CHRISTCHURCH CALL

TO ELIMINATE TERRORIST & VIOLENT EXTREMIST CONTENT ONLINE

Community Guidance: Government Transparency and Reporting

Table of contents

Introduction 0
About this guidance 1
Scope 2
Why is government transparency important? 3
Key Pieces of Community Advice 5
CDT's framework for meaningful transparency 5
Santa Clara 2.0 5
Tech Against Terrorism's Guidelines for Government 6
Government Transparency Reporting 6
Principles 6
Categories and Content 8
Examples and Good Practices 10
1: The Internet Hotline Centre, Japan 10
2: Independent Review, ARCOM, France 10
3: EU TCO Framework for Transparency 11
4: Transparency Reporting: EU Internet Referral Unit, Europol 13
5: Transparency Reporting: Department of Internal Affairs, New Zealand 13
6: Transparency Reporting, Ofcom, UK 14

Introduction

It is a priority for the Christchurch Call Community that supporters – both governments and tech companies – provide meaningful transparency around how they are acting to eliminate terrorist and violent extremist content online while respecting and promoting human rights and a free, open and secure internet. This objective was clearly articulated and set in the Community work plan for transparency endorsed by Leaders in 2021.

To date, most efforts to improve transparency have focused on and been from within tech companies. This started in the early 2000s in response to growing civil society concerns about the relationships between these companies and governments.[1] The Berkman Klein Center for Internet and Society developed the earliest version of the Lumen database more than 20 years ago, in 2002. In 2010 Google published the first report on government requests it received. Interest in tech companies' disclosure of this sort of information only intensified through the mid-2010s, following the Snowden revelations.

Transparency reporting is now a well-established and reasonably widespread practice among the tech companies that support the Call. This reporting has developed beyond basic data on government legal requests to cover company content policy and enforcement, and information operations disclosures.[2] There is also expert guidance from academia and civil society[3] and from various regulatory and voluntary initiatives, including some specific to terrorist and violent extremist content, that aim to make tech companies' transparency reporting more widespread, consistent, and meaningful.[4]

By contrast, despite transparency being a well-established democratic principle, there has been less emphasis on government transparency in relation to content moderation.[5] Only a handful of the Call's government supporters regularly publish transparency reports.[6]

Yet government transparency is a subject of legitimate interest and concern to all Community members. It is not possible to understand the whole online content ecosystem without visibility of governments' influence in determining what stays online and can be easily accessed, and what gets removed or restricted, whether because it is unlawful or a breach of a company's terms of service.

---

[1] For more, see History of Transparency Reports – Trust & Safety Professional Association (tspa.org)

[2] Integrity Institute Shining a Light on Platform Transparency Best Practices July 2023

[3] Ibid. See also, e.g: Lai, S, Shiffman, N and Wanless, A Operational Reporting By Online Services: A Proposed Framework Carnegie Endowment for International Peace: May 2023; Spandana, S and Bankston, K The Transparency Reporting Toolkit: Content Takedown Reporting New America Open Technology Institute: October 2018; Llanso, E and Morgan, S Getting Specific About Transparency, Privacy, and Free Expression Online Center for Democracy and Technology / Global Network Initiative: November 2014.

[4] Examples include Tech Against Terrorism's mentoring to meet GIFCT membership criteria, the OECD voluntary transparency reporting framework and transparency reporting provisions in the EU Terrorist Content Online Regulation.

[5] See e.g. paragraphs 50-51 of the Office of the Privacy Commissioner's 2017 report, Transparency Reporting, which notes the difficulties of getting comprehensive reporting out of government agencies.

[6] Noted in Vogus, C and Llanso, E "Making Transparency Meaningful: A Framework for Policy Makers", Centre for Democracy and Technology, December 2021. This small group includes Europol's Internet Referral Unit and the Department of Internal Affairs in New Zealand. Australia's e-Safety Commissioner's annual report also includes some relevant metrics.

Government involvement in content moderation can give rise to significant human rights risks[7]. Those risks include the potential for circumventing the law and due process,[8] introducing bias into company policy- and decision-making,[9] using tech companies as proxies to silence legitimate grievances and dissent, and interfering with users' access to effective remedies. Disclosing information about how, why, and when governments exercise influence over digital content can help mitigate these risks by building trust, fostering accountability, and facilitating access to remedy for affected users.

All Call supporters have committed to work together with civil society to improve transparency and ensure that Call commitments are implemented in a manner consistent with a free, open, and secure internet, and international human rights law.

