

HUMAN  
RIGHTS  
CENTER  
UC Berkeley School of Law



# DIGITAL LOCKERS

Archiving Social Media Evidence  
of Atrocity Crimes



**HUMAN  
RIGHTS  
CENTER**

UC Berkeley School of Law

# **DIGITAL LOCKERS**

Archiving Social Media Evidence  
of Atrocity Crimes

**2021**

## HUMAN RIGHTS CENTER

The Human Rights Center at the University of California, Berkeley, School of Law conducts research on war crimes and other serious violations of international humanitarian law and human rights. Using evidence-based research methods and innovative technologies, we support efforts to hold perpetrators accountable and to protect vulnerable populations. We also train students and advocates to research, investigate, and document human rights violations and turn this information into effective action.

2224 Piedmont Avenue, Berkeley, CA 94720

Telephone: 510.642.0965 | Email: [hrc@berkeley.edu](mailto:hrc@berkeley.edu)

[Humanrights.berkeley.edu](https://humanrights.berkeley.edu) | [Medium.com/humanrightscenter](https://medium.com/humanrightscenter) | [@HRCBerkeley](https://twitter.com/HRCBerkeley)

FRONT COVER PHOTO: A man films a protest against President Sebastián Piñera's government and police brutality in Santiago, Chile, on February 12, 2021. (Photo by Vanessa Rubilar /SOPA Images/Sipa USA)(Sipa via AP Images).

DESIGN AND LAYOUT: Nicole Hayward

# CONTENTS

**INTRODUCTION / 2**

**RESEARCH QUESTIONS / 5**

**METHODOLOGY / 6**

**BACKGROUND / 7**

PART I – THE STAKEHOLDERS / 7

Stakeholders / 7

Stakeholder Relationships / 8

PART II – TYPOLOGY OF DIGITAL ARCHIVES / 10

Social Media Platforms as “Accidental Archives” / 10

Traditional Archives / 11

Digital Archives / 15

Model 1: The Legal Compulsion Model / 17

Model 2: The Voluntary Partnership Model / 23

Model 3: The Independent Collection Model / 29

Model 4: The Hybrid Model / 37

PART III – LEGAL, TECHNICAL, AND OPERATIONAL CHALLENGES / 41

Defining Terms and Scope / 41

Legal Compliance / 43

Automated Detection of Graphic Content / 46

**DISCUSSION / 48**

**RECOMMENDATIONS / 51**

**CONCLUSION / 53**

**ACKNOWLEDGMENTS / 54**

# INTRODUCTION

Given the use of social media by people living in areas of armed conflict or severe repression, social media platforms have become accidental and unstable archives for human rights content.<sup>1</sup> The last two decades have witnessed a fundamental shift in how people around the world communicate. During this period, the proliferation of smartphones and the rise of social media platforms have enabled increased identification, collection, and sharing of digital information related to international crimes and human rights violations. Whereas human rights researchers once struggled to find online content relevant to their investigations, today researchers may find themselves drowned in a tsunami of content with potential evidentiary value,<sup>2</sup> as well as utility for the documentation of atrocities more generally—including for advocacy, research, and development of an historical record of world events. With 6,000 tweets generated every second<sup>3</sup> and 500 hours of video content uploaded to YouTube every

minute,<sup>4</sup> the challenge is figuring out how to find the “signal” by siphoning out the online “noise,” as well as how to find reliable information buried in a digital environment replete with misinformation and disinformation.<sup>5</sup>

While human rights researchers and investigators have been pioneering new methods for mining online environments for reliable information,<sup>6</sup> they have frequently found themselves in a race against platforms’ efforts to police their websites.<sup>7</sup> Survivors and bystanders often post videos and images to social media platforms with the hope of alerting the

---

1 “Removals of Syrian Human Rights Content: May 2019,” Syrian Archive, accessed May, 2021, <https://syrianarchive.org/en/tech-advocacy/may-takedowns.html>.

2 By evidentiary value, we mean information that may be used to help establish the facts necessary to satisfy the elements of crimes or other legal violations in a court of law.

