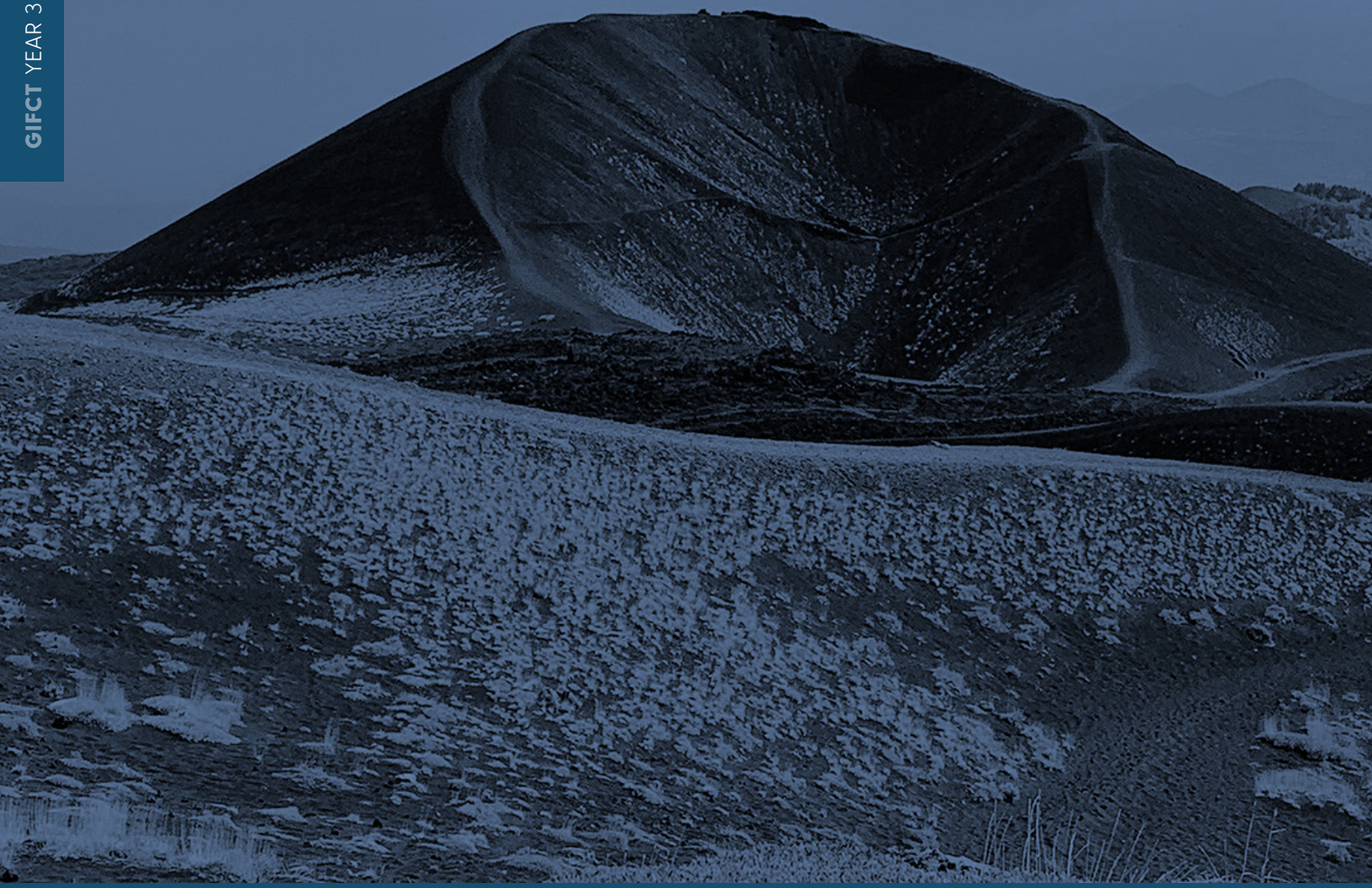
GIFCT YEAR 3 WORKING GROUP OUTPUT

Qualitative Indicators of Transparency During an Incident Response

GIFCT Incident Response Working Group

September 20, 2023



GIFCT
Global Internet Forum to Counter Terrorism

Laura DeBenedetto
Independent Researcher

About GIFCT Year 3 Working Group Outputs

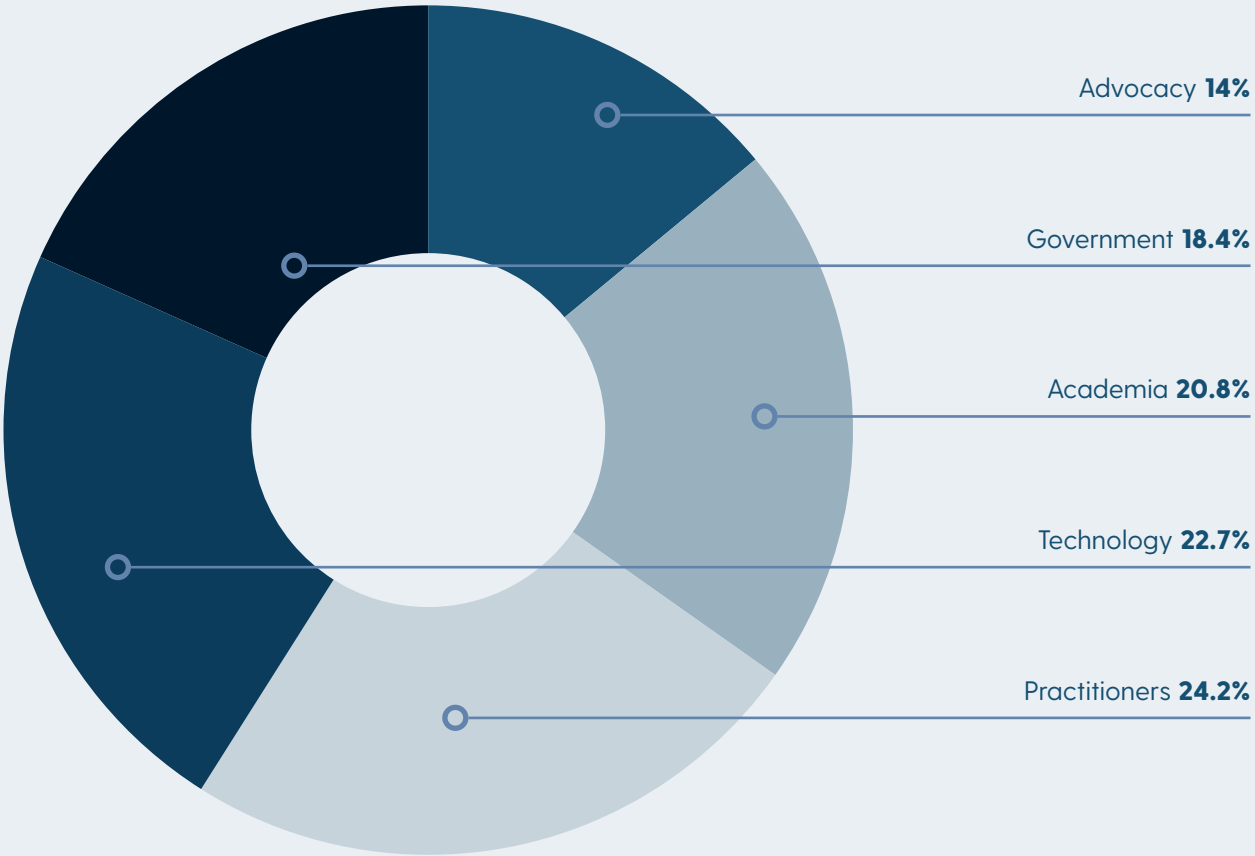
By Dr. Nagham El Karhili, Programming and Partnerships Lead, GIFCT

In November 2022, GIFCT launched its Year 3 Working Groups to facilitate dialogue, foster understanding, and produce outputs to directly support our mission of preventing terrorists and violent extremists from exploiting digital platforms across a range of sectors, geographies, and disciplines. Started in 2020, GIFCT Working Groups contribute to growing our organizational capacity to deliver guidance and solutions to technology companies and practitioners working to counter terrorism and violent extremism.