Call-supporting governments have also made relevant commitments elsewhere. For example, all have signed up to Sustainable Development Goal 16.[10] Most are members of the Open Government Partnership and have therefore committed to increase the availability and usefulness of information about government activities at all levels. Many are members of the Freedom Online Coalition, which has called on governments everywhere to promote transparency and independent, effective oversight around the use of content take-down notices, restrictions on online content or user access, and similar measures, while committing to do the same and to preserve and strengthen the multistakeholder model of internet governance.[11]

The focus of this guidance is on government transparency, and reporting in particular. That is not to suggest that there is no need for further improvement on the tech company side. The Integrity Institute judges that "to date, voluntary transparency efforts from the industry have not provided a satisfactory level of transparency". As key pieces of regulation come into effect, the Call Community needs to evaluate their impact on transparency and continue to lift standards and address gaps, including around the impacts of algorithms on user journeys,[12] and safety policies and tools for emerging technologies.

About this guidance

In 2023, the Call Community started a series of discussions on government transparency. The focus was on things governments could do to improve transparency in their efforts to detect terrorist and violent extremist content online, and associated legal requests and other referrals to tech companies to remove, restrict access, or otherwise take action on it and on associated accounts.

The objectives of the Community sessions were:

- To identify the key stakeholders for government transparency, what they need to know, and how best to deliver that information in a timely and accessible way;

---

[7] Office of the UN High Commissioner for Human Rights "Moderating online content: fighting harm or silencing dissent?" July 2021

[8] Pielemeier, J and Sheehy, C "Understanding the Human Rights Risks Associated with Internet Referral Units," Feb 2019

[9] See Oversight Board overturns Meta's decision in "UK drill music" case | Oversight Board

[10] SDG 16 is "Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable and inclusive institutions at all levels".

[11] FOC-recommendations-consensus.pdf (freedomonlinecoalition.com)

[12] Christchurch Call Initiative on Algorithmic Outcomes | Christchurch Call

- To socialise existing guidance and identify any gaps;
- To explore some current approaches to government transparency;
- To develop Community guidance to support more consistent and meaningful transparency by governments, with good practice examples.

Meaningful transparency is most likely to emerge in a framework or regime that encompasses different actors, types of information and data, modalities, and formats. Governments have a role in shaping the overall framework, through voluntary and cooperative initiatives like the Christchurch Call, through co-regulation and regulation. Governments, as actors in content moderation, also need to provide transparency around their own policies, processes and activities within the framework. That transparency can take various forms, depending on the audience and purpose.

This guidance begins by considering an overall framework for transparency, and the different forms that it can usefully include, before honing in on one particular form: reporting. It synthesises some principles for government transparency reporting as well as key categories and content to cover, as put forward in the nascent body of expert advice published by Community members and articulated in our Community sessions. The guidance also contains case studies that highlight some existing good practices among the Call's government supporters.

Scope

The key relevant Call commitments are set out in the table below.

Table 1: Call Commitments Relevant to Content Moderation and Transparency Reporting

Online service providers commit to: Set transparent terms of service, describing policies on terrorist and violent extremist content and consequences for users who share it. Take transparent, specific measures to prevent upload, detect and remove such content. Implement regular and transparent reporting on the quantity and nature of such content being detected and removed.	Governments commit to: Ensure the effective enforcement of applicable laws that prohibit the production and dissemination of terrorist content.
Online service providers and governments commit to: Recognise the important role of civil society in supporting the Call, including working with supporters to increase transparency and to implement the commitments in a manner consistent with international human rights law and a free, open and secure internet.	

Transparency, participation, and accountability in the policy-making and legislative processes that lead to the passing and enactment of the "applicable laws" is particularly important.[13] While this is an essential part of open government, how to do that well is a huge topic on which much has been published and it is beyond the scope of this guidance. Rather, the focus here is on transparency around implementation and enforcement of existing laws and policies focused on the availability of

[13] Tech Against Terrorism calls for greater transparency in legislative processes in the Online Regulation Series 3.0 (2023 edition).

TVEC on information and communication technology services. Also in scope are government requests and referrals under company terms of service.

In some countries, it is a criminal offence for an individual to possess, produce or distribute illegal terrorist content. Governments and online service providers that support the Call commit to ensure appropriate cooperation with law enforcement agencies for the purposes of investigating and prosecuting illegal online activity regarding detected and/or removed terrorist and violent extremist content. Where the agency that is responsible for issuing a takedown notice to an online service provider hosting illegal terrorist content also has the power to investigate the individual who posted it, transparency (e.g. reporting) would usefully cover requests for user data. In countries where relevant investigative powers sit with other agencies, it would be useful to include or at least reference relevant data or reporting by those agencies. However, government demands and requests to tech companies for other law enforcement and security purposes are beyond the scope of the Call and this guidance.