3 David Sayce, “The Number of Tweets per Day in 2020.” David Sayce, December, 2019, <https://www.dsayce.com/social-media/tweets-day/>.

---

4 “YouTube for Press.” Blog.YouTube, accessed January, 2021, <https://blog.youtube/press/>.

5 For a helpful overview of various forms of “information disorder,” see, Claire Wardle, “Understanding Information Disorder,” FirstDraft, October, 2019, [https://firstdraftnews.org/wp-content/uploads/2019/10/Information\\_Disorder\\_Digital\\_AW.pdf?x76701](https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x76701).

6 Paul Meyers, “How to Conduct Discovery Using Open Source Methods,” in *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability*, eds. Sam Dubberley, Alexa Koenig and Daragh Murray. (New York: Oxford University Press, 2020), 168-199.

7 Dipayan Ghosh, “Are We Entering a New Era of Social Media Regulation?,” Harvard Business Review, January, 2021, <https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation>. See also, Chloe Mathieu Phillips, “Regulating social media: legislation or self-policing?,” The Social Element, November, 2018, <https://thesocialelement.agency/regulating-social-media-legislation-or-self-policing>

world to atrocities on the ground,<sup>8</sup> yet companies' terms, conditions, and community guidelines prohibit an important subset of this content, such as violent, graphic or sexually explicit imagery.<sup>9</sup> In recent years, these platforms have increased their use of automated tools to detect and remove content that violates their terms of service at a rate that outpaces human investigators.<sup>10</sup> For example, between July

and September 2020, Facebook's algorithm detected and "actioned"<sup>11</sup> 99.5 percent of violent and graphic content before users reported such content.<sup>12</sup> This pace of detection means that human rights actors are increasingly losing the race to identify and preserve information that may have legitimate human rights and historical value before it is removed.<sup>13</sup>

Companies have important reasons for removing certain categories of social media content; for example, propaganda from internationally recognized terrorist groups or material that sexually exploits children. Many platforms' terms of service reflect concerns about the privacy and security of platform users, negative user experience, as well as their own legal liability.<sup>14</sup> However, content removals remain

---

8 Sharngan Aravindakshan and Radhika Kapoor, "The Potential and Hurdles of Fighting Atrocities in the Age of Social Media," *The Wire*, April, 2020, <https://thewire.in/tech/social-media-atrocities-evidence>. See also Belkis Wille, "'Video Unavailable': Social Media Platforms Remove Evidence of War Crimes," *Human Rights Watch*, September, 2020, <https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes>.

9 "Community Standards," Facebook, accessed March, 2021, <https://www.facebook.com/communitystandards/introduction>; "Your commitments to Facebook and our community," Facebook Terms of Service, accessed March, 2021, <https://www.facebook.com/terms.php>; "Community Standards Enforcement Report," Facebook Transparency, February, 2021, <https://transparency.facebook.com/community-standards-enforcement>; "The Twitter Rules," Twitter, accessed March, 2021, <https://help.twitter.com/en/rules-and-policies/twitter-rules>; "Transparency," Twitter, accessed March, 2021, <https://transparency.twitter.com>.

10 For Twitter see, "An update on our continuity strategy during COVID-19," Twitter Blog, March, 2020, [https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html); "Insights from the 17th Twitter Transparency Report," Twitter Blog, January, 2021, [https://blog.twitter.com/en\\_us/topics/company/2020/ttr-17.html](https://blog.twitter.com/en_us/topics/company/2020/ttr-17.html); Alyssa Newcomb, "Twitter Says A.I. Is Removing Over Half of the Site's Abusive Tweets Before They're Flagged," *Fortune*, October, 2019, <https://fortune.com/2019/10/24/twitter-abuse-tweets/>. For Facebook see, Jeff King and Kate Gotimer, "How We Review Content," Facebook Blog, August, 2020, <https://about.fb.com/news/2020/08/how-we-review-content/>; James Vincent, "Facebook is now using AI to sort content for quicker moderation," *The Verge*, November, 2020, <https://www.theverge.com/2020/11/13/21562596/facebook-ai-moderation>. For Instagram see, Jacob Kastrenakes, "Instagram now uses AI to block offensive comments," *The Verge*, June, 2019, <https://www.theverge.com/2017/6/29/15892802/instagram-ai-offensive-comment-filter-launches>. For YouTube see, "Protecting

---

our extended workforce and the community," YouTube Creator Blog, March, 2020, <https://youtube-creators.googleblog.com/2020/03/protecting-our-extended-workforce-and.html?m=1>; "Featured Policies: Violent Extremism chapter," Google Transparency Report, January 2020–March 2020, <https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism?hl=en>.

