

A Guide to Transparency under the EU Digital Services Act

Foreword

The Tech Coalition envisions a digital world where children are free to play, learn, and explore without fear of harm. We have long understood that transparency is a crucial component for achieving this vision, making it one of the original pillars of Project Protect. Transparency underpins our mission to unite the global tech industry to protect children from online sexual exploitation and abuse.

In order to promote industry transparency, we developed and launched *Trust: Voluntary Framework for Industry Transparency* in 2022. By providing guidance to tech companies on transparency reporting, the Trust Framework aims to ensure that industry is proactively being accountable to the global community in honoring our commitment to combat online CSEA.

While the Trust Framework laid out a set of voluntary commitments for industry, the transparency landscape has shifted in recent years with regulation such as the European Union's Digital Services Act (DSA). With this in mind, we partnered with **Tremau**, a Trust & Safety solution provider and DSA expert, to create a guide explaining how the Trust Framework interacts with the DSA's transparency requirements with respect to CSEA and how to operationalize reporting in this new era. We believe that mandatory transparency reporting provides an opportunity for companies to go above and beyond in demonstrating their belief in the importance of transparency.

Table of contents

- 01 Background
- 02 The Digital Services Act
- 03 Pillar 1: How should services design transparent OCSEA policies?
- 04 Pillar 2: How should services conduct transparent moderation of OCSEA?
- 05 Pillar 3: How should services report OCSEA within their DSA transparency reports?
- 06 Annex

Executive Summary

- The Tech Coalition’s Guide to DSA Transparency is a guide for all things transparency under the European Union’s (EU) Digital Services Act (DSA). It operationalizes the DSA’s transparency requirements according to **three pillars**: (1) transparent **policies**, (2) transparent content **moderation** and (3) transparent **reporting**. After contextualizing the scope and objectives of the DSA, each section explores one pillar, allowing service providers of all sizes and categories to understand what the DSA requires of them.
- In line with the Tech Coalition’s mission to combat online child sexual exploitation and abuse (OCSEA), this guide focuses on **OCSEA**: transparent OCSEA policies, transparent moderation of OCSEA content, and transparent OCSEA reporting. Under the DSA, services have transparency obligations regarding all types of content and activity on their service. Still, the impact of OCSEA warrants a focus – especially for those services taking important measures to protect minors.
- The guidance for Pillar 3 (transparency reporting) was drafted taking into account the European Commission’s Delegated act on DSA transparency reporting,¹ which further develops the transparency obligations established by the DSA.
- Service providers familiar with the Tech Coalition’s **Trust Framework** for transparency reporting may find many of the DSA’s requirements familiar. This guide maps both frameworks. **Gray-colored boxes** provide a high-level overview of areas where the DSA overlaps with the Trust Framework and where the DSA requires additional elements. They also highlight areas where the Trust Framework goes beyond the DSA – outlining how they could be incorporated into DSA practices to go the extra mile. Where the Trust Framework does not directly address the relevant DSA duty, this is indicated with ‘N/A’.
- The different sections are **color coded** to allow service providers to identify what transparency duties apply to them. The legend is as follows:
 - All intermediary services.
 - Hosting services.
 - Online platforms.
 - Very large online platforms (VLOPs)/Very large search engines (VLOSEs).

¹ See [here](#).

01 Background

Transparency is a key feature of emerging online safety regulatory frameworks, including the European Union's (EU) Digital Services Act (DSA). Many Tech Coalition members will have experience in voluntary transparency reporting, however the DSA implements specific and obligatory transparency requirements. Broadly, these obligations aim to make providers accountable for diligent content moderation and risk management through public scrutiny.

This guide aims to explain the DSA transparency requirements. It showcases where members can apply aspects of the Trust Framework to meet the demands of the DSA, and where they will need to implement totally new practices. The Trust Framework remains a valuable source of alignment for transparency reporting across the industry. This guide supplements it by highlighting areas where the DSA overlaps with the Trust Framework, where additional requirements need to be met and where the Trust Framework recommends going further to demonstrate commitment to combating child sexual exploitation and abuse (OCSEA).

This guide is relevant to all of the Tech Coalition's members; from big to small, and for all types of online intermediaries. Not all requirements will apply to all services; transparency obligations within the DSA are tiered, with the most comprehensive requirements applying to services with the largest risk profile (i.e. very large online platforms and search engines). These tiers are highlighted below so that members can identify their applicable obligations. At the same time, this guide can serve as a roadmap for smaller online platforms who might want to assess what new duties will apply to them as they grow or if they begin to offer new services that change their provider category.

Three transparency pillars

This guide separates the DSA's transparency elements into three pillars:

- 1. Transparent policies:** refers to transparent policies that services must establish and make available to their users about content and behavior allowed on their service, through their terms and conditions (T&Cs).
- 2. Transparent content moderation:** relates to the information that services must provide to users affected by content moderation restrictions.
- 3. Transparent reporting:** concerns the public reporting of content moderation actions taken.

Under each pillar, the specific DSA requirements are explained, with a focus on OCSEA. Throughout the guide there are gray boxes which map the DSA's requirements with the Tech Coalition's Trust Framework. As introduced, in these gray boxes members will read:

- **Trust & DSA overlap:** which Trust Framework elements are applicable to the DSA.
- **Novel DSA requirements:** which additional elements are not addressed by the Trust Framework.
- **Beyond the DSA—Trust Framework elements to go the extra mile:** where the reporting elements from the Trust Framework go beyond the requirements of the DSA, providing an opportunity to demonstrating additional commitment to transparency.

Tech Coalition's Trust Framework

In 2022, the Tech Coalition launched '[Trust](#)', its voluntary framework for industry transparency. The Trust Framework provides flexible guidance to tech companies seeking to build trust and demonstrate accountability by providing transparency reporting concerning their efforts to combat OCSEA.

The framework provides suggestions and guidance for the three major sections of a report:

- Policies and Practices.
- Processes and Systems.
- Outcomes.

Each section includes a list of potential categories and/or metrics to include, with indications for which are recommended and which may be more aspirational.