Why is government transparency important?

Transparency is fundamental to building citizens' trust and participation in democracies, and to effective and accountable government, consistent with rule of law and the protection and promotion of all human rights.

Democratic states typically have well-developed institutions and processes to create transparency about executive government, as a prerequisite for citizens to hold elected and non-elected officials accountable. Those include the hearing and inquiry powers of Legislatures and Parliament, independent oversight bodies like auditors and Ombudsmen, departmental or agency annual reporting requirements, and freedom of information and privacy laws. However, the information provided, e.g. in government departments' annual reports, is typically high level. It may lack the specificity and granularity to enable groups and individuals to understand, assess, debate and challenge laws and activities that interest them or impact their rights.

The Christchurch Call aims to eliminate terrorist and violent extremist content online while protecting and respecting human rights and a free, open, and secure internet. It is well recognised that government laws, policies, programmes, and actions in this space can impact – directly and indirectly, positively and negatively – on the full enjoyment of all fundamental rights and freedoms.

In the Call Community, we seek to centre the voices and experiences of victims, survivors, their families, friends, and communities. The rights to life and to security of the person must be protected and promoted through effective action against terrorist and violent extremist content online. This is a fundamental role for governments. However, individuals and communities most impacted by terrorism and violent extremism are also often negatively impacted by counter terrorism activities. BSR has noted, e.g. that the victims of overbroad restrictions on online speech have tended to be from Muslim communities, which in turn impacts other rights and freedoms, including access to information and to effective remedy.[14] It is also the case that many states abuse counter-terrorism laws to silence and oppress opponents and minorities.

If transparency is important for individuals and groups unfairly targeted and stigmatised by counter-terrorism, it is also important for those who *feel* unfairly targeted. Post-pandemic, many countries have experienced increasing distrust in government and 'big tech'. Sometimes, there is legitimate concern and debate about authorities' relationships with digital platforms, and the possible flow-on

[14] See e.g. BSR "Human Rights Assessment: Global Forum to Counter Terrorism", pg 27 July 2021

effects on freedom of speech.[15] In other cases, mistrust and anti-authority views can be exacerbated by the dis- and misinformation and conspiracy theories that can flourish in the absence of open, accessible, and specific information from governments.

This is why, at a high level, it is important for governments to provide transparency around their laws, policies and actions to remove or restrict terrorist and violent extremist content online.

More specifically, such transparency:

- Informs the public, including vulnerable communities and individuals, about what the government is doing to address terrorist and violent extremist content online and prevent harm.
- Deters officials from acting outside the law or applying it unfairly, to censor speech, to block access to information, or to obtain user data, illegally, for illegitimate reasons or in overly broad or discriminatory ways.
- Builds public trust and confidence that laws are being implemented and enforced fairly and effectively.
- Empowers everyone to exercise their rights online and – in conjunction with legislative and independent oversight bodies, the judiciary, media, and civil society – to assess and hold governments accountable for wrongful decisions, illegal actions, and ineffective policies and programmes.
- Leads to better laws, policies and practices over time, by forcing internal reflection, establishing an evidence base on how well the status quo is working, and generating improvements informed by diverse perspectives.
- Enhances domestic legitimacy while also demonstrating adherence to international human rights law and counter-terrorism standards for a global audience that includes other states.

Clearly, government action to combat terrorist and violent extremist content online impacts online service providers and users all around the world. In the context of the Christchurch Call, which is grounded in a global multi-stakeholder approach, it is incumbent on government supporters to improve transparency alongside the tech sector, especially as they seek to impose mandatory reporting and other transparency obligations on it. Such transparency is the foundation for genuine dialogue in our Community, for the development of mutual understanding and trust, and for effective collective action.

[15] See e.g. Tech Policy Press "What Should We Know About Government Influence on Content Moderation" February 2023; also

Key Pieces of Community Advice

Over the last couple of years, experts in industry, academia and civil society have begun to publish guidance specifically for governments, focussing on their role in enabling and creating meaningful transparency. In this section, we summarise these existing resources.

