Introducing 2022 GIFCT Working Group Outputs

GIFCT
Global Internet Forum
to Counter Terrorism

Dr. Erin Saltman
Director of Programming,
GIFCT

In July 2020, GIFCT launched a series of Working Groups to bring together experts from across sectors, geographies, and disciplines to offer advice in specific thematic areas and deliver on targeted, substantive projects to enhance and evolve counterterrorism and counter-extremism efforts online. Participation in Working Groups is voluntary and individuals or NGOs leading Working Group projects and outputs receive funding from GIFCT to help further their group's aims. Participants work with GIFCT to prepare strategic work plans, outline objectives, set goals, identify strategies, produce deliverables, and meet timelines. Working Group outputs are made public on the GIFCT website to benefit the widest community. Each year, after GIFCT's Annual Summit in July, groups are refreshed to update themes, focus areas, and participants.

From August 2021 to July 2022, GIFCT Working Groups focused on the following themes:

- Crisis Response & Incident Protocols
- Positive Interventions & Strategic Communications
- Technical Approaches: Tooling, Algorithms & Artificial Intelligence
- Transparency: Best Practices & Implementation
- Legal Frameworks

A total of 178 participants from 35 countries across six continents were picked to participate in this year's Working Groups. Applications to join groups are open to the public and participants are chosen based on ensuring each group is populated with subject matter experts from across different sectors and geographies, with a range of perspectives to address the topic. Working Group participants in 2021–2022 came from civil society (57%), national and international government bodies (26%), and technology companies (17%).

Participant diversity does not mean that everyone always agrees on approaches. In many cases, the aim is not to force group unanimity, but to find value in highlighting differences of opinion and develop empathy and greater understanding about the various ways that each sector identifies problems and looks to build solutions. At the end of the day, everyone involved in addressing violent extremist exploitation of digital platforms is working toward the same goal: countering terrorism while respecting human rights. The projects presented from this year's Working Groups highlight the many perspectives and approaches necessary to understand and effectively address the ever-evolving counterterrorism and violent extremism efforts in the online space. The following summarizes the thirteen outputs produced by the five Working Groups.

Crisis Response Working Group (CRWG):

The GIFCT Working Group on Crisis Response feeds directly into improving and refining GIFCT's own Incident Response Framework, as well as posing broader questions about the role of law enforcement, tech companies, and wider civil society groups during and in the aftermath of a terrorist or violent extremist attack. CRWG produced three outputs. The largest of the three was an immersive virtual series of Crisis Response Tabletop Exercises, hosted by GIFCT's Director of Technology, Tom Thorley. The aim of the Tabletops was to build on previous Europol and Christchurch Call-led Crisis Response events, with a focus on human rights, internal communications, and external strategic communications in and around crisis scenarios. To share lessons learned and areas for improvement and refinement, a summary of these cross-sector immersive events is included in the 2022 collection of Working Group papers.

The second output from the CRWG is a paper on the Human Rights Lifecycle of a Terrorist Incident, led by Dr. Farzaneh Badii. This paper discusses how best GIFCT and relevant stakeholders can apply human rights indicators and parameters into crisis response work based on the 2021 GIFCT Human Rights Impact Assessment and UN frameworks. To help practitioners integrate a human rights approach, the output highlights which and whose human rights are impacted during a terrorist incident and the ramifications involved.

The final CRWG output is on Crisis Response Protocols: Mapping & Gap Analysis , led by the New Zealand government in coordination with the wider Christchurch Call to Action. The paper maps crisis response protocols of GIFCT and partnered governments and outlines the role of tech companies and civil society within those protocols. Overall, the output identifies and analyzes the gaps and overlaps of protocols, and provides a set of recommendations for moving forward.

Positive Interventions & Strategic Communications (PIWG):

The Positive Interventions and Strategic Communications Working Group developed two outputs to focus on advancing the prevention and counter-extremism activist space. The first is a paper led by Munir Zamir on Active Strategic Communications: Measuring Impact and Audience Engagement. This analysis highlights tactics and methodologies for turning passive content consumption of campaigns into active engagement online. The analysis tracks a variety of methodologies for yielding more impact-focused measurement and evaluation.

The second paper, led by Kesa White, is on Good Practices, Tools, and Safety Measures for Researchers. This paper discusses approaches and safeguarding mechanisms to ensure best practices online for online researchers and activists in the counterterrorism and counter-extremism sector. Recognizing that researchers and practitioners often put themselves or their target audiences at risk, the paper discusses do-no-harm principles and online tools for safety-by-design methodologies within personal, research, and practitioner online habits.

Technical Approaches Working Group (TAWG):

As the dialogue on algorithms and the nexus with violent extremism has increased in recent years, the Technical Approaches Working Group worked to produce a longer report on Methodologies to Evaluate Content Sharing Algorithms & Processes led by GIFCT's Director of Technology Tom Thorley in collaboration with Emma Llanso and Dr. Chris Meserole. While Year 1 of Working Groups produced a paper identifying the types of algorithms that pose major concerns to the CVE and counterterrorism sector, Year 2 output explores research questions at the intersection of algorithms, users and TVEC, the feasibility of various methodologies and the challenges and debates facing research in this area.

To further this technical work into Year 3, TAWG has worked with GIFCT to release a Research Call for Proposals funded by GIFCT. This Call for Proposals is on Machine Translation. Specifically, it will allow third parties to develop tooling based on the gap analysis from last year's TAWG Gap Analysis. Specifically, it seeks to develop a multilingual machine learning system addressing violent extremist contexts.

Transparency Working Group (TWG):

The Transparency Working Group produced two outputs to guide and evolve the conversation about transparency in relation to practitioners, governments, and tech companies. The first output, led by Dr. Joe Whittaker, focuses on researcher transparency in analyzing algorithmic systems. The paper on Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence reviews how researchers have attempted to analyze content-sharing algorithms and indicates suggested best practices for researchers in terms of framing, methodologies, and transparency. It also contains recommendations for sustainable and replicable research.

The second output, led by Dr. Courtney Radsch, reports on Transparency Reporting: Good Practices and Lessons from Global Assessment Frameworks. The paper highlights broader framing for the questions around transparency reporting, the needs of various sectors for transparency, and questions around what meaningful transparency looks like.

The Legal Frameworks Working Group (LFWG):

The Legal Frameworks Working Group produced two complementary outputs.

The first LFWG output is about Privacy and Data Protection/Access led by Dia Kayyali. This White Paper reviews the implications and applications of the EU's Digital Services Act (DSA) and the General Data Protection Regulation (GDPR). This includes case studies on Yemen and Ukraine, a data taxonomy, and legal research on the Stored Communications Act.

The second LFWG output focuses on terrorist definitions and compliments GIFCT's wider Definitional Frameworks and Principles work. This output, led by Dr. Katy Vaughan, is on The Interoperability of Terrorism Definitions. This paper focuses on the interoperability, consistency, and coherence of terrorism definitions across a number of countries, international organizations, and tech platforms. Notably, it highlights legal issues around defining terrorism based largely on government lists and how they are applied online.

Research on Algorithmic Amplification:

Finally, due to the increased concern from governments and human rights networks about the potential link between algorithmic amplification and violent extremist radicalization, GIFCT commissioned Dr. Jazz Rowa to sit across three of GIFCT's Working Groups to develop an extensive paper providing an analytical framework through the lens of human security to better understand the relation between algorithms and processes of radicalization. Dr. Rowa participated in the Transparency, Technical Approaches, and Legal Frameworks Working Groups to gain insight into the real and perceived threat from algorithmic amplification. This research looks at the contextuality of algorithms, the current public policy environment, and human rights as a cross-cutting issue. In reviewing technical and human processes, she also looks at the potential agency played by algorithms, governments, users, and platforms more broadly to better understand causality.