Check it out here:

- [Trust: Voluntary Framework for Industry Transparency.](#)
- [Trust: Transparency Reporting Implementation Guide.](#)
- [Transparency Reporting Template](#)

02 The Digital Services Act

Overview

The Digital Services Act has applied to all online intermediary services since February 17, 2024. It applies to online service providers that offer their services within the European Union and act as intermediaries in connecting users with goods, services, and content ('intermediary services'.) This includes internet service providers, file-hosting services, social media platforms and others.

The DSA aims to create a safer online environment, protect fundamental rights, and ensure transparency and accountability. Often labeled a 'process-based' regulation, it primarily regulates the processes that underpin content moderation, rather than the removal of specific content. It requires providers to adhere to due process requirements in moderating, and imposes systemic risk management obligations on the largest platforms and search engines.

Tiered and cumulative obligations

The DSA applies to a broad range of online services, covering all online intermediaries involved in providing access to, hosting, transmitting, and indexing content created by others. **Within all intermediary services, the DSA sets out sub-categories of services**, depending on their activity and (sometimes) size. The DSA's obligations are tiered and cumulative. The best way to visualize this is through an inverted pyramid: all intermediaries have the same baseline duties, but obligations get added for services higher up on the pyramid.

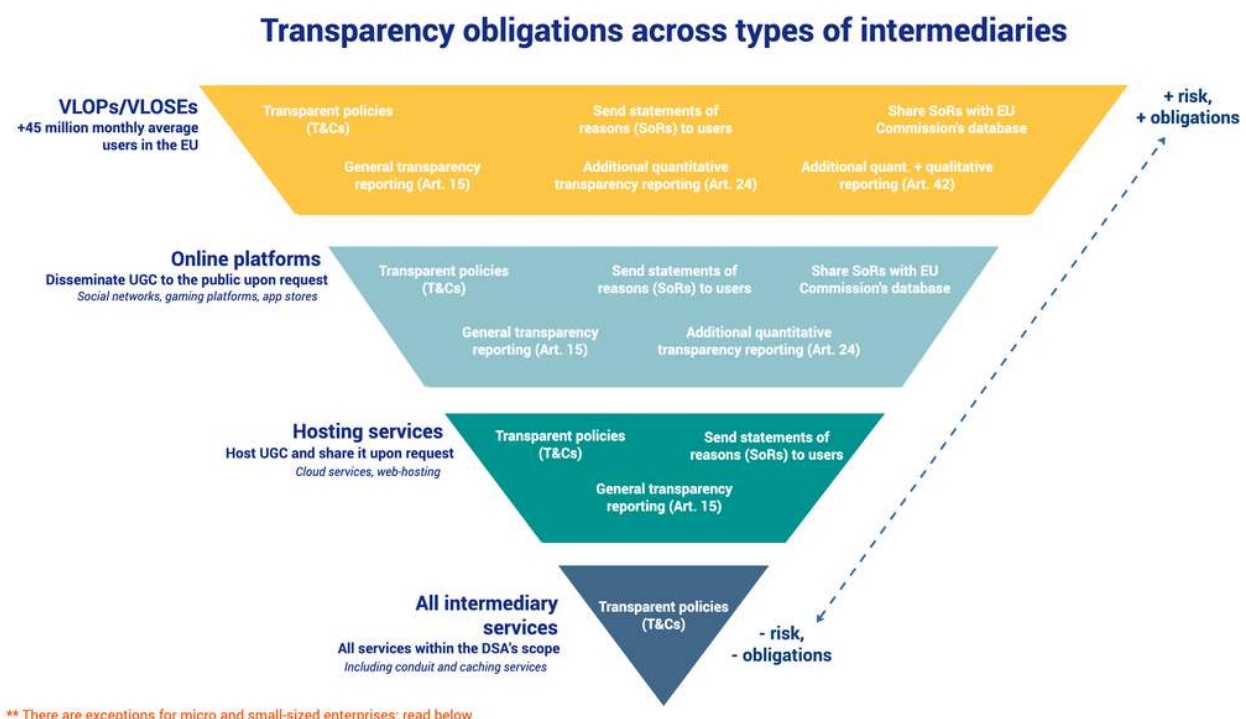


Figure 1: Transparency obligations across types of online intermediaries.

Throughout this Guide, sections have been color coded to allow services to skim through the obligations most relevant to them, depending on their subcategory.

■ **All intermediary services:** this refers to all services within the scope of the DSA. It covers both services at the 'top' of the pyramid (e.g. the largest social platforms) and **mere conduit** (including internet exchange points, wireless access points, DNS, VPN), **caching** (content delivery networks, reverse proxies) and hosting services (including online platforms).

■ **Hosting services:** services that host user-generated content (UGC) and share it upon the user's request. This includes cloud and web hosting services. Hosted UGC can take many forms: photos, videos, messages, comments, code...

■ **Online platforms:** a subset of hosting services who, in addition to hosting UGC, disseminate it to the public at the user's request. Includes social media, gaming platforms, dating apps, marketplaces, app stores, etc. Note however that platforms who qualify as small or micro enterprises under EU law benefit from exemptions (read ahead or see [Table 1](#)).

■ **Very large online platforms (VLOPs)/Very large online search engines (VLOSEs):** online platforms and search engines with over 45 million monthly average recipients in the EU.

→ **Micro and small-sized services:** the DSA reduces the burden for intermediaries who qualify as micro or small enterprises under EU law (below **10 million euros** in their turnover or balance sheet **AND** less than **50 employees**, see [here](#)). As Table 1 explains, micro and small enterprises are exempt from transparency reporting duties (although they are still bound by other transparency pillars).