Overall, this year's five thematic Working Groups convened 207 participants from 43 countries across six continents with 59% drawn from civil society (14% advocacy organizations, 20.8% academia, and 24.2% practitioners), 18.4% representing governments, and 22.7% in tech.

WG Participants

Sectoral Breakdown



Beginning in November 2022, GIFCT Year 3 Working Groups focused on the following themes and outputs:

1. **Refining Incident Response: Building Nuance and Evaluation Frameworks:** This Working Group explored incident response processes and protocols of tech companies and the GIFCT resulting in a handbook. The handbook provides guidance on how to better measure and evaluate incident response around questions of transparency, communication, evaluation metrics, and human rights considerations.
2. **Blue Teaming: Alternative Platforms for Positive Intervention:** After recognizing a gap in the online intervention space, this GIFCT Working Group focused on highlighting alternative platforms through a tailored playbook of approaches to further PVE/CVE efforts on a wider diversity of platforms. This included reviewing intervention tactics for approaching alternative social media platforms, gaming spaces, online marketplaces, and adversarial platforms.
3. **Red Teaming: Assessing Threat and Safety by Design:** Looking at how the tech landscape is evolving in the next two to five years, this GIFCT Working Group worked to identify, and scrutinizes risk mitigation aspects of newer parts of the tech stack through a number of short blog posts, highlighting where safety-by-design efforts should evolve.
4. **Legal Frameworks: Animated Explainers on Definitions of Terrorism and Violent Extremism:** This Working Group tackled questions around definitions of terrorism along with the impact that they have on minority communities through the production of two complementary animated videos. The videos are aimed to support the global counterterrorism and counter violent extremism community in understanding, developing, and considering how they may apply definitions of terrorism and violent extremism.
5. **Frameworks for Meaningful Transparency:** In an effort to further the tech industry's continued commitment to transparency, this Working Group composed a report outlining the current state of play, various perspectives on barriers and risks around transparency reporting. While acknowledging the challenges, the Working Group provided cross sectoral views on what an ideal end state of meaningful transparency would be, along with guidance on ways to reach it.

We at GIFCT are grateful for all of the participants' hard work, time, and energy given to this year's Working Groups and look forward to what our next iteration will bring.

To see how Working Groups have evolved you can access Year One themes and outputs **HERE** and Year Two **HERE**.

Qualitative Indicators of Transparency During an Incident Response

The GIFCT Incident Response Working Group explored incident response processes and protocols of tech companies and GIFCT resulting in a Handbook on Measuring the Impact of Incident Response. The handbook provides guidance on how to better measure and evaluate incident response around questions of (1) communication, (2) qualitative and (3) quantitative transparency metrics, (4) human rights evaluation frameworks, (5) potential inclusions on measuring bystander footage, (6) and how to assess virality. This represents one section of the wider Handbook. All Working Group outputs are made available on the GIFCT Working Groups page.

Executive Summary

This year, the GIFCT Incident Response Working Group (IRWG) reviewed how government and technology organizations communicate their approaches to violent extremist incidents. The aim of this review is to establish recommendations for qualitative transparency, which here refers to the type and manner of information communicated externally regarding a violent extremist incident. It encompasses what information is communicated and how it is conveyed, with a focus on ensuring clarity, accuracy, and openness in order to foster public and platform user trust. This document will discuss six key qualitative indicators identified by the Working Group: Audience, Frequency, Outlets, Impacts, Feedback, and Iteration. This output complements the GIFCT members' resource guide[1] and builds on the work outlined in the GIFCT Transparency Report.[2] We limited the scope of this discussion to qualitative transparency indicators to external stakeholders during and after an incident; quantitative transparency indicators linked to key metrics will be addressed in a separate output.