We at GIFCT hope that these fourteen outputs are of utility to the widest range of international stakeholders possible. While we are an organization that was founded by technology companies to aid the wider tech landscape in preventing terrorist and violent extremist exploitation online, we believe it is only through this multistakeholder approach that we can yield meaningful and long-lasting progress against a constantly evolving adversarial threat.

We look forward to the refreshed Working Groups commencing in September 2022 and remain grateful for all the time and energy given to these efforts by our Working Group participants.

Participant Affiliations in the August 2021 - July 2022 Working Groups:

Tech Sector	Government Sector	Civil Society / Academia / Practitioners	Civil Society / Academia / Practitioners
ActiveFence	Aqaba Process	Access Now	Lowy Institute
Amazon	Association Rwandaise de Défense des Droits de l'Homme	Anti-Defamation League (ADL)	M&C Saatchi World Services Partner
Automattic	Australian Government - Department of Home Affairs	American University	Mnemonic
Checkstep Ltd.	BMI Germany	ARTICLE 19	Moonshot
Dailymotion	Canadian Government	Australian Muslim Advocacy Network (AMAN)	ModusIzad - Centre for applied research on deradicalisation
Discord	Classification Office, New Zealand	Biodiversity Hub International	New America's Open Technology Institute
Dropbox, Inc.	Commonwealth Secretariat	Bonding Beyond Borders	Oxford Internet Institute
ExTrac	Council of Europe, Committee on Counter-Terrorism	Brookings Institution	Partnership for Countering Influence Operations, Carnegie Endowment for International Peace
Facebook	Department of Justice - Ireland	Business for Social Responsibility	Peace Research Institute Frankfurt (PRIF); Germany
JustPaste.it	Department of State - Ireland	Centre for Analysis of the Radical Right (CARR)	PeaceGeeks
Mailchimp	Department of State - USA	Center for Democracy & Technology	Point72.com
MEGA	Department of the Prime Minister and Cabinet (DPMC), New Zealand Government	Center for Media, Data and Society	Polarization and Extremism Research and Innovation Lab (PERIL)
Microsoft	DHS Center for Prevention Programs and Partnerships (CP3)	Centre for Human Rights	Policy Center for the New South (senior fellow)
Pex	European Commission	Centre for International Governance Innovation	Public Safety Canada & Carleton University
Snap Inc.	Europol/EU IRU	Centre for Youth and Criminal Justice (CYCJ) at the University of Strathclyde, Scotland.	Queen's University
Tik Tok	Federal Bureau of Investigation (FBI)	Cognitive Security Information Sharing & Analysis Center	Sada Award, Athar NGO, International Youth Foundation
Tremau	HRH Prince Ghazi Bin Muhammad's Office	Cornell University	Shout Out UK
Twitter	Ministry of Culture, DGMIC - France	CyberPeace Institute	Strategic News Global
You Tube	Ministry of Foreign Affairs - France	Dare to be Grey	S. Rajaratnam School of International Studies, Singapore (RSIS)
	Ministry of Home Affairs (MHA) - Indian Government	Dept of Computer Science, University of Otago	Swansea University
	Ministry of Justice and Security, the Netherlands	Digital Medusa	Tech Against Terrorism
	National Counter Terrorism Authority (NACTA) Pakistan	Edinburgh Law School, The University of Edinburgh	The Alan Turing Institute

	Organisation for Economic Co-operation and Development (OECD)	European Center for Not-for-Profit Law (ECNL)	The Electronic Frontier Foundation
	Office of the Australian eSafety Commissioner (eSafety)	Gillberg Neuropsychiatry Centre, Gothenburg University, Sweden,	The National Consortium for the Study of Terrorism and Responses to Terrorism (START) / University of Maryland
	Organization for Security and Co-operation in Europe (OSCE RFoM)	George Washington University, Program on Extremism	Unity is Strength
	Pôle d'Expertise de la Régulation Numérique (French Government)	Georgetown University	Université de Bretagne occidentale (France)
	North Atlantic Treaty Organization, also called the North Atlantic Alliance (NATO)	Georgia State University	University of Auckland
	Secrétaire général du Comité Interministériel de prévention de la délinquance et de la radicalisation	Global Network on Extremism and Technology (GNET)	University of Groningen
	State Security Service of Georgia	Global Disinformation Index	University of Massachusetts Lowell
	The Royal Hashemite Court/ Jordanian Government	Global Network Initiative (GNI)	University of Oxford
	The Office of Communications (Ofcom), UK	Global Partners Digital	University of Queensland
	UK Home Office	Global Project Against Hate and Extremism	University of Salford, Manchester, England,
	United Nations Counter-terrorism Committee Executive Directorate (CTED)	Groundscout/Resonant Voices Initiative	University of South Wales
	UN, Analytical Support and Sanctions Monitoring Team (1267 Monitoring Team)	Hedayah	University of the West of Scotland
	United Nations Major Group for Children and Youth (UNMGCY)	Human Cognition	Violence Prevention Network
	United States Agency for International Development (USAID)	Institute for Strategic Dialogue	WeCan Africa Initiative & Inspire Africa For Global Impact
		International Centre for Counter-Terrorism	Wikimedia Foundation
		Internet Governance Project, Georgia Institute of Technology	World Jewish Congress
		Islamic Women's Council of New Zealand	XCyber Group
		JOS Project	Yale University, Jackson Institute
		JustPeace Labs	Zinc Network
		Khalifa Ihler Institute	
		KizBasina (Just-a-Girl)	
		Love Frankie	

GIFCT WORKING GROUPS OUTPUT 2022

Privacy and Data Protection / Access

GIFCT Legal Frameworks Working Group

GIFCT
Global Internet Forum
to Counter Terrorism

Dia Kayyali
Mnemonic

Introduction

Last year's Legal Frameworks Working Group whitepaper focused on "the issues relating to the work of technology companies disrupting terrorist and violent extremist content (TVEC) and the intersection with access to data" by parties such as "industry, research, academic, or civil society actors."[1] It identified high-level issues associated with this topic and provided some recommendations. This paper dives deeper into the topic through case studies, further legal and literature reviews, and interviews.

Online platforms, in particular social media platforms, host a variety of content that they might, for one reason or another, determined to be "terrorist or violent extremist content." The process of designating content as TVEC can be accelerated through participation in GIFCT, both through access to the hash database and information sharing. This content and associated data, whether or not it actually violates any laws or platforms' rules, is valuable for a variety of important purposes. Those purposes include journalism, research on content moderation practices, industry research (for example, small platforms learning about content moderation), misinformation/disinformation research, probes into specific violent attacks with online components, and finally open-source investigations into human rights violations. However, providing free and open access to this data is a clear non-starter due to privacy and security issues.

Each of these uses of data presents unique challenges. This paper was initially meant to be very broad but it became clear through research that there is no "one size fits all" solution. In the process of this research, the re-escalation of hostilities against Ukraine started, emphasizing the importance of open-source investigations and the susceptibility of this content to removal. Furthermore, a GIFCT Content Incident Protocol was activated on May 15 after a perpetrator livestreamed himself at a grocery store carrying out a mass shooting that targeted Black Americans.[2]

Thus, this paper focuses on the use of content hosted online for open-source investigations. The paper also outlines broad concerns regarding data preservation and access and touches on some of the challenges posed by government requests for data associated with offline attacks with an online component.