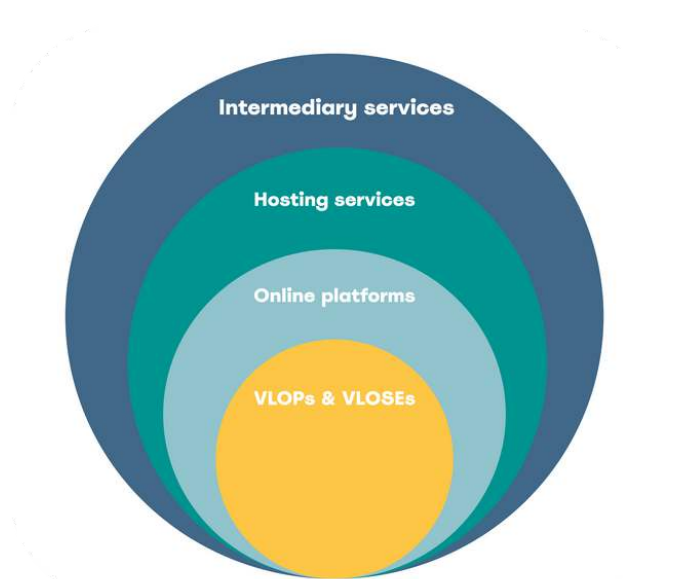


Figure 2: Types of services under the Digital Services Act

Transparency pillars	Obligation	All intermediary services	Hosting services	Online platforms	VLOP/VLOSEs	Micro & small-sized services	
						Micro or small-sized services that are <i>conduit or caching</i> services	Micro or small-sized services who are <i>hosting services or online platforms</i> ¹
1. Policies	Article 14 (T&Cs)	✓	✓	✓	✓	✓	✓
2. Content moderation	Article 17 (send statements of reasons, or SoRs)		✓	✓	✓		✓
	Article 24 (send SoRs to the Commission's SoR database)			✓	✓		
3. Reporting	Article 15 (Quantitative & qualitative reporting)	✓	✓	✓	✓		
	Article 24 (Additional quantitative reporting)			✓	✓		
	Article 42 (additional quantitative and qualitative reporting)				✓		

¹ Micro or small online platforms who are also VLOPs (i.e. they have over 45 million monthly average EU users) would still have to meet VLOP obligations.

03 Pillar 1: How should services design transparent OCSEA policies?

1.1 Why transparent policies?

The DSA requires services to have transparent policies—including with respect to OCSEA—through their terms and conditions (T&Cs). This obligation aims to promote consistent and foreseeable content moderation practices. T&Cs cover any clauses, whatever their name or form, that govern the user-service relationship. For example, they may include the Terms of Service, content policies and community guidelines.

1.2 What to explain?

The DSA sets a high benchmark for what needs to be disclosed through T&Cs. Through reading the T&Cs, a user should be able to answer at least all of the following questions:

- **What behavior is and isn't allowed on this service?** Users should have a comprehensive view of what they can and cannot do. Typically, this is done through community guidelines or content policies. Best practice is to include examples of behavior that is or is not allowed.
- **What sanctions can be imposed against my content or accounts?** The DSA requires informing users of 'any restrictions' possible. For example, indicate if you may remove content; restrict its visibility; suspend an account; ban it; etc.
- **Does this service use automated moderation tools?** Services must indicate if they use automated tools when moderating, such as tools that proactively scan or auto-block content. The DSA doesn't specify a level of detail (e.g. "must I publicly specify the name of the tool I use?"). Providers have some discretion in balancing transparency with protecting the integrity of their content moderation systems.
- **How can I appeal moderation decisions against me?** Services must indicate what are the available redress mechanisms. If they have an internal complaints mechanism, services should inform users about it and about the rules that govern it: what decisions are open to appeal? For how long? How and where can users access it?
- It is also recommended to include in your T&Cs information on **how users can report content that they find harmful or illegal.**

Applying the Trust Framework

Trust Framework & DSA Overlap

Although the Trust Framework focuses on periodic transparency reporting, many of its reporting elements could be repurposed for inclusion in T&Cs to help meet DSA requirements. Members could include the following elements in their T&Cs:

From the 'Policies and Practices' section:

- Violative content: description of the service's policies with respect to OCSEA, including information on what constitutes **violative content and conduct**.
- Sanctions: description of the **consequences of breaching these policies**.
- **Appeals**: description of the policy and process for appeals related to OCSEA.

From the 'Processes and Systems' section:

- **OCSEA prevention and detection** (including a description of the technologies used to detect OCSEA): this could be repurposed for the DSA when reporting about the use of automated means in content moderation.
- **OCSEA moderation and enforcement** (a description of the moderation options that may be taken in response to a breach): the DSA requires T&Cs to describe sanctions and enforcement applied through moderation by the service.

Novel DSA requirements (not addressed by the Trust Framework)

These Trust Framework reporting elements broadly cover the types of information that intermediary services must include in their T&Cs under the DSA, with respect to OCSEA. However, it will be important for providers to meet the standard and format of detail required by the DSA (read subsection 1.3 below).

Of course, providers' T&Cs must be developed in relation to all types of activity and content that can be present on the platform, not only OCSEA.

Beyond the DSA—Trust Framework elements to go the extra mile

Some reporting elements from the Trust Framework go beyond the strict requirements of the DSA. They provide a pathway for developing T&Cs that demonstrate additional commitment to transparent OCSEA policies. Some reporting elements from the Trust Framework that could be included in T&Cs for these purposes include:

- The company's processes for responding to law enforcement requests.
- The company's membership in relevant industry organizations, child safety partnerships or any other relevant cross-industry collaboration.

1.3 How to explain?

The format of T&Cs must also meet certain standards. The DSA's requirements include:

- **Public and electronically available T&Cs.**
- **Clear language:** understandable, clear, and user-friendly T&Cs. Some best practices for understandability include:
 - Highlighting prominent terms.
 - Providing information in bite-sized pieces. For instance, by adding a summary ('TLDR') box at the top of each T&C section.
 - Making the terms searchable.

- **Minor-friendly language:** any service accessible to minors should as a best practice consider writing their T&Cs in a way understandable by them. Under the DSA, minor-friendly T&Cs are a duty for any service “directed or predominantly used” by minors. So far, platforms have made their own self-assessments of this point, but the European Commission may release guidance in the future.