CDT's framework for meaningful transparency

In 2021, the Centre for Democracy and Technology (CDT) put forward a framework for policymakers to help make transparency more meaningful for stakeholders.[16] It recognises that there are many different forms of transparency, produced by different actors, for different audiences and purposes. Those forms include:

- Transparency **reporting** containing aggregate data and qualitative information about moderation actions, disclosures and other actions concerning user generated content and accounts.
- User **notifications** about government demands for moderation of their content or their data.
- **Access to data** held by intermediaries for independent researchers, public policy advocates and journalists.
- Public facing **analysis, assessments, and audits** of tech companies with respect to user speech and privacy.

Each of these forms of transparency has its own strengths and limitations. Brought together in a considered way, these approaches can reveal the 'big picture' i.e. how the policies, processes, decisions and actions of governments and tech companies interact, shape online content, and affect users.[17]

As part of the Call's work on government transparency, CDT suggested the Call framework include descriptions of laws, policies, and administrative arrangements, which could be provided on a government department's website or as part of periodic reporting. CDT also suggested that governments could directly and proactively provide all these forms of transparency, except user notifications. Although we are not aware of any that provide researcher access to data in practice, there are certainly relevant precedents[18] and this is an area that could be explored by the Community in more detail. Governments also shape the overall framework (including mandating or blocking tech companies' provision of user notifications), through regulation, co-regulation and in voluntary cooperative initiatives with the tech sector - see EU example below.

Santa Clara 2.0

The Santa Clara Principles on Transparency and Accountability were first developed by a group of human rights organisations and academics – including CCAN members – in 2018, focusing on tech companies and the rights of their users, and the information needs of academics and civil society in helping defend and promote those rights. The principles were updated in 2021, with one aim being to address the role that governments play alongside tech companies in content moderation and the risks and concerns that arise.

---

[16] Vogus, C and Llanso, E "Making Transparency Meaningful: A Framework for Policy Makers" Centre for Democracy and Technology December 2021

[17] ibid

[18] See e.g. New Zealand Police Research – External Researcher Requests

Santa Clara 2.0 says that there should be transparency around government involvement in the development and enforcement of a tech company's rules and policies. Users should know when a state has been involved in action on their content or account, whether on legal grounds or under the company's own terms of service. This is a foundational principle that applies to all companies, no matter the type, size or age. Tech companies need to provide user notifications to enable this kind of transparency.

The updated principles also state that governments should recognise and minimise their roles in obstructing tech company transparency and must also provide transparency around their own involvement in content moderation decisions, including demands for content removal or restriction. Transparency reporting by governments is therefore also an important part of the overall framework for delivering against the Santa Clara principles.

Tech Against Terrorism's Guidelines for Government

In 2021 Tech Against Terrorism published guidelines for government transparency reporting on counter-terrorism efforts. They complement separate guidelines for tech company transparency reporting and are based on a 'waterfall' model that covers:

- Legal basis i.e. the law(s) under which government agencies refer content or accounts to tech companies for action, and/or request user data.
- Processes and systems that the government agencies use to detect terrorist and violent extremist content and accounts, and to refer them for action either on the basis of local law or the tech company's terms of service. This also covers record keeping and redress mechanisms.
- Outcomes as revealed by quantitative metrics for referrals and requests; broken down by means of discovery, ground for action, company and outcome; and appeals and complaints received as well as the outcomes.

The guidelines contain templates and examples of the qualitative and quantitative information that should be included in a government transparency report. They go a long way towards operationalising the Santa Clara principles as they relate to terrorist and violent extremist content. The guidelines are therefore a particularly useful resource for Call supporter governments and the wider Community.

Government Transparency Reporting

Principles

Drawing on the above resources, the wider literature and our Community sessions, the following principles are intended to guide government supporters of the Call in their efforts to promote transparency in the form of reporting:

- Consider the whole online content ecosystem, the different actors and stakeholders involved, and how best to provide them with relevant information and data. Think about which elements of CDT's framework for meaningful transparency are most relevant to promoting understanding, accountability, and continuous improvement in your country, and in the multistakeholder forums your country participates in. Consider how periodic reporting fits in the framework and adds value for stakeholders.