The Legal Frameworks Working Group recommends that civil society organizations and government work together to find a legislative solution that would provide a legal avenue for the International Criminal Court and United Nations investigative bodies to access removed content in a way that would not require platforms to violate the Stored Communications Act (SCA) or the European Union General Data Protection Regulation. It also recommends increased transparency from platforms and further research into access for other purposes. Finally, the working group urgently recommends that GIFCT commission (or work with governments and civil society organizations to produce) research into the kinds of data governments need to understand terrorist and violent extremist use of the

[1] Legal Frameworks Working Group, "Legal Frameworks Report," Global Internet Forum to Counter Terrorism, July, 2021, https://gifct.org/wp-content/uploads/2021/07/GIFCT-LegalFrameworks-WGroup.pdf.

[2] Global Internet Forum to Counter Terrorism, "Update: Content Incident Protocol Activated in Response to Shooting in Buffalo, New York United States," May 17, 2022 https://gifct.org/2022/05/14/cip-activated-buffalo-new-york-shooting/.

Internet after attacks. Between the time this paper is written and the time it is published, the Content Incident Protocol debrief may provide answers to some of these questions, but regardless, this final recommendation should be a priority specifically for GIFCT as the most relevant forum for discussions about online components of offline attacks.

Open-source investigations

Organizations like Bellingcat and Syrian Archive do open-source investigations (OSI) into human rights violations. OSI rely on user-generated content and are often key in investigating violations in places where traditional investigations may not be possible for a variety of reasons. Organizations discover content through various methods, archive it, and verify that it actually depicts what it says it does, putting together public or private data sets to aid in investigating specific cases. This content is particularly susceptible to being removed (whether correctly or not) by content moderation practices for a variety of reasons. The content can be very graphic and related to groups that are designated as terrorist organizations by U.S. or UN lists, by platforms themselves, or by other governments. Furthermore, considerable amounts of human rights related content is not in English but instead in non-Latin languages such as Arabic, Burmese, and Ukrainian. These factors in content removal have been explored by myriad content moderation and natural language processing experts and are not the focus of this paper, although GIFCT member companies would benefit from more focused research on these topics.

Ultimately, data that could be used by the International Criminal Court or other bodies may be lost and destroyed forever. This is especially challenging because it may be the only evidence available. However, asking platforms to preserve this content, or to provide access beyond their existing procedures for law enforcement, raises many legal and ethical issues.

The Berkeley Protocol on Digital Open Source Investigations, co-published by the United Nations and the Human Rights Center at the University of California, Berkeley, School of Law provides indispensable guidance for using this content, including consideration of ethical and legal issues.[3]

The Mnemonic Experience

In July of 2017, the NGO Syrian Archive and myriad Syrian human rights defenders that had amassed vast collections of documentation from Syria noticed that their content was being deleted and accounts being suspended more rapidly than ever before on YouTube.[4] These mass removals started less than a month after Google announced it would now be using machine learning to detect content it deemed to be "terrorist and violent extremist content."[5]

[3] United Nations Office of the High Commissioner and the Human Rights Center at the University of California Berkeley School of Law, "Berkeley Protocol on Digital Open Source Investigations," HR/PUB/20/2, 2022, https://www.ohchr.org/sites/default/files/2022-04/OHCHR_BerkeleyProtocol.pdf.

[4] Raja Althabani and Dia Kayyali, "Vital Human Rights Evidence in Syria is Disappearing from YouTube," WITNESS blog, August 30, 2017, https://blog.witness.org/2017/08/vital-human-rights-evidence-syria-disappearing-youtube/.

[5] The YouTube Team, "An update on our commitment to fight violent extremist content online," YouTube Official Blog, October 17, 2017, https://blog.youtube/news-and-events/an-update-on-our-commitment-to-fight/.

Many news outlets reported on the issue, and Syrian Archive started to track removals from their own collection, as well as assist account owners whose content had been removed. YouTube voluntarily restored thousands of videos over the next several years as Syrian Archive, along with the NGO WITNESS, worked to address the causes of these removals.[6]

The use of automation to detect and remove content, including the GIFCT hash database, has undisputedly continued to lead to high rates of removal. Statistics and statements from Meta, YouTube, and Twitter from the last five years confirm this.[7]

Mnemonic is the parent organization for Syrian Archive. Its "lost and found" project has tracked content removals to the greatest extent possible by comparing archived collections of content against the original URLs of that content and determining what is still online. For example, currently about a quarter of Syrian Archive's collection gathered from YouTube is no longer online. Some platforms actually provide enough information to determine the cause of removal while some do not. Mnemonic also relies on insight from relationships with users and organizations that post content and share why that content is removed.

Mnemonic sometimes sees direct spikes in removal in relation to external events. In Yemen, for example, Mnemonic found a large and unexplainable jump in content removals of Houthi accounts after the Trump administration indicated that it would be designating Houthis as a "foreign terrorist organization" (FTO).[8] From October 1 to November 30, 2020, "53,846 tweets became unavailable. That's highly abnormal. By comparison, it's more common for [Yemeni Archive] to see a few thousand tweets become available per month." These accounts were important sources of human rights evidence. Twitter said that these removals were due to its "Platform Manipulation and Spam" policy, but the timing of the removals remains suspicious. In Palestine, content removals often skyrocket during periods of intensified IDF activity. The Palestinian organization 7amleh has received reports of removals from users, and Human Rights Watch has also tracked extensive removals from their own investigations.[9]

What kind of data and how is it used?

Content hosted by online platforms and associated data, in particular social media platforms, is increasingly important for international and domestic justice mechanisms. As noted above,

[6] Hadi Al Khatib and Dia Kayyali, "YouTube is Erasing History," New York Times, October 23, 2019, https://www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.html.

[7] Monika Bickert and Brian Fishman, "Hard Questions: What Are We Doing to Stay Ahead of Terrorists?," Meta Newsroom, November 18, 2018, https://about.fb.com/news/2018/11/staying-ahead-of-terrorists/; Erin Saltman, " Identifying and Removing Terrorist Content Online: Cross-Platform Solutions," The Raisina Edit (blog series), 2022, https://www.orfonline.org/expert-speak/identifying-and-removing-terrorist-content-online/; Zoe Strozewski, "Twitter Suspended 44K Accounts for Promoting Terrorism, Violent Orgs in First Half of 2021," Newsweek, January 25, 2022, https://www.newsweek.com/twitter-suspended-44k-accounts-promoting-terrorism-violent-orgs-first-half-2021-1672868.

[8] Dia Kayyali, "What happens when the US decides to designate a group as a terrorist organization? Insights from Mnemonic," Mnemonic, February 18, 2022, https://mnemonic.org/en/content-moderation/What-happens-terrorist-designation.

[9] 7amleh- The Arab Center for the Advancement of Social Media, "The Attacks on Palestinian Digital Rights," May 21, 2021, https://7amleh.org/2021/05/21/7amleh-issues-report-documenting-the-attacks-on-palestinian-digital-rights; Belkis Wille, "'Video Unavailable', Social Media Platforms Remove Evidence of War Crimes," Human Rights Watch, September 10, 2020, https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes.

investigators discover user-generated content and associated data, such as the identity of users posting the content, through OSI.

The use of online content is not always readily apparent from outside of justice mechanisms. It can be used directly as evidence when it can meet applicable legal standards. However, it can also be used in a variety of other ways. It can be very helpful in early stages, where it can inform case building – i.e., help investigators know where to focus. It can also help courts and other international mechanisms create operational security and witness protection plans, locate witnesses and corroborate their statements, and establish timelines. This is important, because these uses of data are not readily visible to stakeholders outside of these mechanisms, and in fact these mechanisms cannot always be forthcoming about their uses of data for security or other reasons.