■ VLOPs and VLOSEs have some **additional obligations** in respect of how they provide and communicate T&Cs. They must:

- Provide a concise, easily accessible and machine-readable **summary of T&Cs**
- Publish in the **official languages** of all EU member states in which they offer services.

Other resources

There are several resources available to guide services in building more child-friendly online services. Adapting the language on your T&Cs and setting strong child safety policies (see subsection below) are a key aspect of this. These resources include:

- IEEE ([2021](#)) Standard for an Age Appropriate Digital Services Framework.
- 5Rights ([2024](#)) Best practices baseline for implementing the DSA for children.
- UK ICO ([2020](#)) Age Appropriate Design Code

Applying the Trust Framework

Trust Framework & DSA Overlap

Under the Trust Framework, members are already advised to report on their overall approach to OCSEA “in language that is easy to understand.” Practices that have been developed under this guidance may help providers to meet the clarity requirements of the DSA.

Additionally, the Trust Framework advises members to provide a “description of any policies or any other relevant information that are available specifically for children or young people”. This may help members to develop T&Cs in minor-friendly language for DSA purposes.

Novel DSA requirements

The Trust Framework is focused on transparency reporting but, as this section describes, the DSA also requires transparency through T&Cs. T&Cs need to be displayed and continuously kept up-to-date outside reporting cycles on the relevant webpage or interface.

While the Trust Framework acknowledges in a general sense that companies’ transparency efforts will vary depending on their size and resources, the DSA imposes specific additional T&Cs formatting obligations on VLOPs and VLOSEs: providing summaries and T&Cs in all relevant MS languages.

Beyond the DSA—Trust Framework elements to go the extra mile N/A

1.4. Notifying T&C changes

A final requirement is to inform users of **any significant changes** to T&Cs. Complying with this duty requires setting up at least two processes:

- Process for defining **when** a significant change has occurred: this should be formalized so that changes cannot be implemented without notifying users.
- Process for informing the **user**: a pop-up notification or an email may suffice.

Applying the Trust Framework

Trust Framework & DSA Overlap N/A

Novel DSA requirements

The Trust Framework recommends disclosing updates to relevant policies that happened during the relevant reporting period. For DSA compliance, significant changes to content moderation policies (i.e. T&Cs) must also be reported within the qualitative template (see Pillar 3, 'Reporting' and the Annex to the guide). However, in addition to reporting on changes, the DSA requires reflecting changes through updates to the public T&Cs and directly notifying users of any significant changes.

Beyond the DSA—Trust Framework elements to go the extra mile N/A

1.5 Child safety policies: How must your policies protect children?

Online platforms have extra duties to protect minors and ensure “a high level of privacy, safety, and security” for them (Article 28). Their T&Cs and policies should articulate measures to comply with this obligation.

At minimum this includes prohibiting ad profiling to minors. Beyond this, Article 28 offers limited detail on measures expected. Nonetheless, online platforms can anticipate further guidance from the Commission in the near future. In the meantime, there are other best practices that services can consider:

- Automatically set minors' accounts to the highest safety settings.
- Minimize data collection of minor users.
- Set limits on the types of content and accounts that are recommended to minors.
- Adapt reporting mechanisms so they are minor-friendly.
- Insert age-appropriate interstitials.

Applying the Trust Framework

Trust Framework & DSA Overlap

Under 'Policies and Practices', the Trust Framework recommends that providers report on “any policies... available specifically for children or young people.” This information may be useful for detailing existing child safety policies within T&Cs for DSA purposes.

Novel DSA requirements

The Trust Framework is focused on periodic reporting of existing child-safety policies while giving services flexibility in defining their content. By contrast, the DSA requires online platforms to adopt measures to protect minors (Article 28), and these measures should be detailed in their T&Cs.

Beyond the DSA—Trust Framework elements to go the extra mile N/A

04 Pillar 2: How should services conduct transparent moderation of OCSEA?

2.1 Why transparent content moderation?

Building on the first pillar of transparent policies, the DSA's second transparency pillar aims to make service providers accountable to individual users when carrying out content moderation. When a provider imposes restrictions on a user's account or content, those restrictions must be identified and explained to users through, in DSA language, "**statements of reasons**" (SoRs). Benefiting from this information, users are empowered to submit appeals and hold providers accountable for fair content moderation.

■ 2.2 What are statements of reasons (SoRs)?

■ SoRs are notifications that hosting services, including online platforms, must send to users who have been affected by a restriction on their content, account, or use of the service. These notifications are often sent in the form of an email or an on-platform message. They must:

- **Identify** the restriction/s imposed;
- **Explain** why they have been imposed, with reference to the law or terms of service;
- **Disclose** whether **automated means** were used; and
- Outline possibilities for **redress**. For example, if your service has an internal appeals mechanism (which is mandatory for platforms), your SoRs must indicate it.

■ 2.3 When do I need to send SoRs to affected users?

■ Under the DSA, essentially all content moderation restrictions imposed under terms of service or due to illegality require an SoR, including:

- **Visibility restrictions**— such as content removals, disabling access or demoting.
- **Monetary restrictions**— any suspensions or terminations of the functionality.
- **Service restrictions**— any suspension or termination of the provision of the service.
- **Account restrictions**— any suspension or termination of a user's account.

The provider will need to explain the terms of service or legal grounds within the SoR. As for timing, SoRs should be sent to affected users when the restriction is imposed. A single SoR can cover several restrictions imposed against the same user simultaneously.

2.4. When should I not send SoRs?

There are three circumstances in which service providers do not need to send SoRs to users. These are:

1. **Orders from authorities:** Articles 9 and 10 describe a different type of notification that needs to be sent to users affected by a removal order or information request.
2. **Deceptive high-volume commercial content:** providers do not need to send SoRs following actions taken against advertising spam.
3. If a provider does not have the **electronic contact details** of an affected user.

