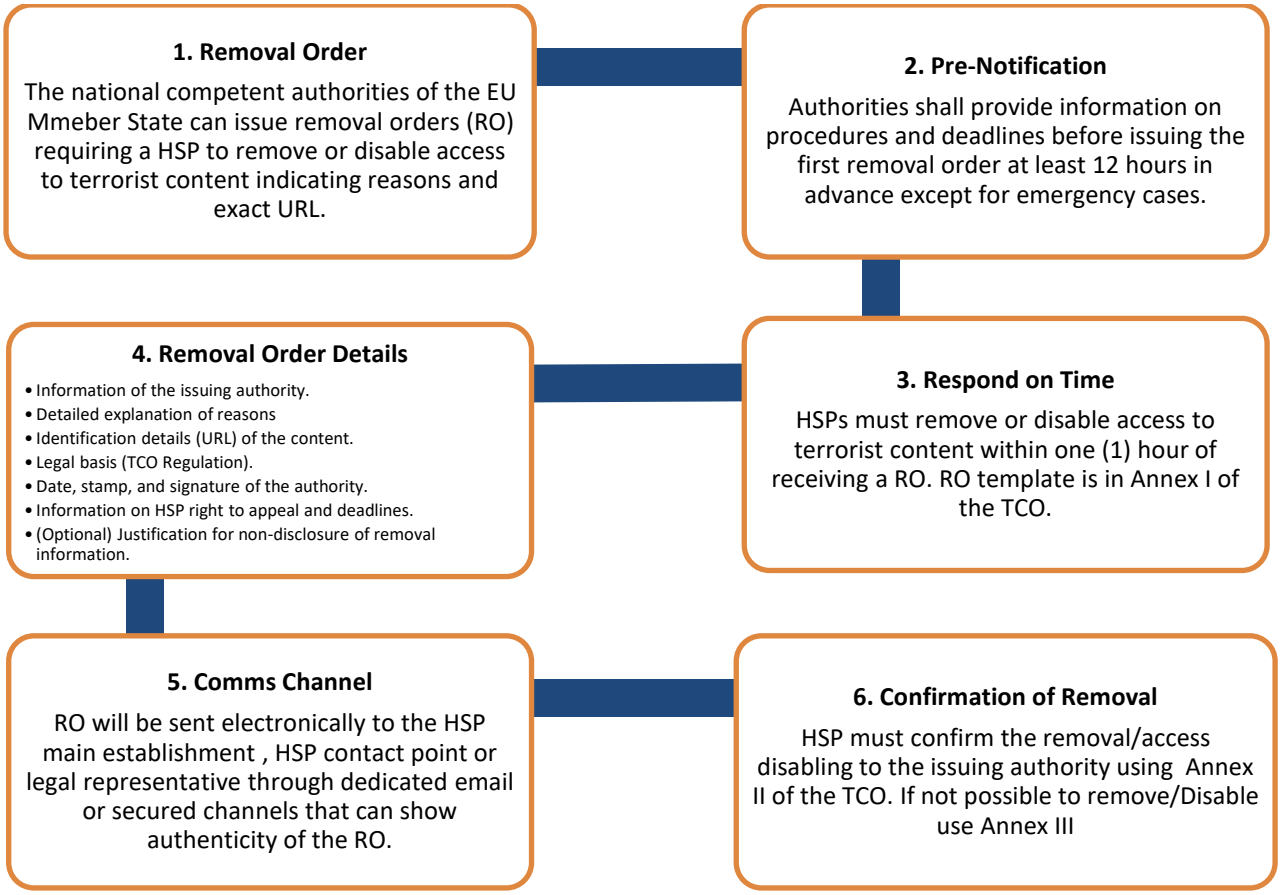
The TCO provides key points to consider:

- **Harmonized Procedures:** A consistent approach across the EU when receiving removal orders.
- **One-Hour Removal:** Terrorist content identified must be removed or access disabled within one hour from receipt, across all EU Member States where your service operates.