Last year's paper identified the following kinds of data which this paper refines. As noted last year, these categories are not mutually exclusive:

- Personally identifiable information (PII): this would include things like people's legal names, images of their faces, recordings of someone's voice, or addresses.
- TVEC: It should be noted that "TVEC" is a disputed term because it relies on the varied definitions of terrorism and violent extremism used by different platforms and governments. It's also not completely clear what part of the content "TVEC" refers to. Instead, the term "user-generated content" is more helpful. This term refers to the actual content itself, i.e., the words of a post or the video file or image file of a post, as well as associated data such as the title of a video. This would also include comments posted by other users.
- Metadata relating to content: This would be information like the IP address a particular piece of content was posted from.
- Non-personal descriptive data: This includes hashtags, key phrases, titles of broadcasts, etc. For purposes of this paper, this category is subsumed under "user-generated content."
- Contextual information about the sources of data.

As noted, these various categories of data can be used in different ways, and they implicate different ethical and legal concerns. For purposes of this paper, "data" is equivalent to "user-generated content" when referring to open source investigations, and is the main focus. The United Nations and the International Criminal Court, cases for accountability in Syria, and the growing number of justice efforts for war crimes in Ukraine provide helpful examples of the use of this data.

United Nations

The United Nations has a long history of investigating human rights violations through special investigatory bodies designed for that purpose, but only with the Syrian conflict did OSI start providing meaningful insight. The International, Impartial, and Independent Mechanism on Syria (IIIM) was established by the UN General Assembly (UNGA) in 2016 after vetoes in the UN Security Council prevented the referral of the Syrian situation to the International Criminal Court (ICC). The majority of evidence preserved by the IIIM is digital, and the IIIM works closely with civil society organizations to collect that evidence.

Similarly, the UN established the Independent Investigative Mechanism on Myanmar (IIMM) in 2018. Like the IIIM, the IIMM has relied heavily on digital evidence. Unlike the IIIM, the IIMM has made its struggle with social media content very public, excoriating Meta in particular for its role in the genocide of the Rohingya and calling on Meta to preserve evidence. According to the UN's IIMM's first report, Facebook was found to be "the internet" in Myanmar, with Myanmar officials being able to spread anti-Muslim and anti-Rohingya hate speech and disinformation.[10] Meta has voluntarily complied with some of the IIMM's requests, but the struggle to obtain further evidence has been well documented in The Republic of the Gambia v. Facebook, Inc.[11]

There have been several interesting developments for the UN. First, France recently passed legislation enabling "information to be transmitted from French courts" to the IIIM.[12] This could set a precedent for other national governments, including the United States. Second, legal experts from Oxford just released a report recommending that the UN provide permanent support for UN-mandated investigations, and presents two options: "the establishment of a standing, independent UN investigative support mechanism (ISM)" or "the establishment of a permanent investigative support division (ISD) within the Office of the High Commissioner for Human Rights (OHCHR)."[13] As noted in the recommendations section of this paper, either of these options would make it easier to craft due process respecting transfers of evidence from social media platforms to the UN. **If the standing mechanism were to become reality, it could make it much easier to ensure data is collected and stored in a timely, legal, and ethical manner that protects the safety and security of people whose PII is included in that data.**

Cases for accountability in Syria

As noted above, the use of data from online sources is not always readily apparent, but it has been a part of every case for accountability in Syria. In a historical first, in January 2022, the Koblenz Higher Regional Court in Germany convicted a senior Assad government official for crimes against humanity.[14] However, this was not the first Syria-related case. Syrian Archive has worked with other NGOs to file cases in several courts in France, Sweden, and Germany about the use of chemical

[10] United Nations Human Rights Council, "Report of Independent International Fact-Finding Mission on Myanmar," A/HRC/39/64, September 12, 2018, https://www.ohchr.org/en/hr-bodies/hrc/myanmar-ffm/reportofthe-myanmar-ffm.

[11] The Republic of Gam. v. Facebook, Inc., Civil Action 20-mc-36-JEB-ZMF (D.D.C. September 22, 2021).

[12] The Ministry for Europe and Foreign Affairs (France), "Jurisdiction of French courts over crimes against humanity," February 9, 2022, https://www.diplomatie.gouv.fr/en/french-foreign-policy/international-justice/news/article/jurisdiction-of-french-courts-over-crimes-against-humanity-9-feb-2022.

[13] Federica D'Alessandra et al., "Anchoring Accountability for Mass Atrocities: The Permanent Support Needed to Fulfil UN Investigative Mandates," The Oxford Institute for Ethics, Law, and Armed Conflict, May, 2022, https://www.bsg.ox.ac.uk/sites/default/files/2022-05/Anchoring%20Accountability%20for%20Mass%20Atrocities%20Report.pdf.

[14] European Center for Constitutional and Human Rights, "First criminal trial worldwide on torture in Syria before a German court," June 2, 2022, https://www.ecchr.eu/en/case/first-criminal-trial-worldwide-on-torture-in-syria-before-a-german-court/

weapons by the Syrian government, seeking to hold the government accountable.[15] Syrian Archive has been a party to every one of these cases, and the organization's extensive open-source chemical weapons database has provided indispensable evidence for each one.[16] Even before these cases were filed, evidence from Syrian Archive was successfully used in a 2019 prosecution of three Belgian firms that provided a sarin gas precursor to the government.[17] Open-source evidence has also been used in Swedish and German courts in several war crimes trials.

Ukraine

Like Syria, the escalation of Russian aggression toward Ukraine is being thoroughly documented on social media platforms. Unlike Syria, however, there has been a very rapid response from governments and international bodies interested in prosecuting war crimes, and an immediate acknowledgment of the importance of digital evidence. Civil society organizations have been able to take lessons learned about finding and archiving content – particularly from Syria, where Russian military and equipment carried out many attacks on civilian infrastructure – and apply those lessons to Ukraine. The civil society response has been almost immediate. Ukraine provides a clear example of the importance of content, as well as the dangers of deletion of that content.

There are myriad overlapping investigations, including an ICC investigation, a UN Commission of Inquiry, and investigations by the Ukrainian General Prosecutor.[18] In fact, the first trial for a war crimes case in Ukraine concluded after the Russian soldier on trial pled guilty.[19] The European Commission is also supporting ICC and creating own mechanism for storing content, and jurisdictions, and Germany's Federal Prosecutor has also started an investigation into war crimes.[20] Finally, on May 17, the United States Department of State announced the establishment of the Conflict Observatory, a program that will document, verify, preserve, analyze and share open-source evidence: "publicly and

[15] Syrian Center for Media and Freedom of Expression, "Survivors and the Syrian Center for Media and Freedom of Expression, with Support from Syrian Archive and the Justice initiative, Seek French Criminal investigations of Chemical Weapons Attacks in Syria," March 2, 2021, https://scm.bz/en/scm-statements/survivors-and-the-syrian-center-for-media-and-freedom-of-expression-with-support-from-syrian-archive-and-the-justice-initiative-seek-french-criminal-investigations-of-chemical-attacks-in-syria; Simon Johnson, "Victims of chemical attacks in Syria file complaint with Swedish police," Reuters, April 19, 2021, https://www.reuters.com/world/middle-east/victims-chemical-attacks-syria-file-complaint-with-swedish-police-2021-04-19/.

[16] Syrian Archive, "Chemical Weapons Database," May 18, 2022, https://syrianarchive.org/en/datasets/chemical-weapons.

[17] Syrian Archive, "Antwerp court convicts three Flemish firms for shipping 168 tonnes of isopropanol to Syria," February 7, 2019, https://syrianarchive.org/en/investigations/BI-sentencing; Jeff Deutch and Kristof Clerix, "Belgium illegally shipped 168 tonnes of sarin precursor to Syria," Syrian Archive, 2018, https://syrianarchive.org/en/investigations/belgium-isopropanol.