2.5. How should I explain restrictions to users?

The DSA requires hosting services, including online platforms, to explain two elements in their SoRs: (1) **the facts and circumstances** relied on in taking the decision (ex. action taken following a user report; or following own-initiative content moderation), and (2) the **grounds** for the decision (whether the content is allegedly illegal or incompatible with T&C). Providers have scope to determine how to meet these requirements – expectations for specificity are tempered by the generally automated and at-scale nature of content moderation.

2.6. When should providers send SoRs to the Commission database?

To allow public scrutiny of content moderation decisions, platforms must also send their SoRs without “undue delay” to the European Commission for inclusion in a public database. Unlike SoRs sent to users, these SoRs must be **anonymized** and sent in a **standardized** format.²

Applying the Trust Framework

Trust Framework & DSA overlap N/A

Novel DSA requirements

The Trust Framework does not engage directly with how to explain OCSEA-related restrictions to affected users, although it acknowledges that secrecy may be valued by providers to prevent OCSEA perpetrators from subverting platform controls. However, the DSA does not allow for any exemptions from SoRs based on content type. The DSA’s overriding objective appears to be to enforce transparency and accountability for content moderation through due process.

Beyond the DSA—Trust Framework elements to go the extra mile N/A

² The format is available through an API endpoint [here](#). Due to the technical limitations of standing standardized and anonymised SoRs on an immediate basis, it is acceptable to submit bulk collections of SoRs at semi-frequent intervals.

05 Pillar 3: How should services report OCSEA within their DSA transparency reports?

■ 3.1 Why transparency reporting?

- Under its third transparency pillar, the DSA imposes a number of transparency reporting requirements that allow scrutiny of providers' content moderation and risk management activities. This is achieved in two parts. First, through periodic reporting across quantitative and qualitative criteria on **content moderation activities affecting content and accounts in the EU**. Second, through the release of risk assessments, mitigation measures and independent audit reports.

The European Commission has released a Delegated act specifying when, how, and what to report.³

3.2 When to report?

■ ■ ■ All services (with the exception of VLOPs & VLOSEs) must report once a year. Each report must cover the reporting period between January 1 - December 31. Publication must occur, at the latest, 2 months from the end of the reporting period (i.e. by end-of-February of the following year).

■ VLOPs and VLOSEs follow different rules. They must report twice a year, with 6-month reporting periods (January 1 - June 30; July 1 - December 31) and publication is due within 2 months after the end of each reporting period. The DSA began applying to these service types before others, so VLOPs/VLOSEs have already published several reports. VLOPs will need to draft their transparency reports in-line with the format stipulated in the Delegated act from 1 July 2025 onwards.

■ 3.3 How to report?

■ a. Mandatory templates

- The DSA requires both qualitative and quantitative reporting in a machine-readable format. The Delegated act contains templates in comma separated values (CSVs) which services must follow when reporting. Below are the templates' subject areas. In line with the DSA's tiered logic, not all services need to report on every template; it depends on their tier. For a more detailed breakdown of the metrics and CSVs that each service type must report on, see the Annex.

- **Quantitative templates:**

- Report identification (identity of the service and reporting period). ■ ■ ■ ■
- Member state orders (i.e. removal and information orders from EU states). ■ ■ ■ ■
- Notice and action mechanisms (reports of illegal content from users and trusted flaggers). ■ ■ ■
- Complaint mechanism (appeal mechanism). ■ ■
- Use of automated tools in content moderation (reporting on totally and partially automated decisions). ■ ■ ■ ■

³ In EU law, Delegated acts are a type of legal tool that empowers the European Commission to specify how to implement a higher-level legal obligation. In this case, the Delegated act will further define the transparency reporting obligation created by the DSA. See the Delegated act [here](#).

- Human resources (number of moderators, both internal (i.e. employed) and external (i.e. contracted)). ■
- **Qualitative template:** A CSV with space for services to provide qualitative descriptions to help explain and contextualize the quantitative metrics. ■ ■ ■ ■

CSVs may be machine-readable but they aren't always the most user friendly. Hence, a suggested approach to DSA transparency reports could entail:

1. Reporting using the mandatory templates.
2. In addition, releasing a separate file in a more user-friendly design and language that summarizes the content of the templates. This file would typically include qualitative descriptions and tables reflecting quantitative insights.

According to the Delegated act, the templates are mandatory for all services to follow. The final act will likely establish a transitional period, where services are allowed to report differently, giving them time to build the appropriate data collection pipelines.

b. Content categories

The Delegated act sets out a list of 15 primary content categories that services must use for reporting. For example, when reporting on removal orders from EU states, in addition to indicating the total number, services must break down the orders received by category ([See Table 2](#)).

Reporting period	Category of illegal content	Scope	Nr. removal orders received	...
YYYY-MM-DD/ YYYY-MM-DD	TOTAL	TOTAL	<i>Of all categories, from all EU states</i>	...
YYYY-MM-DD/ YYYY-MM-DD	STATEMENT_CATEGORY_ANIMAL_WELFARE	TOTAL	<i>Of this category, from all EU states</i>	...
...
YYYY-MM-DD/ YYYY-MM-DD	TOTAL	AT	<i>Of all categories, from Austrian authorities</i>	...
YYYY-MM-DD/ YYYY-MM-DD	STATEMENT_CATEGORY_ANIMAL_WELFARE	AT	<i>Of this category, from Austrian authorities</i>	...
...

Table 2: Excerpt of the quantitative template for member state orders, illustrating the different breakdowns required, including (1) by content category (2) by member state.

These **primary categories are exhaustive and mandatory**: services cannot create extra ones. The logic behind this is to support comparability: if services explain their moderation activities in similar terms, stakeholders can better understand and compare them.

Within each category, the Delegated act also sets out subcategories. For example, within category 4 “Illegal or harmful speech”, it distinguishes between 4a “Defamation”, 4b “Discrimination” and 4c “Hate speech”. In contrast to the mandatory primary categories, these subcategories are flexible. Services can create **additional subcategories** if those provided don't properly capture the moderation action. Operationally, to comply, services will have to map their moderation activities to these content categories.