However, HSPs should be aware of exceptions to these points:

- **Emergency Cases:** Competent authorities must provide information and deadlines at least 12 hours before the first removal order, except in emergencies.
    - Emergencies involve situations with imminent threats to life (see below for more details) or ongoing harmful events.
    - The authority issuing the order must define and justify the emergency within the removal order.
- **Force Majeure and Impossibility:** If removing content within one hour is impossible due to unforeseen circumstances (force majeure) or technical/operational limitations, the concerned HSP should notify the issuing authority immediately, and comply as soon as the situation is resolved.

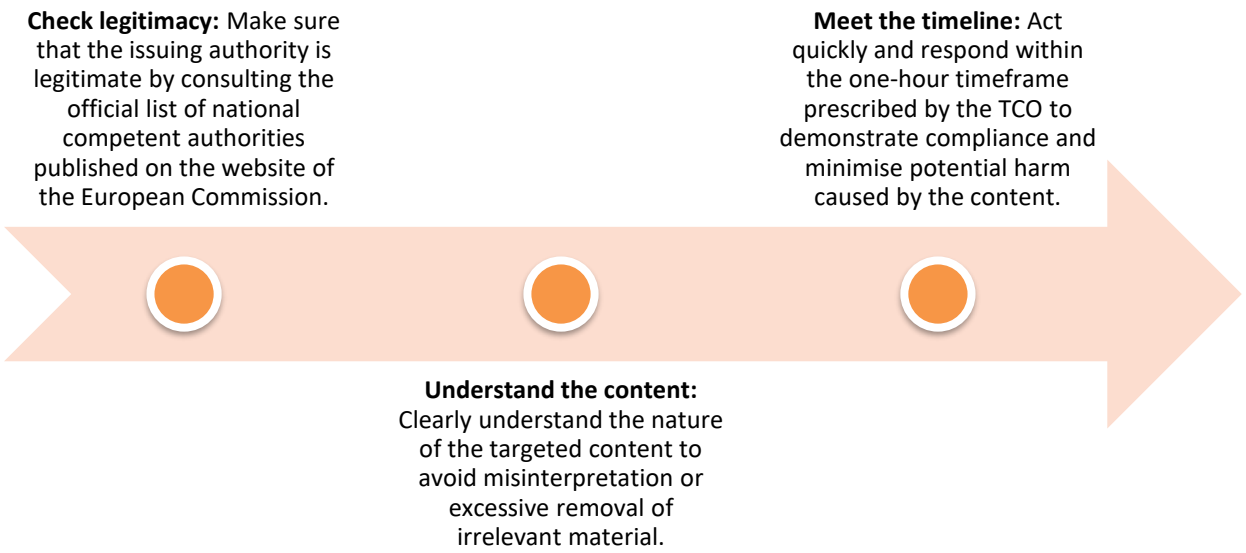
Considering the above key points, it is necessary for HSPs to train staff to identify terrorist content according to the TCO Regulation's definition. This ensures accurate content assessment and efficient removal. Clear internal procedures should be established for receiving, processing, and responding effectively. This includes maintaining clear and up-to-date records of all interactions with authorities and develop a communication strategy for informing users about content removal due to legal orders. This should balance transparency with any legal limitations on disclosing.



SWIFT AND ACCURATE RESPONSES TO OPTIMISE SYSTEMS

- **Establish an Internal Alert System:** Establish a system that immediately notifies a designated Incident Manager upon receiving a RO.
    - Consider implementing machine learning to automatically scan incoming emails for keywords related to terrorism, triggering an alert even outside of regular business hours.
- **Removal Order Checklist:** Develop a clear, internal checklist with specific criteria to assess flagged content.
    - This ensures content meets the removal requirements and avoids misinterpretations.
- **Geo-Blocking Capabilities:** Implement tools for swift geo-blocking of flagged content within the crucial 60-minute window.
    - This minimizes the content's reach while adhering to the TCO's timeframe.
- **Notify content providers:** notify content removal/disabling, explain the reasons and right to challenge, but exceptions exist for public security concerns as determined by authorities.