[18] International Criminal Court, "Ukraine: Situation in Ukraine," March 2, 2022, https://www.icc-cpi.int/ukraine; United Nations Human Rights Council, Forty-ninth session, "Situation of human rights in Ukraine stemming from the Russian aggression," A/HRC/RES/49/1, March 7, 2022, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/277/44/PDF/G2227744.pdf?OpenElement; Greg Myre, "Ukraine begins prosecuting Russians for war crimes," NPR, May 14, 2022, https://www.npr.org/2022/05/14/1098941080/ukraine-begins-prosecuting-russians-for-war-crimes.

[19] Associated Press, "Russian soldier pleads guilty at Ukraine war crimes trial," May 18, 2022, https://apnews.com/article/russia-ukraine-kyiv-moscow-war-crimes-61c89e6c73541f3fa2364dde1498df72.

[20] Bokan Pancevski, "Germany Opens Investigation Into Suspected Russian War Crimes in Ukraine," The Wall Street Journal, March 8, 2022, https://www.wsj.com/livecoverage/russia-ukraine-latest-news-2022-03-08/card/germany-opens-investigation-into-suspected-russian-war-crimes-in-ukraine-bNCphalWE30f2REH8BCi; Directorate-General for Neighbourhood and Enlargement Negotiations, "Russian war crimes in Ukraine: Commission welcomes European Parliament's adoption of Eurojust's reinforced mandate," European Commission, May 19, 2022, https://ec.europa.eu/neighbourhood-enlargement/news/russian-war-crimes-ukraine-commission-welcomes-european-parliaments-adoption-eurojusts-reinforced-2022-05-19_en.

commercially available information, including satellite imagery and information shared via social media."[21] Although the Conflict Observatory was created specifically to respond to Russian war crimes in Ukraine, it may provide a model for other conflicts.

Mnemonic and Bellingcat are two of the leading organizations working on archiving and verifying Ukrainian content in collaboration with Ukrainian organizations (including the 5 am coalition), and they have observed several trends.[22] First, content from Ukraine has been removed at a lower rate than Syria, for unfortunately obvious reasons – few of the participants in the conflict are designated terrorist organizations, content is not being posted in Arabic, and there is wide public and governmental support for Ukrainians. The presence of the Azov battalion, which is on Meta's DIO list, has had an impact on content. Meta had an exception for discussions of the battalion that the company "dialed back" after receiving negative press.[23] Graphic violence policies also certainly apply, but appear to have been interpreted more loosely than in previous conflicts. However, the official Russian state media channels that have been shut down are also very important sources of evidence for thorough investigations. The only platform that has publicly committed to preserving content is Meta, and not in an announcement but rather as a comment to the BBC.[24]

Meta's willingness to consider the importance of its platform in OSI is encouraging, but unfortunately Meta is not the primary source of evidence in this conflict. Content from Syria, Sudan, Yemen, and other conflicts has been on Facebook and Instagram, but also Twitter and YouTube. Yet in this conflict, Tik Tok has emerged as a very important source. Another significant development is the importance of Telegram. This is new and poses challenges for OSI for a variety of reasons (including TVEC), as channels on the platform are sometimes spaces for far-right organizing. The German government threatened to ban Telegram but finally met with Telegram representatives early this year.[25] Since then, Telegram has shut down at least 64 channels in Germany.[26] Investigators focused on Ukraine report entire channels suddenly gone with no warning, which could be related to government pressure. Telegram has been very resistant in engaging with civil society organizations (much less governments). As a platform-focused body, GIFCT could encourage the company to engage more.

The situation in Ukraine has prompted more interest in OSI and makes it much more likely that the policy solutions proposed in this paper, in particular legislative solutions, will be adopted.

[21] Office of the Spokesperson, "Promoting Accountability for War Crimes and Other Atrocities in Ukraine," United States Department of State, May 17, 2022, https://www.state.gov/promoting-accountability-for-war-crimes-and-other-atrocities-in-ukraine/.

[22] Alfred Landecker Foundation, "Mnemonic - Ukrainian Archive, Preservation, verification, and investigation of open-source documentation concerning human rights violations in Ukraine," May 19, 2022, https://www.alfredlandecker.org/en/article/introducing-mnemonic; Human Rights Centre ZMINA, "Ukraine 5 AM Coalition devoted to documenting war crimes is launched in Ukraine," March 15, 2022, https://zmina.ua/en/event-en/ukraine-5-am-coalition-devoted-to-documenting-war-crimes-is-launched-in-ukraine/; Eliot Higgins, "These are the Cluster Munitions Documented by Ukrainian Civilians," Bellingcat, March 11, 2022, https://www.bellingcat.com/news/rest-of-world/2022/03/11/these-are-the-cluster-munitions-documented-by-ukrainian-civilians/.

[23] Sam Biddle, "Facebook Allows Praise of Ukraine's Neo-Nazi Azov Battalion if It Fights Russian Invasion," The Intercept, February 24, 2022, https://theintercept.com/2022/02/24/ukraine-facebook-azov-battalion-russia/

[24] James Clayton, "Are tech companies removing evidence of war crimes?" BBC, March 31, 2022, https://www.bbc.com/news/technology-60911099.

[25] Reuters Staff, " Germany holds 'constructive' talks with Telegram, plans more," Reuters, February 4, 2022, https://www.reuters.com/article/germany-telegram-idUKKBN2K90Z9.

[26] Deutsche Welle, " Telegram blocks over 60 channels in Germany – report," February 12, 2022, https://p.dw.com/p/46uZT.

International Criminal Court

The ICC has been innovating around the use of OSI for several years now. In 2017, the ICC referred to Facebook videos in an arrest warrant for Mahmoud Mustafa Busayf Al-Werfalli.[27] This reference was groundbreaking for openly referring to social media content, but since then the ICC has continued to develop its approach to OSIs. In the case of The Prosecutor v. Ahmad Al Faqi Al Mahdi, the Court was presented "with a significant quantity of open-source evidence," including YouTube content.[28] In The Prosecutor v. Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu, and Narcisse Arido, "the Prosecution argued that the defendant had bribed a witness to change their testimony. In support of this allegation, the Prosecution submitted evidence of a wire transfer as well as pictures from Facebook showing the two witnesses alleged to have been bribed together."[29]

The ICC relies on OSIs because mutual legal assistance treaties (MLATs) and other formal tools to request evidence are largely not available to the Court. MLATs are agreements between two or more countries that "enable law enforcement authorities and prosecutors to obtain evidence, information, and testimony abroad in a form admissible in the courts of the Requesting State."[30] The ICC's rules of evidence are not as specific as (for example) the United States Federal Rules of Evidence, although these rules do help guide admission of evidence.[31] However, U.S. evidence rules would only allow the use of content obtained through OSI in very limited circumstances. In the ICC, evidence can be established through testimony in a more flexible way. The more markers of authenticity content has, the easier it is to use. This means that any level of preservation of and access to TVEC data could be helpful in prosecutions.

The SCA and GDPR

There are a variety of laws that apply to the retention and access to data. Unsurprisingly, due to platforms' geographical location, the most impactful of these laws are the United States Stored Communications Act (SCA), followed by the European General Data Protection Regulation (GDPR). The SCA is a federal law enacted as part of the Electronic Communications Privacy Act (ECPA) of 1986, an update on the Federal Wiretap Act of 1968. The SCA bans the disclosure of user data to third parties except under certain conditions. The SCA was updated by the 2018 CLOUD Act.