- Identify key audiences, internal and external, domestic and international, and consult them on the purpose and their specific use(s) of information and data that will go into your reporting, as the 'why' should inform the approach and content.[19]
- Contextualise the information and data in the report. Describe the different ways and places your government interacts with tech companies to address TVEC online. Take stock of and explain the unique laws and administrative arrangements for government requests and referrals, and transparency, in your country.
- If multiple departments or agencies make orders and requests, or otherwise engage with companies on content moderation, aim to ensure appropriate coverage of these agencies' efforts, depending on the information needs of key stakeholders. Reference relevant information and reporting from these other sources, as well as from any independent review or oversight bodies. Note that the Santa Clara Principles call for consolidated reporting across different kinds of illegal or policy-violating content (e.g. adult, copyright, defamation, hate, privacy, violence). Consolidated reporting of this kind will be particularly valuable for some stakeholders. Others will derive greater value from subject-specific reporting, which may be more detailed and regular.
- Acknowledge and seek to address limitations and barriers to comprehensive reporting, including legal, financial, and practical impediments.
- Look at tech companies' transparency reports and the data they provide on your government's orders and requests. Seek to understand and explain differences in scope between company and government reporting and apparent anomalies.
- Make information and data accessible and easy to understand. Use plain language as much as possible. Define key terms, especially technical ones. Analyse and interpret the data. Describe the nature of content being referred. Explain how any incidents or trends during the reporting period impacted on the nature and volume of content referred to tech companies. Avoid spin.
- Provide regular updates e.g. on an annual or six-monthly basis.
- Strive for consistency across reporting periods. But do not let that prevent innovations in approach that will better serve stakeholders and their interests. Make sure to highlight any changes in the context for or approach to reporting since the last one and explain any impacts on the information presented.
- Make a start and don't let the perfect be the enemy of the good. As is the case for tech companies, producing a government transparency report can be resource intensive and require new systems and tools as well as coordination across multiple units within and beyond a specific department.[20] Governments will be starting from different places.
- Look to continuously improve. Actively seek feedback from stakeholders. Look to the leading edge of tech company practice, too, for approaches that may better serve them. Tech companies are e.g. evolving away from static PDF reports to comprehensive transparency centres, with a range of interactive, accessible, and machine-readable methods of disclosing information and data.[21]

---

[19] Action Coalition on Meaningful Transparency https://www.meaningfultransparency.tech/post/reporting-summary

[20] ACT Discussion Summary: What goes into preparing a tech transparency report? (meaningfultransparency.tech)

[21] ibid

- Recognise the inherent limitations of transparency reporting[22] and convene multi-stakeholder discussions to enable trusted information-sharing with key stakeholders, as well as dialogue around your government's laws and activities, and the impacts on safety, security, and other fundamental rights and freedoms.

Categories and Content

The table below provides a possible structure and overview of the categories and types of content that governments may consider providing e.g. in transparency reports. It is based on the recommendations in Tech Against Terrorism's guidelines and advice published by other experts in the Community and beyond[23], tailored to the Call context and scope of this guidance, and reflecting the discussions in our Community sessions. It is recommended that governments refer to the source documents for detail and nuance as they work to develop a transparency framework and reporting.

Describe the legal and policy framework in your country

- What are your laws and policies against online terrorist and violent extremist content?
- How do you define terrorism and terrorist content, and violent extremism and violent extremist content? What legislation and/or policy are these definitions set out in? What national and/or intergovernmental terrorist designation list(s) do you use in this context, if any?
- What are the legal provisions that enable officials and/or law enforcement agents to order or request action against online terrorist or violent extremist content and related accounts, and to make related orders or requests for user data?
- Are there any legal provisions that can be used in this context to prevent or delay a tech company notifying users that the government or state agent has ordered or requested action against their content or account, or provision of their data? What are they and when/how often are they used?
- Are there any legal decisions, interpretations or opinions that government officials and/or law enforcement agents rely on in defining content and executing the above powers?
- What international treaties have you signed and ratified that bear on the execution of these powers?

Explain who is involved in implementing your laws and policies

- Which government departments and officials, and/or state entities and agents, are legally mandated and/or able in practice to make relevant orders and requests or referrals against terms of service?
- Are there any legal provisions or policies or practices (e.g. Codes of Conduct) that constrain the ability of officials and state agents to make orders and requests, or referrals

[22] Integrity Institute Shining a Light on Platform Transparency Best Practices July 2023: page 15.

[23] E.g. Llanso, E and Morgan S "Getting Specific About Transparency, Privacy and Free Expression Online" CDT; GCTF "Policy Toolkit: The GCTF Zurich-London Recommendations on Preventing and Countering Violent Extremism and Terrorism Online"; Lai, S, Shiffman, N and Wanless, A "Operational Reporting by Online Services: A Proposed Framework" Carnegie Endowment for International Peace May 2023.

against terms of service?