Upon receipt of a removal order, here is what you need to do:



FAQs ON PRACTICAL ISSUES CONCERNING REMOVAL ORDERS (RO)

*Q: **Who** can issue removal orders?*

A: EU **national competent authorities** can issue removal orders. Check the list of contact points for each EU Member State here.

*Q: **How** can I verify the legitimacy of the removal order?*

A: Check the **list of the official contact points of designated national competent authorities** above. If in doubt, contact the national competent authority in your country directly to verify the request, if this can be done swiftly.

Please note that authorities are unlikely to notify you of a removal order by phone but will send a **standardised email to your designated PoC with the removal order template** and the information defined in Art. 3(4) of the TCO Regulation about the content concerned (incl. a precise URL). Consult the removal order template in Annex I of the Regulation.

It is strongly advised to **become familiar with any country-specific guidelines** or information campaigns issues by national competent authorities on the implementation of the TCO Regulation, as specific procedures might differ from one EU Member State to another.

For example, the German regulator BNetzA has a dedicated page with key information on the TCO Regulation and its implementation in Germany, including contact information and further resources.

*Q: **What if** I do not have the capacity to monitor and respond to incoming emails 24/7?*

A: For HSPs receiving a removal order for the first time, the national authorities are required to send **a pre-warning at least 12 hours before** issuing the removal order, except in emergency situations.

Once you set up PoC and shared it with the respective competent authority, consider implementing as a machine reading software of incoming emails to filter keywords such as terrorism and **trigger an alert system** so that an incident responder would receive an immediate notification outside of business hours.

*Q: **What happens if** an HPS doesn't have a main establishment or legal representative in the EU country that wants to remove content?*

A: If the HSP doesn't have a presence in the country issuing the removal order (Competent Authority), that country must act. They will send a copy of the removal order to the relevant authority in the country where the HSP is established (Residency Competent Authority). This ensures the removal order reaches the right party, even if the HSP operates across borders.

*Q: **Are there any exceptions** to responding to removal orders in the required timeline (one hour)?*

A: Yes, there are a few exceptions to removal orders one hour rule:

Force majeure or technical/operational impossibility: if unforeseen circumstances or technical limitations prevent the removal of content, this can be an exception. However, you must immediately notify the issuing authority of the situation.

Manifest errors or insufficient information: if the RO contains clear errors or lacks crucial information, you can request clarification from the authority before proceeding with removal.

Check: Annex III Information about the impossibility to execute the removal order (Article 3(7) and (8) of the TCO Regulation).

*Q: **Who has the right to challenge** a removal order issued for content under this regulation?*

A: Both hosting service providers who received the order and content providers whose content was removed can challenge it.

4.3. Threat to life

If HSPs become aware of terrorist content that poses an immediate threat to life (Article 14), they must notify the competent authority in the EU country where the threat is located without delay. If the location of the threat is unclear and the EU Member State concerned cannot be identified, the HSP must inform swiftly the contact point in its own country and pass on the information to Europol.

This procedure enables a rapid response to possible terrorist attacks by involving the relevant law enforcement authorities, with Europol acting as a central hub for the exchange of information if the location of the threat is uncertain.

Become familiar with the channels through which you can report content involving imminent threat to life to the national authority in your country, including 24/7 hotlines, dedicated e-mail addresses or web forms. These might differ from the regular communication channels for removal orders.

In case you are unable to identify the Member State concerned by the content, you should notify the competent authority in the EU Member States where you are located and transmit information concerning the terrorist content to Europol.

