



Project Partners;



digiQ



anthro polis



Co-funded by
the European Union



Handout for Youth-Work Organizations

How to Obtain Trusted Flagger Status for Reporting Hate Speech

Introduction: What Is Trusted Flagger Status and Why It Matters

Under the Digital Services Act (DSA), civil-society organizations can apply to become **Trusted Flaggers** — officially recognized entities whose reports of illegal content (including hate speech) must be given **priority treatment** by online platforms. Trusted Flaggers have enhanced reporting channels, faster review times, and a structured relationship with both platforms and the national **Digital Services Coordinator (DSC)**.

For youth-work organizations, especially those dealing with hate speech, cyberbullying, discrimination, online harassment, or the protection of vulnerable groups, Trusted Flagger status can significantly increase their impact. Many youth-serving NGOs already submit reports to social networks, but without special recognition, their reports move through slow or inconsistent review processes. As a Trusted Flagger, your organization's reports are treated with elevated credibility and urgency, which allows harmful content affecting young people to be removed more quickly and reliably.

Youth organizations also play a uniquely important role: they have access to real experiences of young users, understand evolving online language (slang, coded terms, memes), and are often among the first to observe new forms of hate speech or harassment. Trusted Flagger status helps formalize this expertise and integrate it into the broader online safety ecosystem.

This handout explains what the criteria are, how to prepare your organization, how to apply, and what responsibilities you should be ready to assume once recognized.

Understanding the Trusted Flagger Criteria

The Digital Services Act sets out clear criteria for becoming a Trusted Flagger. Youth organizations should assess whether they meet — or can build — the necessary capacity before applying.

1. Demonstrated Expertise in a Relevant Field

Your organization must show a sustained track record in areas related to illegal hate speech, harmful content, youth protection, online safety, or digital citizenship. This includes experience such as operating a reporting portal, running anti-hate-speech programs,



conducting monitoring or research, handling cases of online harassment, or providing support to marginalized groups affected by hate speech.

2. Independence and Political Neutrality

Trusted Flaggers must be independent from political influence and should demonstrate that their decision-making is guided by expertise, not political agendas. Youth organizations should have transparent ownership, governance structures, and no partisan alignment.

3. Strong Reporting and Verification Procedures

Your organization must have internal systems to verify that the reported content is illegal or harmful under EU or national hate-speech laws. This requires trained staff, a clear escalation mechanism, and the ability to document evidence (URLs, screenshots, context explanations, timestamps).

4. Capability to Process Reports Systematically

Trusted Flaggers must handle a significant number of reports professionally and consistently. You will need a sustainable workflow for receiving, assessing, and forwarding cases. This may include case management tools, internal guidelines, quality assurance processes, and staff trained in hate-speech classification.

5. Compliance with Data Protection and Privacy Rules

Because your organization will be processing sensitive content involving young people, you must demonstrate GDPR compliance. This includes data minimization, anonymization, secure storage, and clear retention policies.

6. Integrity and Accuracy of Reporting

Platforms and DSCs expect high accuracy. You must be able to distinguish illegal hate speech (as defined by law) from harmful but legal expressions. Clear internal criteria, legal consultation, and staff training will be essential.

If your organization can meet these criteria — or is prepared to develop them — you can proceed toward application.

Preparing Your Organization for the Application

The application process requires demonstrating capacity, expertise, and internal structures. Before applying, youth organizations should prepare the following elements:

1. Document Your Experience and Expertise



Prepare a dossier summarizing your work related to hate speech and youth protection. This may include:

- descriptions of relevant programs (anti-bullying, hate-speech prevention, digital citizenship education)
- annual reports
- statistics on cases handled
- previous cooperation with platforms or authorities
- research or publications on online harms

Clear documentation strengthens your credibility and shows the DSC that you understand the digital risk landscape.

2. Build or Update Reporting Protocols

A Trusted Flagger must use structured reporting mechanisms. Prepare or refine internal protocols that include:

- a verification process for each report
- categories of illegal content
- documentation standards (screenshots, metadata, context)
- quality control procedures
- a workflow for responding to urgent cases involving youth

This ensures consistency and demonstrates that your organization is ready to handle priority reporting channels responsibly.

3. Train Staff on Legal Frameworks

Staff members who will submit reports must understand:

- national hate speech laws
- relevant EU legislation
- distinctions between illegal hate speech and harmful, but legal, content
- contextual factors (protected characteristics, incitement, threats)



Training can be done internally, with external experts, or in partnership with digital-rights organizations.