Categories Mandatory use when reporting.	Subcategories (Examples) These subcategories are suggested by the DSA but services can create their own. For the full list of subcategories, refer to the Delegated act .
1. Animal welfare	<ul style="list-style-type: none"> • 1a. Animal harm. • 1b. Unlawful sale of animals.
2. Consumer information infringements. <i>Particularly relevant for online marketplaces.</i>	<ul style="list-style-type: none"> • 2a. Hidden advertisement or commercial communication, including by influencers. • 2b. Insufficient information on traders. • 2c. Misleading information about the characteristics of the goods and services. • 2d. Misleading information about the consumer's rights.
3. Cyber violence	<ul style="list-style-type: none"> • 3a. Cyber bullying and intimidation. • 3b. Cyber harassment. • 3c. Cyber incitement to hatred or violence. • 3d. Cyberstalking.
4. Cyber violence against women	<ul style="list-style-type: none"> • 4a. Cyber bullying and intimidation against girls. • 4b. Cyber harassment against women. • 4c. Cyber staking against women. • 4d. Gendered disinformation.
5. Data protection and privacy violations	<ul style="list-style-type: none"> • 5a. Biometric data breach. • 5b. Data falsification. • 5c. Missing processing ground for data. • 5d. Right to be forgotten.
6. Illegal or harmful speech	<ul style="list-style-type: none"> • 6a. Defamation. • 6b. Discrimination. • 6c. Illegal incitement to violence and hatred based on protected characteristics (hate speech).
7. Intellectual property infringements	<ul style="list-style-type: none"> • 7a. Copyright information. • 7b. Design infringements. • 7c. Geographical indications of infringements. • 7d. Patent infringements. • 7e. Trade secret infringements. • 7f. Trademark infringements.
8. Negative effects on civic discourse or elections	<ul style="list-style-type: none"> • 8a. Misinformation, disinformation, foreign information manipulation and interference. • 8b. Violation of EU law relevant to civic discourse or elections. • 8c. violation of national law relevant to civic discourse or elections.
9. Protection of minors <i>Key category for reporting OCSEA content moderation</i>	<ul style="list-style-type: none"> • 9a. Age-specific restrictions concerning minors. • 9b. Child sexual abuse material. • 9c. Child sexual abuse material containing deepfake or similar technology. • 9d. Grooming/sexual enticement of minors. • 9d. Unsafe challenges.

10. Risk for public security	<ul style="list-style-type: none"> • 10a. Illegal organizations. • 10b. Risk for environmental damage • 10c. Risk for public health. • 10d. Terrorist content.
11. Scams and/or fraud	<ul style="list-style-type: none"> • 11a. impersonation or account hacking. • 11b. Inauthentic accounts. • 11c. Inauthentic listings. • 11c. Inauthentic user reviews. • 11e. Phishing. • 11f. Pyramid schemes.
12. Self-harm	<ul style="list-style-type: none"> • 12a. Content promoting eating disorders. • 12b. Self-mutilation. • 12c. Suicide.
13. Unsafe, non-compliant or prohibited products <i>Most relevant for online marketplaces.</i>	<ul style="list-style-type: none"> • 13a. Prohibited or restricted products. • 13b. Unsafe or non-compliant products.
14. Violence	<ul style="list-style-type: none"> • 14a. Coordinated harm. • 14b. General calls or incitement to violence and/or hatred. • 14c. Human exploitation. • 14d. Human trafficking. • 14e. Trafficking in women and girls.
15. Other violation of provider's terms and conditions <i>Can be a catch-all for moderation that does not fit in the previous categories.</i>	<ul style="list-style-type: none"> • 15a. Adult sexual material. • 15b. Age-specific restraints. • 15c. Geographical requirements. • 15d. Goods/services not permitted to be offered on the platform. • 15e. Language requirements. • 15f. Nudity.

Table 3: High-level content categories (left) and examples of subcategories provided by the DSA. Categories are mandatory and exhaustive while services may create bespoke subcategories.

3.4 What to report?

The DSA's tiered approach extends to reporting. Hence, what to report depends on the service category: regular-sized online platforms have less to report than VLOPs and VLOSEs, but more than mere caching services, etc. The **Annex details which specific reporting requirements apply** to each service type. Reporting must be broken down by **calendar month**.

3.5 Reporting OCSEA under the DSA transparency framework

If platforms detect and moderate alleged OCSEA, then this must be reported through the DSA's reporting templates. Where it is reported will depend on how it was detected, either through a member state order, an illegal content report, or own-initiative investigation (e.g. automated tools). In each case, OCSEA must be reported under the relevant content category, differentiating it from other types of illegal and incompatible content.

a. Sheet 3: Member state orders

If an authority sends a removal order or an information request linked to OCSEA, the service who receives it would report on it on this sheet. Services must report on:

1. The total number of orders/requests received from all EU member states, broken down by content category; and
2. For each EU member state, how many orders/requests they received, also broken down by content category.
3. In both cases, services must report on how many of the received orders were actioned, median times to respond, number of items moderated, etc. See the Annex for more details.

b. Sheet 4: Article 16 notices

A cornerstone duty for hosting services and online platforms is creating a reporting tool for users to report content as being illegal. Increasingly, platforms are creating a specific form (separate to their reporting tools for T&C violations) to meet the DSA's requirements for this mechanism.

Any user notices of illegal content submitted during the relevant period must be reported on and broken down by content category. Services must indicate the number of notices received, the median time to take action, the items moderated, and the number of reports that in the end were actioned based on illegality or T&C incompatibility.⁴

c. Sheet 5: Own-initiative moderation

Any moderation conducted of your own initiative (ex. automated filtering, pre-moderation, etc.) where, as a result, you action content or accounts for sharing OCSEA would be reported here.

d. Content categories

When reporting moderation actions in all of the sheets listed above, all services in the DSA's scope must break the reported data by content category. The most relevant categories linked to OCSEA content may include:

- Category 9 (Protection of minors): while services are able to create new subcategories, the subcategories provided by the Delegated act are:
 - age-specific restrictions concerning minors.
 - child sexual abuse material.
 - grooming/sexual enticement of minors.
 - unsafe challenges.
- Category 15 (T&Cs): if your policies lead services to restrict other types of content that don't fit in any of the first 14 categories (see Table 3 above), they can report on these moderation actions here. For example, services that prohibit nudity or lewd content.