When reporting to the authorities, refer explicitly to the issue of imminent threat to life and provide as much information as possible to enable quick processing by the authorities, including screenshots and URLs, as well as your contacts for follow-up questions.

4.4. Appeal process and complaint mechanism

The TCO regulation strikes a balance between takedown efficiency and user rights. While it exempts HSPs from the legal burden of assessing every reported piece of content, it prioritizes user agency through a robust appeal mechanism. This ensures users have the right to challenge moderation decisions they believe are incorrect. This is crucial for situations where automated systems or human moderators might misinterpret content. The TCO Regulation compels HSPs to establish clear and accessible appeal procedures, allowing users to contest content blocking or removals or any further restrictions they deem unjustified.

To achieve a comprehensive approach, Article 10 of the TCO Regulation provides for the need to establish effective and accessible complain mechanisms. Implementing a user complaint system is key where a content has been removed or access to it has been disabled as a result of a removal order or specific measures pursuant to Article 5.

To allow for proactive reporting of suspected terrorist content:

1. First, you need to make sure that you notify users of content removed or disabled in a timely manner and provide information on the reasons for this.
2. Second, you need to establish a user-friendly complaint mechanism and ensure that complaints are dealt with promptly and in full transparency towards the content provider, such as providing a copy of the removal order itself.

Exceptions exist if the legal authority issuing the removal order deems non-disclosure necessary for public security reasons (e.g., ongoing investigation processes for terrorist

activities). This non-disclosure period can last up to six weeks, with a potential extension under specific circumstances.

For further practical information and guidance, refer to the **FRISCO Toolbox** and the **Process Map tool.** FRISCO's process map is an interactive tool that structures and describes the entire compliance process with the TCO Regulation and related duties for HSPs in a holistic way, from exposure to terrorist content through to transparency reports. Focused on HSPs' operational needs, it provides a precise breakdown of the TCO Regulation, step by step, and is based on a holistic and chronological approach to compliance.

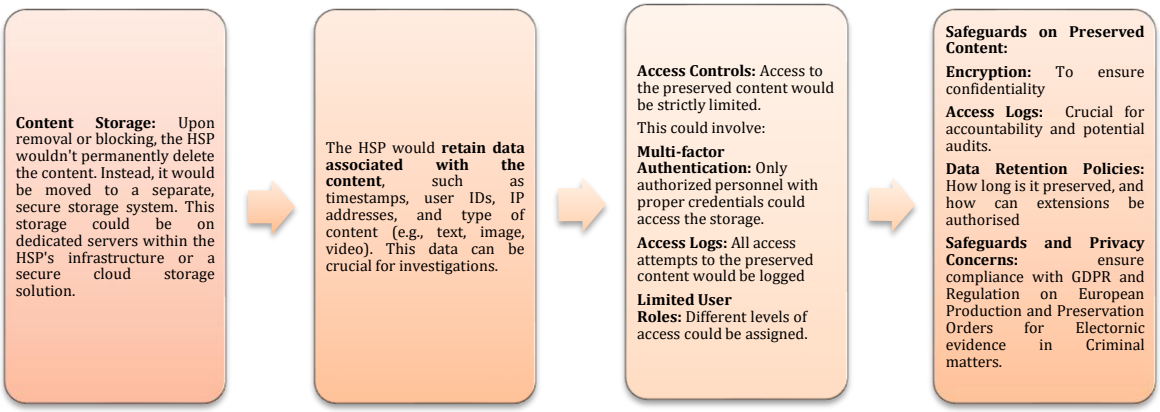
To help with compliance, complete the **FRISCO Self-assessment questionnaire** aimed at helping small and medium HSPs understand their current level of compliance with the TCO Regulation. It provides HSPs with a compliance score, which helps them situate themselves in the path to full TCO compliance.