[27] Alexa Koenig, "Harnessing Social Media as Evidence of Grave International Crimes," Human Rights Center, Berkeley School of Law, October 23, 2017, https://medium.com/humanrightscenter/harnessing-social-media-as-evidence-of-grave-international-crimes-d7f3e86240d.

[28] Róisín A. Costello, "International criminal law and the role of non-state actors in preserving open source evidence," Cambridge Journal of International Law, December 1, 2018, 268–283.

[29] Id. (citing The Prosecutor v Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu and Narcisse Arido (Judgment) ICC-01/05-01/13 (19 October 2016).

[30] United States Department of Justice, "Mutual Legal Assistance Treaties of the United States" US DOJ, Criminal Division, Office of International Affairs, April, 2022, https://www.justice.gov/criminal-oia/file/1498806/download.

[31] Alexa Koenig and Nikita Mehandru, "Open Source Evidence and the Criminal Court," Harvard Human Rights Journal, April, 2019, https://harvardhrj.com/2019/04/open-source-evidence-and-the-international-criminal-court/; International Criminal Court, "Rules of Procedure and Evidence," 2013, https://www.icc-cpi.int/sites/default/files/RulesProcedureEvidenceEng.pdf.

For purposes of this paper, the relevant provision of the SCA is Section 2702, which allows platforms to disclose customer communications to a law enforcement agency "if the contents… appear to pertain to the commission of a crime" or "to a foreign government pursuant to an order from a foreign government that is subject to an executive agreement that the Attorney General has determined and certified to Congress satisfied section 2523."

The SCA regulates just two categories of internet service provider (ISP), reflective of the state of technology at the time the statute was passed: "electronic communication service" (ECS) or a "remote computing service" (RCS). ECS is defined as "any service which provides to users thereof the ability to send or receive wire or electronic communications," and RCS is defined by the statute as "the provision to the public of computer storage or processing services by means of an electronic communications system." However, these distinctions come into play in Section 2703 concerning "required disclosure of customer communications or records." In order to require an ECS provider to disclose content that is in temporary electronic storage for 180 days or less, the government needs a search warrant. For an ECS provider to disclose content in storage for more than 180 days, or to make this request of an RCS provider, the government has three options: a search warrant, subpoena, or court order.

It should be noted that the applicability of any of these provisions to publicly posted content is questionable. Therefore, the SCA largely seems to play a role in the question of preservation and access due to platforms' concerns about potential liability. **That being said, once a platform deletes user-generated content, whether it is posted publicly or privately, people no longer have the ability to change the privacy settings. Thus, regardless of the applicability of the GDPR platforms should treat even public posts with care.**

The European Union (EU) GDPR came into effect in 2018. The GDPR regulates the collection and processing of personal data by organizations and companies like technology companies as well as government agencies. It requires that companies that process data ensure there are specific purposes for doing so and make this clear to individuals when collecting their data (known as purpose limitation). Companies must also establish time limits to erase or review the data stored (storage limitation) and respond to requests from individuals exercising their data protection rights free of charge within 1 month of receipt. Under the GDPR, data subjects have the following privacy rights: to be informed; have access; seek rectification; restrict processing; ensure erasure and data portability; object; and exercise their rights in relation to automated decision-making and profiling. To be protected under the GDPR, a data subject has to be either a citizen of the EU or simply located in the EU, regardless of national identity.

The GDPR could apply to the preservation and access of relevant data. Currently, platforms' terms of service refer only to law enforcement investigations and not the other potential uses of data considered by this paper. It should be noted that personal or household activities, national security, and law enforcement are exempt from the EU GDPR. As noted, the Swedish and German courts that used open-source data to prosecute war crimes did not run into GDPR issues. However, more analysis is needed to determine whether and how the GDPR might apply to data shared with the ICC and United Nations.

High-level principles for solutions

Discussions about the need to solve the problem of disappearing human rights content have been going on for many years now, but thanks to the visibility of social media's role in Myanmar and the increasingly obvious utility of this content, policymakers are taking this issue seriously.

Many groups have been working on this topic. One is an informal coalition composed of Human Rights Watch, WITNESS, Mnemonic, and the Berkeley Human Rights Center. Starting in January 2020, this group hosted a series of workshops to understand the views of various stakeholders. These workshops included a wide range of participants, including lawyers, content moderation experts, representatives of community archives from conflict zones, and privacy experts. The goal was not to propose one specific solution, but rather to consider the principles solutions ought to follow.

In the end, the informal coalition gathered at these workshops reached a consensus position: while there is no one size fits all solution, there are a few things that should be considered in any solution. The most pressing principle to consider is the tension between privacy/security and the purpose of preserving and providing access to data. This is reflected in the principle of necessity and proportionality in international human rights law.

There was no consensus among workshop attendees about how to articulate principles for solutions. However, many of the resources referenced in these workshops inform this paper. In addition to concerns about privacy and security, there were a few other overarching concerns, spanning responsible retention and stewardship of data (including transparency), holding data for limited and clearly defined purposes, and having methods to ensure remedies for parties negatively impacted by solutions. One seldom-noted principle that should be recognized and implemented is the idea that impacted communities should be centered in discussions about possible solutions.[32]

As noted above, the tension between privacy and security and access to justice is particularly important to consider. While it may be easy to simply ask platforms to preserve data, there are always potential harms in preserving and providing access to PII (regardless of whether it was originally posted in a public or private forum).

PII is data that could be used to identify an individual. The GDPR provides a very helpful definition: "personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic,

[32] See, eg, Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M., "The CARE Principles for Indigenous Data Governance." Data Science Journal, November 4, 2020, 19(1), p.43. DOI: http://doi.org/10.5334/dsj-2020-043; Society of American Archivists, "SAA Core Values Statement and Code of Ethics." Approved by the Society of American Archivists, last revised August 2020, https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

cultural or social identity of that natural person."[33]

The vast majority of people whose PII is included in content are not only not accused of any crime but are uniquely vulnerable. Their privacy, security, and even safety can be impacted by the retention of data. Solutions should weigh these potential harms against how the data will be used and who it benefits. PII that is stored by companies should be considered accessible by governments and law enforcement agencies, including those government actors that might misuse the data, or use it in service of human rights violations. This is not hypothetical when it comes to social media data. For example, in 2017 the Egyptian government ramped up its targeting of LGBTQ people after images of attendees waving a rainbow flag at a Mashrou Leila circulated on social media. The Egyptian government has a unit dedicated to arresting and prosecuting LGBTQ people that uses social media content as evidence.[34]

These concerns are in no way limited to authoritarian governments. The United States and the EU have also unfairly targeted vulnerable communities using social media posts and associated data. For example, under President Donald Trump, non-immigrant visa applications were updated to ask applicants for their social media handles, in service of Trump's colloquially named "Muslim Ban."[35] In a case challenging this requirement, plaintiffs said "The Registration Requirement is the cornerstone of a far-reaching digital surveillance regime that enables the U.S. government to monitor visa applicants' constitutionally protected speech and associations not just at the time they apply for visas, but even after they enter the United States."[36] It was expected that President Joe Biden would do away with this practice, but instead his administration has recommended expanding it.[37] In addition to the U.S. government's use of social media, myriad federal agencies use social media monitoring in ways that have led to false arrests and other human rights violations.[38] Providing these agencies with a vast store of data should not be the end result of attempts to preserve human rights documentation and other important data.

European law enforcement agency handling of data raises similar concerns. In January of this year, the European Data Protection Supervisor (EDPS) ordered Europol "to delete data concerning

[33] Article(1), REGULATION (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[34] Declan Walsh, "Egyptian Concertgoers Wave a Flag, and Land in Jail," New York Times, September 27, 2017, https://www.nytimes.com/2017/09/26/world/middleeast/egypt-mashrou-leila-gays-concert.html.