In addition to the proposed indicators, the Working Groups have identified existing gaps in government and technology institutions' transparency, particularly regarding victims and their families, as well as the impact on individuals who share content related to an event in order to raise awareness. While the primary purpose of this section is to provide a set of qualitative indicators to assist both technology and government organizations in evaluating the effectiveness of their qualitative transparency, we also recognize broader opportunities for considering the human rights implications of content and account removal in such situations. This should come from a place of empathy to bridge the gap

1 GIFCT Member Resource Guide.

2 GIFCT Transparency Reports can be found on GIFCT's Transparency page.

with more marginalized communities often left out of these conversations. It is important to note that fully addressing this issue goes beyond the scope of this output, but it serves as a crucial reminder for organizations as they assess their incident-related protocols.

Introduction

The IRWG came together to discuss the current state of qualitative transparency across the government and technology organizations. The forum took perspectives from those working in government, technology, academia, and civil society to address the question of how government and technology institutions should structure communication about violent extremist events to their audiences through the lens of qualitative transparency.

Technology and government organizations have worked to refine their processes over the years to ensure concise and accurate communication. Each institution has its own protocols based on the stakeholders they serve and are able to tailor information that is best suited to their audience. While government and technology companies will likely have different information to share, we have seen increased collaboration to ensure efficiency and consistency. The key methods of collaboration include a number of crisis protocols.[3] It is important to highlight that while there is close collaboration, governments and technology companies have different objectives during a violent extremism incident. Governments are responsible for public safety and security and should focus on official updates and sharing verified information. Technology platforms have a responsibility to address harmful content and maintain the safety and integrity of their platforms. While their objectives differ, collaboration across these two sectors is imperative to taking meaningful action on incident-related content during a violent extremist event and ensuring public safety and awareness.

The following themes emerged from the Working Group conversation in relation to the status quo:

- During an incident, effective communication should follow certain principles. It should be concise, iterative, and on a need-to-know basis. Organizations should clearly communicate their plans regarding the timing of statements, the frequency of updates, and the intended recipients of the information when possible.
- Organizations need to address the potential human rights implications associated with removing incident-related content that individuals share on technology platforms. While quantitative metrics can provide insights into the number of removed pieces of content or accounts, qualitative indicators are essential in assessing how these actions might impact public discourse.
- The level of detail and frequency of communication may vary based on the type of institution (government or technology company) and the intended audience, and modifications may be

3 Examples include the EU Crisis Protocol, Christchurch Call Crisis Response Protocol (CRP), GIFCT Incident Response Framework (IRF), Australia Online Content Incident Arrangement, New Zealand Online Crisis Response Process, and United Kingdom Online Policy Unit Crisis Response Protocol.

necessary to align with the specific context and audience requirements.

Existing Knowledge

GIFCT provides annual transparency reports[4] to its stakeholders to highlight what was done during a past incident and the lessons learned. The Working Group used this as a starting point to tackle some of the gaps in communication and considerations around human rights.[5] The conversation centered on reviewing existing best practices in this space and discussing how governments and tech companies can optimize qualitative communications for their audience by looking at specific indicators.

During a violent extremist incident, both government and technology organizations strive to deliver accurate and timely information to their audiences. The timing of communication becomes crucial due to the immediate societal and media impacts of such incidents. However, a dilemma arises between the need to confirm information before widespread dissemination and the urgency associated with these events. It is worth noting that when organizations are forced to retract or correct information, it creates confusion among their audiences. Therefore, prioritizing accuracy over speed is essential in crisis communication to maintain clarity and avoid misinformation.

The GIFCT Incident Response Framework[6] was established to "address potential content circulating online resulting from an offline terrorist or violent extremist event." GIFCT reviews data from incident responses as noted in the Incident Response Framework (e.g., how quickly the protocol is activated and how much content is removed) to establish and fine-tune this framework after each activation. This Working Group sees this method of continuous improvement as critical to maintaining and improving the efficacy of the GIFCT Incident Response Framework and organizations' responses to violent extremist events.