Removal and Appeal order templates can be found in the following Annexes to the TCO regulation:

- Annex I Removal Order (Article 3 of the TCO Regulation)
- Annex II Feedback following removal of or disabling of access to terrorist content (Article 3(6) of the TCO Regulation)

4.5. Content preservation

- Preserve removed or blocked content for 6 months, and longer if requested by authority (Article 6).
- Apply technical and organisational safeguards on preserved content. Here is how:



5. Specific measures for addressing terrorist content

The TCO Regulation does not require HSPs to proactively monitor terrorist content or search for illegal activity, nor does it mandate the use automated tools for content moderation. However, it does introduce specific measures under Article 5 for HSPs **"Exposed to Terrorist Content".**

A HSP is considered exposed to terrorist content if:

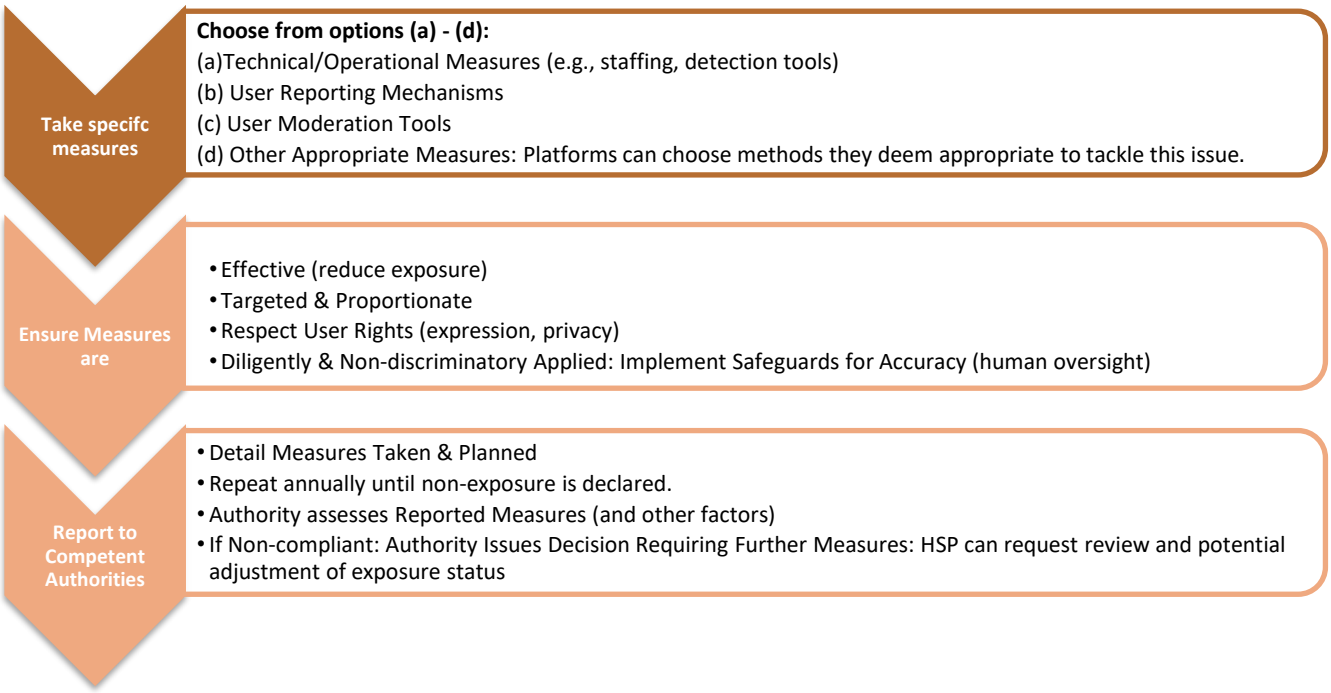
- It has received **two or more final removal orders** for such content within the past year (12 months).
- The relevant authority has notified them of this decision.