Describe how your laws and policies are operationalised and enforced in practice

- What approaches, systems and tools do the relevant officials and/or state agents use to discover terrorist and violent extremist content online?
- How do they submit orders and requests to tech companies?
- How do they make referrals against terms of service?
- Are there any internal review mechanisms to ensure orders, requests and referrals are lawful and well-founded before they are sent?
- What data is recorded and stored about discovery of online terrorist and violent extremist content online, and in relation to orders and requests, and referrals, sent to tech companies?

Describe any structures, mechanisms and provisions for independent oversight, review and redress

- What structures or processes exist to provide independent oversight of the department(s) and/or agencies involved? Are orders, requests and referrals externally reviewed? If so, how?
- What channels and forms of redress are available to users whose content or accounts or data has been wrongfully subject to an order, request or referral?
- How do you support companies in providing transparency and facilitating redress in cases where their content or account has been wrongfully actioned or data wrongfully obtained?

Provide data and explanations on activity and outcomes

- How much terrorist and violent extremist content has been proactively discovered using the various processes and systems? How much has been discovered because of reports from the public?
- Quantity of content that violates law or policy, disaggregated by: source of discovery; nature of violation; terrorist group and designation status, or ideological category; and company.
- Number of legal content-related orders and requests made to tech companies, disaggregated by: company; nature of legal violation; terrorist group and designation status, or ideological category; action ordered or requested; outcome.
- Number of referrals and requests made outside legal channels, disaggregated by company; nature of policy violation; terrorist group and designation status, or ideological category; action requested; outcome.
- If relevant, number of prosecutions against tech companies and/or individuals, disaggregated by: nature of legal violation; outcome, including penalty.
- Appeals or other challenges made against government or state agent orders or requests, disaggregated by: company; type of order or request (legal or ToS referral); success rates; remedies.

Other useful information

Consider including information on:

- Any formal or informal working relationships with tech companies or the multistakeholder community, regular discussions or collaborative projects, in pursuit of the elimination of terrorist and violent extremist content in a transparent and human rights-respecting way.
- Your government's efforts to deliver its other Call commitments, e.g. regulatory and policy developments, activities aimed at awareness-raising and capacity-building for small firms, and crisis response.

Examples and Good Practices

1: The Internet Hotline Centre, Japan

In Japan, The Internet Hotline Center (IHC) is an independent non-governmental organisation responsible for assessing potentially illegal or harmful content online and referring it to online service providers for action.

Any internet user or member of the public can flag such content to IHC. It then assesses the content against operational guidelines and refers illegal or harmful content to the relevant online service provider(s) or website administrators for action. The IHC may request that the online service provider removes or blocks the content if it is illegal or takes voluntary action on harmful content based on terms of service. Although there is no specific prohibition on terrorist or violent extremist content, it would fall under categories of harmful content related to depictions of / incitement to serious violent crimes, and/or sale or manufacture of weapons.

IHC also reports content to the National Police Agency, where it appears to violate national law or be related to a significant crime or represent an imminent threat to life (e.g. suicide).

IHC publishes six-monthly and annual data on its website, including the total number of reports it has received, broken down by the type of illegal or harmful information and action taken by IHC (including reports to Police) and online service providers.

The IHC's work is periodically reviewed and evaluated by an Operating Committee of five external people from academia and civil society.

Good practices

- State actors are not directly involved in content moderation.
- IHC publishes its operational guidelines (including definitions of different kinds of illegal and harmful content) and regular (six-monthly) on activities and outcomes.
- There is provision for independent review of the IHC's work, and multistakeholder evaluation of subtle decisions that can feed into updates to the operational guidelines.

2: Independent Review, ARCOM, France

ARCOM is the French Regulatory Authority for Audio-visual and Digital Communication. Its role is to guarantee freedom of expression and access to information online.

Since 2014, the Central Cybercrime Prevention Office (OCLCTIC) of the Judicial Police has had the power under French law to issue take down notices to online service providers hosting terrorist

content. If the online service provider does not remove the content within 24 hours, OCLCTIC can ask that the content be blocked or dereferenced.

To enable fast action there is no need for OCLCTIC to obtain a court order for these requests. Instead, the Board of ARCOM is empowered to appoint a qualified member of its board to oversee OCLCTIC operations and review all its requests to ensure compliance with the law.

The current appointee is Laurence Pécaut-Rivolier, a judge and member of the Court of Cassation.

She may recommend that OCLCTIC revoke a notice that is not well-founded in law and ask that the relevant content be reinstated. Where there is a difference of opinion between her and OCLCTIC, she may refer the matter for judicial review.