Qualitative Transparency Indicators During Incident Response

The following list of qualitative indicators can help to ensure that organizations are addressing key qualitative concerns that surfaced during the Working Group. The IRWG proposes that these be incorporated by organizations to gauge how well they are addressing key areas of concern:

1. **Audience** - Who the communication is for?
2. **Frequency** - When will there be updates on the incident?

4 See GIFCT's Transparency page.

5 Human rights will be addressed more fully in a separate output.

6 Found in the GIFCT Content Incident Protocol.

3. **Outlets** - Where can individuals find updates?
4. **Impacts** - How the content is impacting society and what is happening to individuals who share incident-related content?
5. **Feedback** - How can individuals get in touch with relevant parties regarding new information about an incident?
6. **Iteration** - How can the process be improved going forward based on lessons learned?

When a violent extremist incident occurs, organizations not only need to execute their internal processes for managing and mitigating risk, but they also need to determine their communication content and cadence to their external audiences as incidents evolve. Information shared across any channel needs to be verified for accuracy. This is the case for both government and technology companies. The key elements to address are the "who what when where why" - this can be summed up quickly and concisely for a broad audience with multiple stakeholders. Updates to communication are common and necessary as an incident evolves and should be done on a need-to-know basis. In some cases, organizations will also conduct a debrief on an incident to review what went well and what can be improved.

The IRWG looked at existing approaches and examined what is going well and where there are opportunities to improve communication. We looked at how organizations manage ongoing communication and incident impact and used that to highlight potential modifications that could help optimize how organizations structure communications.

Due to the varying nature of questions and time constraints, it is not always possible to address all inquiries simultaneously. Because of this, it is crucial to engage in iterative communication during and immediately after an incident. Once organizations have the opportunity to evaluate the outcomes of the incident, conducting debriefs becomes essential. These debrief sessions facilitate in-depth discussions about the event and identify areas that require improvement, thereby ensuring a continuous enhancement cycle. Some good examples of this include GIFCT's debriefs on recent incidents.[7]

One of the essential areas of managing an incident is controlling the spread of related content across tech platforms. Technology platforms are responsible for identifying incident-related content and mitigating its spread by employing an array of tools to detect and remove violating content. In the case of perpetrator-filmed content, tech platforms can use hash-sharing and other forms of automation to mitigate the spread of imagery. While there might be several reasons that incident-related content is shared across platforms, it is important that platforms are transparent about their approach to removing the content. In the case of the Christchurch incident, we saw that some news outlets shared perpetrator content on their social media channels (specifically YouTube) as part of their coverage. Most tech platforms later removed it. There is an educational opportunity for individuals and

7 Examples of multi-stakeholder debriefs include GIFCT Memphis multistakeholder debrief and GIFCT Buffalo multistakeholder debrief.

media organizations who may share content to raise awareness versus those who share it to praise or support the act of violence. Governments are in a position to communicate that content should not be distributed, while technology companies should be clear that this content violates its policies and ensure that there are clear external community guidelines to ensure users understand why certain content is not permitted to be shared. As part of this, tech companies should have clear avenues to an appeals process in case something is incorrectly removed.

The Working Group also discussed the need to tailor information based on the audiences consuming incident information. This is especially relevant for governments. Depending on the organization, it makes sense to review what exactly a given audience needs to learn from a communication and determine the most useful information to convey. This scope and depth of information that organizations share will depend on the audience they aim to inform. While our Working Group can advise on some of the best practices, we also advise that each organization take stock of what has worked well for them and be open to feedback.