Once exposed to terrorist content, the HSP may decide upon a range of specific measures to identify and prevent the dissemination of terrorist content. The TCO Regulation does not define what these should be, but provides examples:

- appropriate technical and operational measures or capacities, such as **appropriate staffing or technical means to identify and promptly remove** or disable access to terrorist content;
- easily accessible and user-friendly **mechanisms for users to report or flag** to the hosting service provider alleged terrorist content;
- any other mechanisms to **increase the awareness of terrorist content** on its services, such as mechanisms for **user moderation**;
- any other measure that the hosting service provider considers to be appropriate to address the availability of terrorist content on its services.

The TCO requires HSPs that are exposed to terrorist content to update their ToS to include the specific measures taken by an HSP to prevent the upload and dissemination of terrorist content.

The TCO in its Art. 5(2) stipulates that "such measures may **include one or more** of the following", In this part, we will summarise options stated under the TCO to respond quickly and proportionately when your platform falls under category of Article 5:



These measures must:

- Be effective in minimizing exposure to terrorist content;
- Be targeted and proportionate, considering the platform's size and capabilities;
- Respect user rights, especially freedom of expression, privacy, and data protection;
- Be applied diligently and without discrimination; and
- Include human oversight and verification when using technical methods to avoid removing non-threatening content.

5.1. Content Moderation

Content moderation is the first line of defence against harmful or inappropriate content within online platforms. It refers to the procedures and organised practices for reviewing user-generated content posted or shared.

Moderation is a proactive and nuanced approach to managing online content, focusing on ensuring that user-generated material complies with established guidelines and local jurisdictions. It involves the removal or restriction of content that violates rules while allowing for diverse perspectives and free expression within acceptable boundaries. Moderation aims to maintain a safe and constructive digital environment. It involves preventing harmful content from spreading online and disabling it once it is shared publicly to reduce its reach. It can include a mix of automated and human processes, both of which constitute specific measures under the TCO, such as:

- **Pre-moderation:** Content is reviewed by a moderator before it is published. High level of control, good for high-risk platforms. However, slows down content publishing and expensive for large communities.
- **Post-moderation:** Content is published immediately and then reviewed by a moderator. This is fast, good for fast-paced communities but expensive for large communities, platform may be liable for harmful content.
- **Reactive moderation:** Community members flag harmful content for moderators to review. This is scalable and can reduce platform liability. Harmful content may remain visible for some time, can damage platform reputation.
- **Distributed moderation:** Moderation responsibility is shared among a group of people.

5.2. Automated Tools

Automated detection tools can be a valuable addition to detecting terrorist content and represent a standard specific measure under the TCO.

Automated online content moderation tools can be categorized based on the technology they use, the types of content they focus on, and their operational approaches. While the TCO does not mandate the use of automated detection tools, many HSPs use a combination of these alongside user reporting, Trusted Flagger programs, and industry collaboration efforts.

Technology Used:

- **Machine Learning Models**: These tools use trained algorithms to identify patterns in content that correlate with inappropriate material. Examples include deep learning models for detecting nuances in text, images, or videos.
- **Natural Language Processing (NLP):** Specifically tailored to analyse text, these tools can understand context, sentiment, and intent, which is crucial for detecting subtle forms of inappropriate content like hate speech or harassment.
- **Computer Vision:** Employed for moderating images and videos, these tools analyse visual content to identify nudity, violence, trademarks, and other visual policy violations.

- **Rule-based Systems:** These systems work on predefined rules or patterns. For example, they might automatically flag or remove content containing specific prohibited words or phrases.

Content Type:

- **Text Moderation Tools:** These analyse written content such as comments, posts, forums, and chats.
- **Image/Video Moderation Tools:** These focus on visual content, using image recognition technologies to detect inappropriate or sensitive visual elements.
- **Audio Moderation Tools:** These analyse spoken words in audio formats to detect issues like profanity or hate speech.