⁴ Implied here is that, even though the service must provide users with an option to report content as illegal, if the reported content is also incompatible with the service's T&Cs, the service can action said content on the basis of its policies.

Applying the Trust Framework

Trust Framework & DSA overlap

Several elements overlap.

- 1. From the 'Processes and Systems' section:** Qualitative description of detection and enforcement processes.
- 2. From 'Outcomes':** OCSEA identified and actioned, user accounts identified and actioned.

3. From 'Additional Outcome Metrics':

- **'Discovery':** total volume broken down by flagging method
 - User reports: note that DSA transparency reporting requires separated reporting of Article 16 illegal content notices received and actioned, whereas regular ToS user reports actioned must be reported under own-initiative CM.
 - EU government or law enforcement reporting (reporting on member state removal and information orders).
 - Proactive tools or other technology (useful for the DSA's own-initiative moderation CSV).
- Some elements of **'action'**:
 - Volume or percentage of OCSEA action broken down by action: remove, disabling, de-indexing (demotion), etc.
 - Volume or percentage broken down by policy violation as applicable (ex. OCSEA and grooming): for DSA reports this would need to be adapted to the DSA's categories.
 - Number of accounts actioned, broken down by type (warning, temporary suspension, closure, etc.): particularly when they follow own-initiative moderation.
 - Timeframes: since the DSA inquires about median times.
- Some elements of **'law enforcement requests'**: member state orders received and number of orders complied with.
 - Divided by type – except, for DSA reporting, instead of distinguishing between subpoenas and search warrants, services would need to distinguish removal and information orders.
- Some elements on **appeals**:
 - Total number of user appeals
 - Appeal success rate: for the DSA services must indicate how many appeals were overturned.
 - Appeals consequences: for the DSA services must report if the restriction was maintained, reversed, or if new restrictions were imposed.
- Geographical **breakdown**: the Trust Framework asks to report data on a geography or regional basis where possible. In several points across DSA reporting, services will need to break down their data by EU member states.

Novel DSA requirements

The DSA mandates reporting on additional elements to those found on the Trust Framework, especially for VLOPs. The best way to pinpoint these is to read Annex, which lists the key metrics required by the DSA templates for each service type. Some of the additional metrics are:

- For all intermediaries, regarding law enforcement requests, the DSA requires breaking down median times to inform the authority that you received the order; as well as median time to give the order effect.

- From hosting services onwards: own-initiative moderation details, including number of items moderated at your own initiative and number of items detected solely using automated means.
- For online platforms, including VLOPs: reporting on disputes submitted to out-of-court dispute settlement bodies and on suspensions taken against repeat offenders who violate misuse policies (either misuse of the reporting or the appeal mechanism).
- For VLOPs, reporting on human resources, breaking down moderators by their linguistic expertise in EU languages of countries where the platform does business.

Beyond the DSA—Trust Framework elements to go the extra mile

Some recommended reporting elements within the Trust Framework go beyond strict DSA requirements. While DSA reporting must be completed through mandatory templates, providers could report on additional metrics independent of the DSA to demonstrate their rigorous commitment to transparency.

From the Trust Framework, this could include:

- Discovery: Trust recommends that platforms provide a full breakdown of content detection, by flagging method. This is distinct from the DSA which only requires reporting on detection indirectly in relation to Article 16 notices and government orders received. Other flagging methods such as terms of service user reports are reported under own-initiative moderation, but only where they are actioned.
- Action: volume actioned by content format: images, videos, chats, livestreams, etc.
- Reporting: data on the statutory reports made and to which relevant authorities (ex. The National Center for Missing and Exploited Children or NCMEC).
- Law enforcement requests by type of information disclosed, whether content, non-content metadata, etc.
- Other insights: OCSEA insights, research findings, trend data, etc.

3.6 Risk assessment reports, mitigations, and independent audits

For VLOPs and VLOSEs, the DSA mandates transparency about their risk management activities at large: how are they evaluating their societal and economic impact? How are they mitigating any risks? Every year, VLOPs and VLOSEs must conduct a **systemic risk assessment** to evaluate risks stemming from the design, (mis)use and functioning of their services, and define mitigation measures. The DSA identifies **four broad systemic risks**:

1. Dissemination of illegal content.
2. Negative effects to fundamental rights, including freedom of expression, media freedom and pluralism, non-discrimination, and rights of the child.
3. Negative effects on civic discourse, electoral processes and public security.
4. Negative effects in relation to gender-based violence, public health, minors, as well as physical and mental wellbeing.

Mitigation measures must be tailored to the service's risk profile. They may include adapting its features, strengthening content moderation processes, or testing algorithmic systems. The full report must be submitted to the Commission for review. In addition, for transparency towards other stakeholders, VLOP/VLOSEs must release a **public version of the report**.

⁵ Under Article 16, hosting services, including online platforms, must put in place mechanisms that allow any individual or entity to notify them of information they believe to be illegal content

Further, these services undergo a **yearly independent audit** where an auditor evaluates, to the highest level of assurance possible in auditing, the service's commitments in the previous risk assessment and mitigation report. Essentially, did the service implement the controls declared? The **audit report** must also be released (in a non-sensitive version) to the public.

VLOPs and VLOSEs finalized their first risk assessment and mitigation reports in August 2023. These will be published along with the first DSA independent audit reports in October 2024 while the second iteration of risk assessment reports will be published in 2025.





Resource:







- European Commission, [Supervision of VLOPs and VLOSEs](#): lists all VLOPs and VLOSEs, their monthly average users and any ongoing enforcement decisions.





Annex

The following tables summarize the scope and key metrics that services must report on using the DSA's mandatory templates for quantitative and qualitative reporting. This table should allow services to understand what types of metrics they must report on depending on their service type. For a **full breakdown of required metrics**, please refer to the Delegated act's templates.