4. Establish Data Protection Safeguards

Create or update:

- privacy policies
- data retention procedures
- secure storage solutions
- internal GDPR compliance guidelines
- staff confidentiality agreements

This is particularly important for organizations working with minors.

5. Build a Case Management Infrastructure

A simple spreadsheet may be enough for small organizations, but those aiming for Trusted Flagger status should consider:

- case management systems
- ticketing tools
- shared folders with controlled access
- secure incident-tracking platforms

This will help you manage reporting volume and maintain audit trails.

The Application Process: Working with the Digital Services Coordinator

When your organization is ready, you submit an application to your national **Digital Services Coordinator**. The process may differ slightly by country, but the core elements are the same across the EU.

Step 1: Submit a Formal Application

Your application will typically include:



- organizational information
- governance structure
- documented expertise in hate speech and youth protection
- description of reporting workflows
- data protection documentation
- staff qualifications and training
- examples of past reporting or monitoring

Some DSCs may also ask for references or evidence of cooperation with platforms or other regulators.

Step 2: Participate in Evaluation and Consultation

The DSC may request additional information, ask for clarifications, or conduct an interview. They may also consult other authorities or civil-society actors to confirm the organization's track record. Youth organizations should be ready to demonstrate:

- their legal understanding of hate-speech definitions
- their ability to analyze context
- their ability to avoid false reporting
- their independence and neutrality

Step 3: Receive Designation and Begin Cooperation

If approved, your organization will be officially listed as a **Trusted Flagger**. The DSC will notify social networks and provide you with:

- priority reporting channels
- platform-specific instructions
- communication points of contact
- expectations for reporting accuracy and case volume



Trusted Flaggers must then begin submitting reports through the designated channels, following the agreed rules.

Step 4: Ongoing Monitoring and Accountability

Trusted Flagger status is not permanent or unconditional. DSCs may periodically review your organization's performance, accuracy, and independence. This ensures the Trusted Flagger ecosystem remains credible, effective, and resistant to misuse.

Youth organizations that maintain consistent, accurate, and ethical reporting will retain status long-term and may receive additional opportunities for collaboration, such as participation in systemic risk assessments or policy consultations.

Operating as a Trusted Flagger: Responsibilities and Best Practices

Once granted Trusted Flagger status, youth organizations gain significant influence — but they also assume new responsibilities. Trusted Flaggers must operate at a high professional standard and engage constructively with platforms and regulators.

1. Maintain High-Quality, Evidence-Based Reports

Each report must be contextualized, precise, and backed by clear evidence. Poor-quality reports undermine the system. Trusted Flaggers should always document:

- the exact illegal expression
- the protected characteristic targeted
- the context in which the content appears
- any patterns or repeated behaviours
- potential risks to the young person

This ensures the platform can act quickly and confidently.

2. Prioritize Youth Safety and Wellbeing

When reporting hate speech involving minors, youth organizations must handle cases with care. They should minimize emotional burden on the young person, explain the reporting process, and coordinate with mental-health or school support when relevant. Trusted Flagger responsibilities never override child protection principles.

3. Engage with Platforms Constructively



Trusted Flaggers often communicate regularly with online platforms. This can include providing explanations of local context, highlighting emerging trends, or clarifying ambiguous cases. Constructive dialogue helps platforms improve content moderation systems.

4. Share Insights with the DSC and Contribute to Broader Policy Work

Trusted Flaggers play a role not only in reporting individual pieces of content but also in identifying systemic failures. Youth organizations should consider submitting:

- quarterly summaries of hate-speech trends
- youth-centered insights on coded language or new harassment patterns
- observations about platform responsiveness

These insights help DSCs assess risks and improve enforcement across the digital landscape.

5. Maintain Transparency, Independence, and Ethical Standards

Trusted Flagger status requires integrity. Youth organizations must avoid political bias, conflicts of interest, or practices that could damage trust. They should uphold clear ethical guidelines and remain impartial when evaluating content.

Conclusion

Becoming a Trusted Flagger is both a responsibility and an opportunity. For youth-work organizations, it provides a powerful mechanism to protect young people from online hate speech, violence, and harassment. The recognition strengthens your organization's influence, improves collaboration with platforms and authorities, and contributes to a safer digital environment for all young users.

With strong preparation, transparent governance, reliable expertise, and robust procedures, youth organizations can successfully obtain Trusted Flagger status and become central actors in the new online safety framework established by the Digital Services Act.