Detection Method:

- **Keyword Detection:** Uses specific banned or sensitive words as triggers for moderation actions.
- **Anomaly Detection:** Identifies outliers or abnormal patterns in content which may suggest inappropriate or unusual behaviour.
- **Sentiment Analysis:** Determines the sentiment or emotional tone behind a piece of content to help identify negative interactions like bullying or toxic behaviour.
- **Semantic Analysis:** Goes beyond keyword detection to understand the meaning and context, which is particularly useful in identifying sophisticated forms of misuse like disguised profanity or coded language.

Operational Approach:

- **Real-time Moderation:** Analyses and takes action on content as it is posted, aiming to prevent inappropriate content from ever appearing on the platform.
- **Batch Processing:** Analyses content in bulk, typically used for large datasets where real-time processing is not required or feasible.

These categories help in tailoring the moderation tools to the specific needs of different platforms, considering factors like the type of content, the volume of data, and the specific risks associated with the platform's context.

Despite the potential of automated tools, human review remains essential for contextual understanding of complex content that can be nuanced, satirical, or created to deliberately evade known content moderation tactics.

Learn more about content moderation by taking **FRISCO Training Module 1**.

For an overview of **industry technical tools and solutions** for content detection and moderation, please refer to the **FRISCO Brochure Tools and Approaches for small HPSs to address terrorist content online.**

5.3. Partnerships and industry collaboration

The FRISCO survey revealed a significant knowledge gap exists among HSPs regarding communication with LEAs (69.7% unaware). There's a lack of established mechanisms for smooth interaction.

To bridge this gap, we need a multi-pronged approach. Establishing partnerships and promoting cooperation among the various stakeholders are essential elements of the collective effort to combat terrorist content online and represent specific measures under the TCO. It is important for HSPs to invest in collaborative initiatives with large online platforms, governments, non-governmental organisations and international bodies. By working together, these entities can share valuable information, data, and expertise, allowing for a deeper understanding of extremist trends and countermeasures. For example, HSPs should:

- **Make use of tools and programmes** for ready made solutions:
    - Europol's PERCI [15]- a cloud-based single system allowing cooperation among competent authorities and Europol, allowing HSPs to receive removal orders and referrals in a unified and standardised manner, in a secure way, from a single channel.
    - Trusted flagger programmes: Collaborate with experts and NGOs to identify and address high-risk content effectively.
- **Engage with national and multinational authorities** for clarification on procedures and specific interpretations of content categories.
    - Learn from organizations and networks working collaboratively with wider government-led forums—such as the EU Internet Forum, the EU Internet Referral Unit at EUROPOL, the United Nations' Counter-Terrorism Executive Directorate and the Christchurch Call to Action in an effort to advance tech companies' efforts to self-regulate and increase proactive responses.
- **Public-Private Partnerships (PPPs):** Teaming up big tech companies with smaller HSPs and LEAs creates a powerful force against illegal content, and discuss shared challenges. Collaborate with:
    - **Other platforms and industry organizations** : Existing forums like Global Internet Forum to Counter Terrorism (GIFCT) can be a valuable resource for sharing best practices and tackling online extremism. And industry bodies like the Trust and Safety Professional Association (TSPA) bring together some of the best vendor solutions for these challenges.

- **Academic networks:** such as Voxpol and GNET are valuable sources of insight about emerging trends and terrorist use of the Internet.
- **Other relevant EU networks**: such as the Radicalisation Awareness Network which connects frontline practitioners across Europe with one another, and with academics and policymakers, to share knowledge on preventing and countering radicalisation and violent extremism.

5.4. Transparency Reporting

Under Article 7 of the TCO Regulation, hosting providers (HSPs) must publish yearly transparency reports detailing actions taken against terrorist content. This applies to both voluntary removals and those required by law.

Reports must be publicly available by March 1st of the following year.

These transparency reports help HSPs build trust by demonstrating commitment to good governance and content moderation. They also provide valuable information about content removal practices. A good Transparency Report should include:

Measures taken by HSP to identify and remove terrorist content

Clear data and statistics on how the platform is meeting its obligations under the TCO.