11 Actioned content includes material on Facebook and Facebook Messenger that was covered with a warning label or removed from the platforms.

12 "Community Standards Enforcement Report," Facebook Transparency, accessed January, 2021, <https://transparency.facebook.com/community-standards-enforcement#graphic-violence>

13 See, e.g., Wille, "'Video Unavailable': Social Media Platforms Remove Evidence of War Crimes."; Alexa Koenig, "Big Tech Can Help Bring War Criminals to Justice," *Foreign Affairs*, November, 2020, <https://www.foreignaffairs.com/articles/united-states/2020-11-11/big-tech-can-help-bring-war-criminals-justice>.

14 For more information on the type of content subject to removal from Facebook see, "Community Standards," Facebook, accessed January, 2021, <https://www.facebook.com/communitystandards/introduction>. See also, "The Twitter Rules," Twitter, accessed January, 2021, <https://help.twitter.com/en/rules-and-policies/twitter-rules>. For more information on Twitter's rules on content in violation that are subject to removal. See, "YouTube Community Guidelines & Policies—How YouTube Works," YouTube Community Guidelines & Policies, accessed January, 2021, <https://www.youtube.com/howyoutubeworks/>

of deep concern to human rights researchers, legal investigators, and historians, who recognize that the information posted to social media may include critical data for proving the elements of crimes and preventing further abuse—and in some cases may be the only documentation of such events.

This report provides an overview of various models that have previously been used to archive social media content and other digital information. Part I identifies key **stakeholders**—their missions, values, and interest in the preservation and accessibility of social media content. Part II establishes a **typology of archives** that have been used as repositories of online digital content. We include at least one case study for each model, and discuss several legal and operational considerations that stakeholders may want to assess when designing one or more ways forward. Each case study is followed by a brief summary highlighting who provides, holds, and

---

policies/community-guidelines/ (guidelines on platforms, videos, thumbnails, and links that are subject to removal).

can access the content, as well as a brief overview of legal obligations, challenges, and end-uses. Part III provides a summary of the broader **legal and technical context** that surrounds the debate around what form an evidence locker or other human rights archive should take, drawing on national and international norms to aid future conversations. We conclude with **recommendations** for next steps.

Importantly, this report does not identify or advocate for a specific model or sketch a single way forward, but simply illustrates some of the ways that previous digital archives have been constructed. It is our hope that this report will advance an ongoing and longstanding conversation among human rights organizations, social media companies, diverse government actors, researchers and others, and inform a collaborative and multidisciplinary effort for ensuring that the preservation of online information with human rights value can more effectively serve the goals of legal accountability and justice—as well as the diverse needs of affected communities worldwide.



# RESEARCH QUESTIONS

Our research was driven by the following questions:

1. What models for archiving digital information—and especially social media content—already exist?
2. How are these models structured, funded, and managed?
3. What lessons can we learn from these models about challenges to and opportunities for preserving and archiving online content for evidentiary and other human rights purposes?
4. What legal, political, technical, financial, and operational challenges are likely to arise in the creation of a digital evidence locker or new legal framework—including how to prospectively identify what subset of online information needs saving?

# METHODOLOGY

From January 2020 through June 2020, nine students from diverse departments on the UC Berkeley campus<sup>15</sup> worked with researchers at the Human Rights Center at UC Berkeley School of Law to identify and analyze various social media repositories that might

---

<sup>15</sup> The students came from Global Studies, Public Policy, Development Practice, Information Management, Advanced Law (LLM), Media Studies, Interdisciplinary Studies, Political Science and a multidisciplinary international program.

offer valuable insights into opportunities and challenges associated with creating a digital archive—or “evidence locker”—for social media content at risk of deletion. The team identified and analyzed precedents from disparate but related contexts, such as terrorism and human trafficking, and conducted interviews to fill gaps in the desk research. A second team of center staff conducted supplemental research and edited the report between October 2020 and May 2021.