Recommendations and Implementation

Communication should continue to be brief and on a need-to-know basis. This is something that most organizations do well, but it is important to highlight the need to maintain this approach as part of the Working Group's recommendations. We would also like to highlight the following guiding communication themes for both governments and technology companies:

- Based on the institution and the intended audience, there may need to be modifications in terms of the level of detail and frequency of communication.
    - This can take the form of messaging that highlights the 5Ws and explicitly notes what information is relevant to different parties. Governments should be explicit about who the information pertains to when they release statements. When possible, we recommend parsing out communication for the following audiences:
        - **Impacted individuals** - These are individuals directly affected by an incident, such as those present at the location or injured. Communication should include information about the incident's impact, affected areas, and available emergency services.
        - **People in the area of the incident** - This refers to members of the public residing or present in the incident area. Communication should cover details about the affected area, safe spaces, available emergency services, and other pertinent information.
        - **Family members of impacted individuals** - This communication aims to inform individuals who may not be in the incident area but have a connection to someone who is affected. It should provide contact methods, emergency services information, how to reach relevant authorities, expected update timings, and additional resources

depending on the incident's nature.

  - **Media and general public** - This communication should be the broadest and focus on answering the who, what, when, where, and why (if such information is available). Safety information should be prioritized (e.g., such as areas to avoid). General details about the incident, including perpetrators, response efforts, anticipated updates, and other relevant information, should also be included.

- For governments, it is imperative that they understand the landscape of content being shared on platforms, so close collaboration with the technology sector is key. Governments should note how they view re-sharing content related to an incident in their communication (e.g., are there any instances where re-sharing is helpful or permitted?) and ensure alignment with platform policies.
- Technology companies should be transparent about the implication of sharing content. Because their primary audience is platform users, tech companies should be clear about the implications of incident-related content (e.g., there should be clear external guidelines around any context or parameters where technology companies allow sharing incident-related footage).

In addition to the themes above, it is recommended that both government and technology companies assess their communications using the six qualitative indicators derived from feedback received from the Working Group. The table below serves as a concise guide outlining these recommendations based on organization type. Addressing the questions under each qualitative indicator helps to gauge whether communication adequately addresses the primary areas of concern for different audiences during an incident.

Organization Type	Audience(s)	Frequency	Outlets	Impacts	Feedback	Iteration
Governments	· Impacted individuals · People in the area of an incident · Family of impacted individuals · General public/ media outlets	· How often will updates be available? · Why that cadence? · How has the incident impacted how governments are communicating?	· Government sites · Traditional media outlets	Discuss how tech companies are addressing incident-related content	Review communication cadence and efficacy	Discuss how to improve processes
Tech Companies	· Platform users	· Highlight what is being done and when updates will be available	· Company blogs · Newsroom posts, media	Highlight potential harms to users and the type of content being shared	Review accuracy in moderations and appeals of content	Discuss how to improve processes

Table 1: Recommendations to Governments and Tech Companies Across Areas of Concern

Although there may be variations in qualitative indicators between government and technology companies, it is crucial to prioritize coordination in order to achieve effective and inclusive communication with the public. By combining qualitative indicators with quantitative measures and taking into account human rights considerations, sustained collaboration can ensure meaningful transparency for all those affected by violent extremist events. Therefore, it is essential for government and technology institutions to maintain their cooperation, strengthen their collaboration, and leverage their respective expertise in order to continuously develop comprehensive strategies that prioritize meaningful transparency.

Copyright © Global Internet Forum to Counter Terrorism 2023

Recommended citation: Laura DeBenedetto, Qualitative Indicators of Transparency During an Incident Response (Washington, D.C.: Global Internet Forum to Counter Terrorism, 2023), Year 3 Working Groups.

GIFCT is a 501(c)(3) non-profit organization and tech-led initiative with over 20 member tech companies offering unique settings for diverse stakeholders to identify and solve the most complex global challenges at the intersection of terrorism and technology. GIFCT's mission is to prevent terrorists and violent extremists from exploiting digital platforms through our vision of a world in which the technology sector marshals its collective creativity and capacity to render terrorists and violent extremists ineffective online. In every aspect of our work, we aim to be transparent, inclusive, and respectful of the fundamental and universal human rights that terrorists and violent extremists seek to undermine.

www.gifct.org **outreach@gifct.org**