Include references to specific policies and procedures that demonstrate how the HSP adheres to regulations

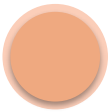
Mention Help centres, community guidelines websites, and blog posts that detail the specific provisions of policies

Provide clear explanations on how policies are enforced, addressing concerns about potential human rights risks

Measures taken by HSP to address the reappearance of terrorist content on your platform

This could include information on the number of removal orders received, the types of terrorist content identified, and the response times for takedown requests.

Acknowledge the ongoing challenges in combating terrorist content and outline specific steps being taken to improve detection and removal mechanisms.



The number of items of terrorist content HSP removed and did not remove following removal orders or specific measures

Information should be clear and available for users. number of content removals, user appeals, and compliance metrics.

The number and the outcome of complaints the HSP handled; judicial and administrative proceedings.

While providing data on content removal is crucial, ensure user privacy is protected by anonymizing the data whenever possible.

Source: authors based on Article 7 of the TCO regulation and industry measures.

Benchmarking: Check transparency reports from leading platforms like Roblox, and Google on addressing the dissemination of Terrorist Content Online and learn from their transparency reporting practices.

Accessibility: Make transparency reports easily accessible to the public on your website. Consider offering them in multiple languages for broader reach.

PRACTICES TO ENDORSE FOR EFFECTIVE TRANSPARENCY AND ACCOUNTIBILITY

In 2018, a group of human rights organisations, advocates, and academic experts developed and launched **the Santa Clara Principles on Transparency and Accountability in Content Moderation.** This is a set of three principles for how best to obtain meaningful transparency and accountability around Internet platforms' increasingly aggressive moderation of user-generated content. Those were endorsed by platforms such as META and Tiktok.

The principles have been further developed and reviewed in consultation with up to 50 experts and professionals. Here are the key principles applicable to HSPs, and guidance on their implementation:

Foundational Principles

- **Human Rights & Due Process:**
  - Transparent information on how human rights are considered in platform rules and enforcement.
  - Clear methods for obtaining support regarding content/account actions.
- **Understandable Rules & Policies:**
  - Easily accessible and detailed guidance on prohibited content (with examples).
  - Clarification on actions beyond removal (downranking) and their triggers. - Defined circumstances for account suspensions (temporary/permanent).
- **Cultural Competence:** Put rules, notices & appeals in user's language
  - Confidence in moderation decisions considering culture and context.
  - Company reports demonstrating language, regional, and cultural competence.
- **Integrity & Explicability:** Inspire confidence in content moderation systems through accuracy and non-discrimination
  - Understanding of automated decision-making and its impact.
  - Control over algorithmic curation and its influence on user experience.

Operational Principles

- **Numbers:** Requires companies to report data on content moderation actions, including appeals and involvement of state actors.

- **Notice:** Requires companies to provide clear and timely notice to users about content removal, account suspension, or other actions.
- **Appeal:** Requires companies to offer users a meaningful opportunity to appeal content moderation decisions through a clear and accessible process.

6. Conclusion

This Manual is designed to provide HSPs and Internet professionals with the knowledge and best practices needed to effectively combat terrorist content online. By implementing the recommendations presented here, HSPs can make a significant contribution to a safer online environment for users and society.

Combating terrorist content online goes beyond immediate security; it strengthens the very foundations of a democratic society. By preventing the spread of extremist ideologies and fostering a responsible online space, the internet sector plays a crucial role in upholding democratic values and promoting security.

A strong public-private partnership is essential for sustainable success. We encourage service providers to actively collaborate with other stakeholders, including industry peers, law enforcement agencies and civil society organisations. By working together, we can leverage the power of technology to create a safer and more inclusive digital space for all.

To learn more about the FRISCO project and the consortium leading it, please visit our website: Frisco (friscoproject.eu)