[35] United States Department of State, "Frequently Asked Questions on Social Media Identifiers in the DS-160 and DS-260," June 4, 2019, https://travel.state.gov/content/dam/visas/Enhanced%20Vetting/CA%20-%20FAQs%20on%20Social%20Media%20Collection%20-%206-4-2019%20(v.2).pdf.

[36] Doc Society et al v. Pompeo, COMPLAINT FOR DECLARATORY AND INJUNCTIVE RELIEF, filed December 12, 2019, https://www.brennancenter.org/sites/default/files/2019-12/Complaint%20Doc%20Society%20v%20Pompeo.pdf.

[37] Anna Diakun and Carrie DeCell, "Why is the U.S. still probing foreign visitors' social media accounts?," Washington Post, April 26, 2022, https://www.washingtonpost.com/outlook/2022/04/26/social-media-surveillance-us-visas-state/.

[38] Rachel Levinson-Waldman, Harsha Panduranga and Faiza Patel, "Social Media Surveillance by the U.S. Government," Brennan Center for Justice, February 7, 2022, https://www.brennancenter.org/our-work/research-reports/social-media-surveillance-us-government.

individuals with no established link to a criminal activity (Data Subject Categorisation)."[39] Sensitive data was "sampled from asylum seekers never involved in any crime."[40]

A note on data related to attacks with an online component

The livestreamed March 2019 attack on mosques in Christchurch, New Zealand and the livestreamed May 2022 attack on a grocery store in a predominantly Black neighborhood of Buffalo, New York exemplify the kinds of incidents that GIFCT and the Christchurch Call to Eradicate Terrorist and Violent Extremist Content Online are meant to respond to. GIFCT and New Zealand government officials have created the Content Incident Protocol (CIP) and the Crisis Response Protocol (respectively) for incidents such as these. Both of these protocols are meant to allow situational information sharing, and the CIP allows companies to share hashes of content created by perpetrators or their accomplices (the CIP in particular is limited in scope to this perpetrator content). What the CIP does not necessarily do is ensure that governments have access to technical information about this content that could help them understand the online component of these attacks.

After the Christchurch shooting, the New Zealand government wanted access to information such as how many times the video had been viewed and the location of users who viewed it. They were not able to obtain that information. It is essential that governments not bypass due process requirements for access to data and further research is needed to address existing relevant legal frameworks for sharing such information. It should be noted that even non-PII could help aid in such post-incident investigations. There seem to be technical limitations to how much information companies can provide, as well as confusion within companies about who should be providing this information. The CIP, as well as the related Christchurch Call Crisis Response Protocol, helps to define points of contact. However, this information is not sufficient for large companies like Meta and Google – these companies need to determine how to provide non-PII and who should be responsible for the necessary steps. GIFCT should aid in this by ensuring that due process requirements and applicable legal frameworks are clarified and made available to small companies.

Company efforts: Twitter's research consortium

It is worth noting that some platforms are already providing more data to researchers than others. Twitter stands out as one of the platforms with the most accessible data. The company has always been notable in providing meaningful API access, but it also has an academic researcher program that allows greater access. Furthermore, in December 2021, the company announced that it would be launching the Twitter Moderation Research Consortium to "provide comprehensive data about attributed platform manipulation campaigns to members of the consortium, who may independently choose to publish their findings on the basis of the data we share and their own research."[41] The

[39] European Data Protection Supervisor, "EDPS orders Europol to erase data concerning individuals with no established link to a criminal activity," January 10, 2022, https://edps.europa.eu/press-publications/press-news/press-releases/2022/edps-orders-europol-erase-data-concerning_en.

[40] Apostoli Fotiadis et al., "'A data 'black hole': Europol ordered to delete vast store of personal data," The Guardian, January 10, 2022, https://www.theguardian.com/world/2022/jan/10/a-data-black-hole-europol-ordered-to-delete-vast-store-of-personal-data.

[41] Gadde Vijaya and Yoel Roth, "Expanding access beyond information operations," Twitter Blog, June 7, 2022, https://blog.twitter.com/en_us/topics/company/2021/-expanding-access-beyond-information-operations-.

consortium has just launched and is available to a limited number of academic researchers. **This** effectively excludes archives like Mnemonic, **and hopefully Twitter will determine standards that do allow broader access**. However, as it grows, Twitter's work in defining standards for access and protection of users could help to inform further GIFCT research on this topic.

What we know about how companies handle data

In the course of conducting research for this paper, a survey was sent to social media companies asking them to explain how they determined where GDPR and SCA applied to the data they hold, what their data retention and deletion practices are, and what their law enforcement request procedures are. Unfortunately, only one company provided a response, leaving us to guess how exactly GIFCT member companies are handling data.

It is impossible to get a clear picture of how all GIFCT member companies handle data because most of them do not provide specific time frames for how long it takes them to delete data-including backup copies after a user deletes it, nor how companies determine what laws apply to specific users (in particular the GDPR).

However, we were able to find some information from public posts and policies about how member companies handle data. Currently, it appears that they have few limitations on storing data. Meta says, "We store data until it is no longer necessary to provide our services and Meta Products, or until your account is deleted – whichever comes first.... When you delete your account, we delete things you have posted, such as your photos and status updates, and you won't be able to recover that information later."[42] It does not provide any specific time frame. Twitter's policy states, "We keep your profile information and content for the duration of your account. We generally keep other personally identifiable data we collect when you use our products and services for a maximum of 18 months." Twitter's policy explains that a user's account information will be held for up to 30 days, and that "[w]here you violate our Rules and your account is suspended, we may keep the identifiers you used to create the account (i.e., email address or phone number) indefinitely to prevent repeat policy offenders from creating new accounts."[43] Google's privacy policy says that they delete content when it is deleted by users, but that the whole process "generally takes around 2 months from the time of deletion," and that data on encrypted backup servers "can remain on these systems for up to 6 months."[44] Links to other GIFCT member privacy policies can be found in the footnotes, but the pattern is clear: these policies do not address how platforms manage data they took down themselves. While it is clear that data is not deleted instantly, the policies leave significant room for platforms to selectively preserve content where appropriate – **keeping in mind that once a platform deletes content, users no longer have the option to change privacy settings.**

No platform has publicly committed to retaining specific types of data after the activation of a CIP,

[42] Meta, "Data Policy," January 4, 2022, https://www.facebook.com/about/privacy/update. Note that the policy is labelled differently for the Meta Platforms Ireland Limited, which processes data for European Union residents, and Meta Platforms Inc. This information is under "Data retention, account deactivation and deletion" for Meta Ireland and "How can I manage or delete information about me?" for Meta.

[43] Twitter, "Privacy Policy," June 10, 2022, https://twitter.com/en/privacy.

[44] Google, "How Google retains the data we collect," June 10, 2022, https://policies.google.com/technologies/retention.

nor in times of crisis where it is widely known that human rights documentation is amassing online (e.g., in Ukraine, but also Iran, Palestine, Sudan, and other crises). **Considering the legal, privacy and security implications raised by storing user data this is understandable, but it's clear that a solution is needed as soon as possible.**

Recommendations

Interviews and literature review during the course of this research reinforced the conclusion that it would be a mistake at this time to simply ask platforms to retain all data or hand it over to most non-governmental parties. Instead, there are a number of limited steps that should be taken, and further research undertaken.