In this role, Ms Pécaut-Rivolier must report annually to the Government and Parliament on the number of orders issued by OCLTIC and the number of pieces of terrorist content taken down or blocked or dereferenced as a result. It is her practice to include information about the instances where she has recommended revocation of a notice and how they concluded.

The 2022 ARCOM report is published online in full and in summary form. Ms Pécaut-Rivolier also convened a meeting of industry and civil society stakeholders to discuss the report and its conclusions.

Under the EU TCO Regulation, Ms Pécault-Revolier is the competent authority for scrutinising cross-border removal orders against French hosting service providers. Arcom oversees the implementation of specific measures by HSPs exposed to terrorist content, and imposes any penalties. ARCOM will issue transparency reports under Article 8, and has contributed data and information to France's national reporting under Article 21.

Good Practices

- Transparency reporting on ARCOM's independent oversight and review functions means the French Parliament and public can be confident that OCLCTIC powers are being exercised in accordance with the law and fundamental rights and freedoms.
- Reporting includes information on:
    - ➢ Legislative framework and administrative arrangements for removal orders, oversight, review and redress.
    - ➢ Activities and outcomes, e.g. characteristics of terrorist content subject to removal orders in a given year.
    - ➢ There is an opportunity for multistakeholder dialogue around OCLCTIC and ARCOM activities and outcomes each year.

3: EU TCO Framework for Transparency

The regulation addressing the dissemination of terrorist content online came into application in June 2022. It empowers designated competent authorities in EU Member States to send removal orders to hosting service providers requesting they remove or disable access to terrorist content. They have one hour to comply. The Regulation also empowers Member States to decide, based on objective factors, whether a hosting service provider is exposed to terrorist content and therefore must take specific measures to address it.

The Regulation establishes a framework for transparency around the use of these powers and actions taken by hosting service providers, as an integral part of the safeguards for human rights. Within that

framework, there are transparency and reporting obligations for hosting service providers, competent authorities, Member States and the European Commission.

The European Commission must publish an online register of all designated competent authorities, based on information from Member States.

Under Article 7, hosting service providers must set out clearly in their terms of service their policy for addressing the dissemination of terrorist content and how they implement it. Any hosting service provider that has received a removal order and/or been required to take specific measures to address terrorist content on their service must issue a transparency report containing the number of removal orders received, details of specific measures taken, complaints, judicial or administrative reviews, and outcomes. They are also asked to provide a meaningful explanation of the functioning of specific measures.

Under Article 8, national competent authorities must publish annual transparency reports mirroring those required from hosting service providers and containing information on the number of removal orders sent, decisions to require specific measures against terrorist content by hosting service providers, judicial or administrative reviews, outcomes, and any penalties imposed.

Under Article 11, hosting service providers must notify users whose content has been taken down or blocked because of a removal order. The content provider can request reasons or a copy of the order. A competent authority can delay such notification for six weeks, where necessary and proportionate, to enable investigation and prosecution of a terrorist offence. That can be extended for another six weeks.

Under Article 21, Member States must bring together information from both hosting service providers and their national competent authorities and submit an annual monitoring report to the European Commission on the implementation of the regulation in their territory.

Under Articles 21 and 22, the European Commission must work with Member States to establish a programme for monitoring the outputs, results and impacts of the regulation and submit an overall implementation report to the European Parliament and the European Council in 2023.

Under Article 23, the European Commission must submit an evaluation by June 2024 to the European Parliament and European Council on the impact of implementing the regulation on security and fundamental human rights and freedoms.

The transparency reports from hosting service providers and competent authorities, and the European Commission’s implementation and evaluation report, are public. The Regulation sets out the minimum required information, although hosting service providers and national competent authorities can provide more.

Europol has developed a digital platform, PERCI, to connect competent authorities and hosting service providers and help them fulfil their obligations under the regulation. Competent authorities issue removal orders and referrals through the platform, enabling coordination across Member States and with Europol. Hosting service providers enter information about their response e.g. the time at which the content is taken down or blocked. This will support the production of annual transparency reports. PERCI will be further developed to provide meaningful information about actions taken by HSPs because of a referral or removal order.

The European Commission also chairs the EU Internet Forum. It is a platform for EU Member States, European agencies like Europol and the Fundamental Rights Agency, tech companies, partners like the GIFCT and Tech Against Terrorism, and experts from civil society to come together and discuss approaches to combating terrorist and violent extremist content online, in a voluntary and cooperative format. Discussions take place under the Chatham House Rule, to enable trust and open information-sharing. Information about the EUIF's objectives, rules of procedure, events and activities is published on its website.