Quantitative templates

CSV nr. & name	Reporting area	Scope	Examples of metrics
1 - Report identification	<i>Identifies the reporter and the reporting period</i>	All services 	<ul style="list-style-type: none">Name of service provider.Publication date.Start and end dates of the reporting period.
2 - Content categories	<i>Lists the content categories that services must use to report</i>	All services 	This sheet reiterates the categories and subcategories established by the delegated act. If services create any additional subcategories when reporting, these should be logged here.
3 - Member state orders	<i>Removal and information orders received from EU member state authorities</i>	All services 	<ul style="list-style-type: none">Nr. orders received.Nr. orders complied with.Nr. items moderated.Median time to inform authority of receipt.Median time to give effect to the order.Number of specific items of information included in the total number of orders to act against illegal content.
4 - Notice and action mechanism	<i>Reports of illegal content by users and trusted flaggers</i>	Hosting services 	<ul style="list-style-type: none">Nr. notices received, categorized by the type of alleged illegal content.Nr. notices submitted by trusted flaggers.Nr. items moderated.Median time to take action.Nr. actions taken on the basis of the law.Nr. actions taken on the basis of T&Cs.

5 - Own-initiative moderation	<i>Content moderation activities at the service's own initiative</i>		All services 	<ul style="list-style-type: none"> Nr. items moderated at own-initiative. Nr. items detected solely using automated means. Total nr. restrictions taken. Nr. restrictions taken broken down by type: <ul style="list-style-type: none"> Removals. Demotions. Age restrictions. Labeling. Monetary restrictions including suspensions and terminations. Account suspensions or terminations. Provision of the service suspensions or terminations.
6 - Overall figures	Complaint mechanism	<i>Appeal mechanism for users</i>	Online platforms 	<ul style="list-style-type: none"> Total nr. complaints submitted. Decisions upheld. Decisions reversed. Decisions omitted.
	OOC bodies	<i>Disputes submitted to out-of-court dispute settlement bodies</i>	Online platforms 	<ul style="list-style-type: none"> Nr. disputes submitted. Decisions upheld. Decisions reversed. Decision omitted. Median time. Percentage of outcomes implemented.
	Suspensions against repeat offenders	<i>Actions under misuse policies</i>	Online platforms 	<ul style="list-style-type: none"> Nr. suspensions enacted for providing manifestly illegal content. Nr. suspensions for abusing the reporting system. Nr. suspensions for abusing the appeal system.
	Use of automated tools in content moderation	<i>Items processed using automation, accuracy & error rates of automated</i>	All services 	Total numbers: <ul style="list-style-type: none"> Nr. items solely/partly/not processed by automated means. Accuracy rates of items processed by automated means.
			Hosting services 	Use of automation linked to the notice & action mechanism (i.e. the illegal content reporting tool): <ul style="list-style-type: none"> Nr. items solely/partly/not processed by automated means.

		<i>means</i>		<ul style="list-style-type: none"> Accuracy rates of items processed by automated means.
			Online platforms 	Use of automation linked to the internal complaints mechanism and to own-initiative moderation: <ul style="list-style-type: none"> Nr. items solely/partly/not processed by automated means. Accuracy rates of items processed by automated means.
	Human resources	<i>Number of human moderators</i>	VLOPs and VLOSEs 	<ul style="list-style-type: none"> Nr. items solely/partly/not processed by automated means. Accuracy rates of items processed by automated means.
7 - Internal complaints	<i>Reports on the number of appeals submitted to the internal appeals mechanism and on the type of restriction that was appealed</i>		Online platforms 	<ul style="list-style-type: none"> Nr. visibility restrictions broken down by type: Removal, Disable, Demotion, Age restriction, Interaction restriction, Labeling, etc. Nr. monetary restrictions broken down by type: Suspension, Termination or Other. Nr. provisions of the service restrictions by type: Suspension or Termination. Nr. account restrictions by type: Suspension or Termination.
8 - By country and language	<i>Breakdown of moderators with sufficient linguistic expertise and of accuracy/error rates of automated tools by EU languages</i> <i>Number of monthly average users (MAUs) broken down by country (note that MAUs are used to determine VLOP/VLOSE status, see p. 8)</i>		VLOPs and VLOSEs 	<ul style="list-style-type: none"> Number of average monthly users during the reporting period, broken down by EU state. Nr. of internal moderators employed. Nr. of external moderators contracted. Nr. moderators with sufficient linguistic expertise, broken down by official EU languages. Nr. items processed using automation broken down by official EU languages. Accuracy rate of items processed and error rates, broken down by EU languages.

Qualitative template

CSV nr and name	Reporting area	Scope
9 - Statements	Summary of own-initiative moderation	All services 
	Meaningful and comprehensible information regarding the content moderation engaged in	All services 
	Qualitative description of the automated means	All services 
	Qualitative description of indicators of accuracy and possible rate of error of automated means	All services 
	Specification of the precise purposes to apply automated means	All services 
	Safeguards applied to the use of automated means	All services 
	High-level description of the content moderation governance structure	VLOPs and VLOSEs 
	Qualifications of the human resources dedicated to content moderation	VLOPs and VLOSEs 
	Training given to human resources dedicated to content moderation	VLOPs and VLOSEs 
	Support given to human resources dedicated to content moderation	VLOPs and VLOSEs 
	Methodology used to compute the number of human resources dedicated to content moderation	VLOPs and VLOSEs 



About Tech Coalition

The Tech Coalition facilitates the global tech industry's fight against the online sexual abuse and exploitation of children. We are an alliance of technology companies of varying sizes and sectors that work together to drive critical advances in technology and adoption of best practices for keeping children safe online. The Tech Coalition convenes and aligns the global tech industry, pooling their knowledge and expertise, to help all our members better prevent, detect, report, and remove online child sexual abuse content. This coalition represents a powerful core of expertise that is moving the tech industry towards a digital world where children are free to play, learn, and explore without fear of harm.

To learn more visit www.technologycoalition.org