First, there is one problem that has a clear solution: the need for a mechanism to allow the ICC and UN investigative bodies to request data. The most likely form this mechanism would come in is a very limited exception to the SCA disclosure limitations. This would likely have to be a direct amendment to the SCA, and it should be written in the narrowest possible way. It should provide an exception only for content relevant to a limited set of international human rights law violations, including war crimes, crimes against humanity, and genocide. Scholars at Yale and Boston College have proposed just such a solution.[45] However, due to the massive privacy implications of tampering with the SCA, this paper suggests that any amendment should be written in the most narrow way possible, and should comport with due process requirements. It could require the ICC and UN to apply with a magistrate judge in line with Rule 41 of Federal Rules of Criminal Procedure.[46] It could establish standards for a data request that conform as much as possible to 4th Amendment warrant requirements, which require applications for warrants to be justified by probable cause, supported by oath or affirmation, and in particular describe the place to be searched and the persons or things to be seized. Additionally, where necessary, platforms should refine their policies to clarify that data could be used for the purpose of prosecuting the same limited set of international crimes in order to comply with the GDPR. Further steps may be needed to comply with the GDPR, and GDPR experts should be consulted.

Further research is needed to determine exactly what gaps exist for governments to request data from platforms about content after a terrorist attack with an online component. The purpose of the GIFCT CIP and the Christchurch Crisis Response Protocol is partly to enable information sharing, but it should not be done in a way that bypasses due process. Some of the metadata governments are interested in is not necessarily PII: for example, platforms could provide governments with information about the locations of accounts or the number of times a piece of content was viewed without implicating privacy concerns. GIFCT is well-positioned to address this topic, and it should be the focus of the working group next year.

[45] Joshua Lam and David Simon, "To Support Accountability for Atrocities, Fix U.S. Law on the Sharing of Digital Evidence," Just Security, April 20, 2022, https://www.justsecurity.org/81182/to-support-accountability-for-atrocities-fix-u-s-law-on-the-sharing-of-digitial-evidence/; Rebecca J. Hamilton, "Platform-Enabled Crimes," Boston College Law Review, November 12, 2021, https://ssrn.com/abstract=3905351 or http://dx.doi.org/10.2139/ssrn.3905351.

[46] FRCP 41 "Search and Seizure," https://www.law.cornell.edu/rules/frcrmp.

Similarly, further investigation and analysis is needed to determine how to provide increased civil society access to data, either by directly providing copies of user-generated content depicting human rights violations or providing data for other research purposes (such as understanding the spread of misinformation). Twitter's experience could be very valuable here.

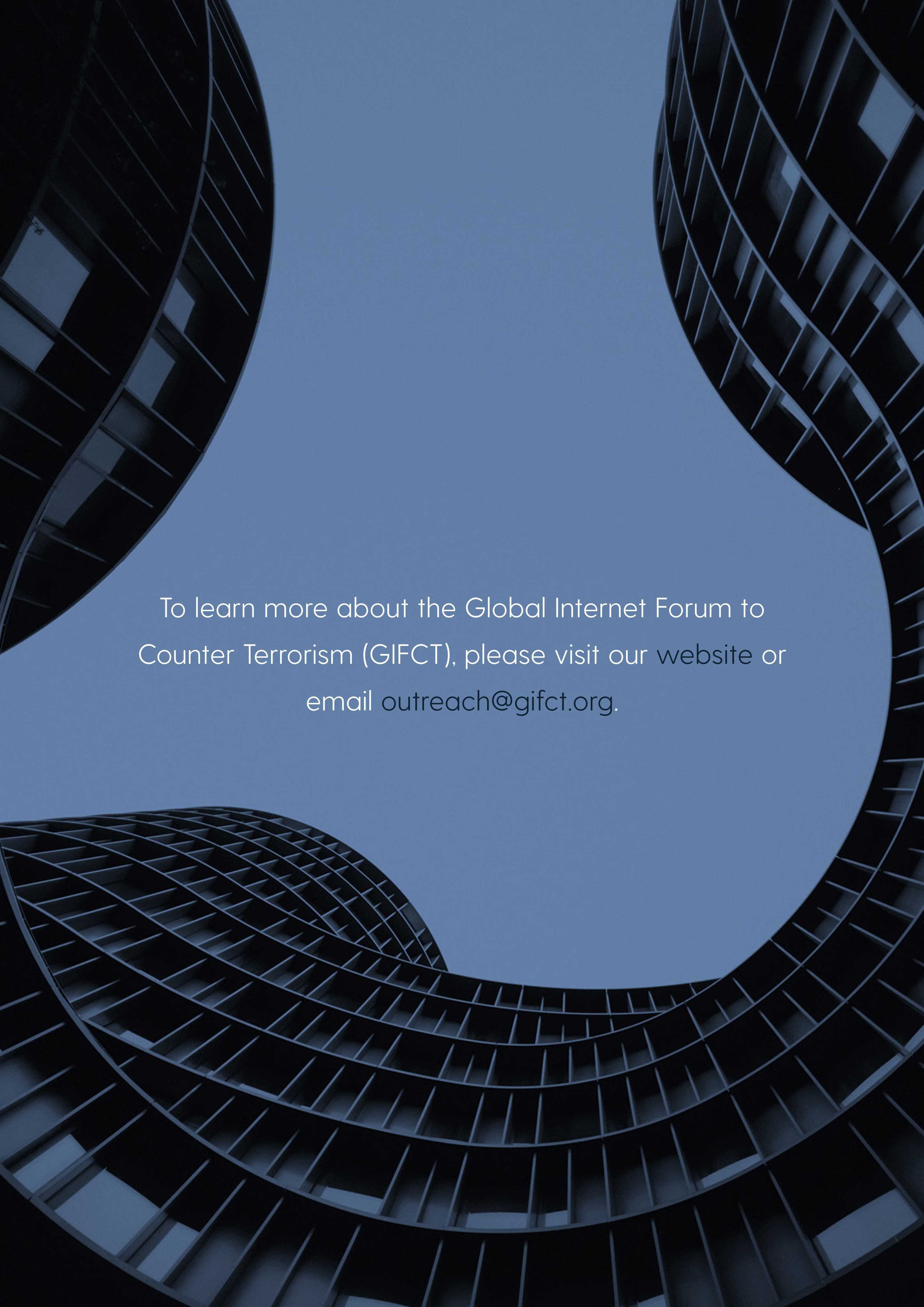
Finally, this research reinforced the need for increased transparency from platforms in their privacy policies or terms and conditions. Platforms should clarify how long to retain data when a user has deleted it, and explain exactly how long they handle data from content that they themselves have removed. They should also explain with more clarity what legal frameworks they apply to what users. Broadly, they should also commit (or recommit) to the Santa Clara Principles On Transparency and Accountability in Content Moderation, **which provides in-depth and operationalizable standards for transparency from platforms**.[47]

Ultimately, it is clear that OSIs are a compelling necessity that should be addressed as soon as possible. Similarly, in the wake of the Buffalo shooting, it is clear that more information about how perpetrator content travels in the media ecosystem is necessary to understand the online aspect of such violent attacks. These are achievable goals that should be prioritized by civil society organizations, governments, and companies.

Interview list

Hadi al Khatib (Syrian Archive/Mnemonic), David Shanks (former), Lindsay Freeman (UC Berkeley Human Rights Center), Nick Waters (Bellingcat) , Aaron Zelin (Jihadology/Brandeis), Sun Kim (IIMM but interviewing in her personal capacity), Yvonne McDermot (Swansea), Nathaniel Raymond (Yale), Libby McAvoy (Mnemonic), Anonymous activists

[47] The Santa Clara Principles on Transparency and Accountability in Content Moderation, 2021, https://santaclaraprinciples.org/.



To learn more about the Global Internet Forum to Counter Terrorism (GIFCT), please visit our website or email outreach@gifct.org.