Good Practices

- Establishes a framework for transparency and reporting that includes user notifications and reporting by tech companies and governments. Monitoring and implementation reports should also help enable the "big picture" to emerge.
- PERCI ensures systematic record keeping for relevant decisions and actions by all actors and should mean their reports are mutually consistent.
- Includes an evaluation of the impact of the regulation on security and human rights.
- EUIF provides a forum for multistakeholder discussion on emerging challenges and threats to develop joint action on a voluntary basis.

4: Transparency Reporting: EU Internet Referral Unit, Europol

The EU IRU has published annual transparency reports on a voluntary basis since 2017.

The reports describe EU IRU's preventive work across three workstreams: monitoring and analysing terrorist propaganda; referring terrorist content and content related to migrant smuggling to online service providers for assessment against their terms of service; and work in public private partnerships e.g. in the EU Internet Forum and GIFCT Crisis Response Working Group.

The TCO Regulation specifies that the EU IRU (and Member States) retain the ability to refer content to online service providers for assessment and voluntary action under terms of service. The section of the EU IRU's transparency report on referrals sets out the legal mandate for this aspect of the IRU's work, provides a brief description of the processes and systems that the IRU uses – including any thematic and platform-specific referral action days held during the year - and gives the number of pieces of content assessed as being related to terrorism and violent extremism, and subsequently referred to an online service provider for consideration and action. There is no breakdown by e.g. platform or outcome.

See the 2021 report for an example.

Good Practices

- Gives information on the mandate for and approach to ToS referrals.
- Describes where and how the EU IRU works with industry and other partners on issues of mutual concern e.g. crisis response, and research on emerging threats.

5: Transparency Reporting: Department of Internal Affairs, New Zealand

The Department of Internal Affairs set up its Digital Violent Extremism Team (DVET) in 2019, as part of the New Zealand Government's response to the Christchurch terror attack. The Department has

published transparency reports for 2021 and 2022. The aim is to help the public understand how DVET works to prevent exposure to harmful and illegal content.

The reports contain information on the legal framework in New Zealand, as well as data on referrals against terms of service, take down notices for illegal content, and search warrants and productions orders issued for the purpose of investigating offences under the Films, Videos and Publications Classification Act. Data is presented in graphs with explanatory material. Case studies help readers understand how the team goes about its work, the kind of content it has dealt with during the year, and how domestic and international events have influenced the online environment. Additional information is provided on incident response, Official Information Act requests and media queries.

The Department has worked closely with Tech Against Terrorism's and domestic partners (including the New Zealand Police and the Classifications Office) to develop these reports. Tech Against Terrorism reviews each report against the guidelines and makes public recommendations for further improvements.

Good Practices

- Transparency reports contain most of the categories and types of information and data recommended in Tech Against Terrorism's guidelines.
- Data and information is presented in a way that helps different audiences interpret and understand it.
- The Department is also looking at innovations to make the data and information in the report even more readily accessible to key audiences, and for digital savvy youth.

6: Transparency Reporting, Ofcom, UK

Ofcom is the United Kingdom's independent communications regulator, currently responsible for regulating UK-based video-sharing platforms and due to take on broader online safety powers following the passage of the Online Safety Bill. Ofcom works with platforms to ensure they have effective systems and processes in place to protect all users from relevant harmful material, as required by law. Ofcom does not make requests or referrals in relation to specific pieces of content. Rather, it focuses on platforms' policies, systems and tools.

Ofcom published a transparency report on its first year of video-sharing regulation in October 2022. The report sets out the organisation's mandate and approach, as well as a summary of its key findings about regulated services for the year and priorities for the year ahead. The report also includes more detail on Ofcom's work with and assessment of the video-sharing platforms notified to it under the legislative regime, as well as with international partners. Ofcom has also published more detailed reports on specific incidents – such as the Buffalo, New York, terror attack in May 2022 – to analyse and address risks in the sector, as well as on its work with individual platforms to address particular harms.

With the passage of the Online Safety Bill in 2023, Ofcom will be required to set up a mandatory transparency framework for the regulatory regime, which will cover user-to-user platforms and search engines. Ofcom will consult on this framework but intends reporting requirements to be tailored to each specific regulated platform's functionality, user base and unique characteristics.

Good Practices

- Ofcom publishes detailed information on its work with individual video-sharing platforms to improve online safety, as a basis for public trust and confidence.

- Ofcom is required to consult on the transparency framework under the Online Safety Bill. It intends that reporting requirements will be tailored to specific types of platforms and information and data will be made available in various formats, including static reports and real-time dashboards.