# BACKGROUND

## Part I – The Stakeholders

### STAKEHOLDERS

The following are key stakeholders who have an interest in the preservation and disclosure of social media content related to human rights violations:

- 1. Content creators, subjects and users:** Individuals who create and/or upload human rights content to social media platforms—as well as those depicted in that content—have an interest in what happens with the material. Importantly, the privacy rights of these individuals are likely to be implicated in any sort of digital locker, as are their interests in legal accountability, advocacy, and creating a record of events.
- 2. Social media companies:** Technology companies that operate social media platforms are third-party intermediaries that have an interest in what happens with their users' data. These companies must comply with national and international laws, especially those related to data protection and privacy, as well as financial, ethical, and operational constraints. These companies have diverse and sometimes conflicting obligations to their users and their shareholders. Any type of legal mechanism, archive, or digital locker would likely create obligations for social media companies to preserve human rights content and potentially be responsible for its long-term storage and sharing.
- 3. Inter-governmental organizations:** Inter-governmental organizations (IOs) such as United Nations bodies and international courts are mandated to investigate violations of international law. There are several UN investigative mechanisms, fact-finding missions, and commissions of inquiry that investigate violations of international human rights and humanitarian law. In addition, international criminal courts and tribunals investigate violations of international criminal and humanitarian law. These entities have an interest in obtaining user-generated content that can serve as intelligence, lead information, or evidence related to their investigations.
- 4. Non-governmental organizations:** As the largest and most diverse stakeholder group, non-governmental organizations (NGOs) are interested in the preservation of and access to user-generated content on social media platforms for use in human rights documentation, advocacy, research, and reporting. In some cases, NGOs might also be interested in preserving content as potential evidence for future accountability processes. Any sort of mechanism or legal framework that might serve NGOs will need to clearly define which categories of NGOs qualify as human rights NGOs for purposes of these efforts, since the general category of NGO is very broad.

**5. Academic researchers:** Social science and other academic researchers have an interest in the preservation of and access to human rights content for the purpose of study and establishing an historical record. Such research might be conducted for private or public uses.

For ease of reference, the term “human rights practitioners” is used to collectively refer to IOs, NGOs, and academic researchers, although their interests may vary or even conflict. For example, some individuals and organizations are interested in the content staying on the platform in public view, some are interested in it being preserved but not necessarily made public, and some are interested in having it preserved and shared for a range of end uses.

## STAKEHOLDER RELATIONSHIPS

Social media users and human rights practitioners have repeatedly reported their dissatisfaction and concern with technology companies removing human rights-related content from public access without some mechanism for preserving that content outside national law enforcement processes.<sup>16</sup> From removing critical documentation of some of the world’s worst atrocities from the public record to silencing the voices of survivors, these takedowns may distort the information ecosystem in ways that enhance impunity for perpetrators and minimize the possibilities of justice for some of the world’s most egregious crimes—for example, when the original or only video or posting of an event is caught in the dragnet.<sup>17</sup>

---

<sup>16</sup> Wille, “‘Video Unavailable’: Social Media Platforms Remove Evidence of War Crimes.”

<sup>17</sup> Ibid. One example would be videos of extrajudicial killings in Libya allegedly perpetrated by al-Werfalli, for whom an arrest warrant was issued by the ICC. In that case, several videos were removed by Facebook from its platform, but preserved by Bellingcat prior to removal.

At present, content creators and human rights practitioners work with social media companies to address problematic takedowns, but the process is largely informal, inconsistent, and ad hoc. Content creators and human rights practitioners complain that they are often unable to successfully appeal the removal of human rights content by social media companies.<sup>18</sup> Social media companies complain that creators and practitioners don’t fully understand or appreciate the legal and operational constraints within which they’re working. Regardless, current practice is unsustainable. NGOs often feel as though they are functioning as unpaid content moderators for some of the most well-resourced companies in the world and inequities are created among NGOs, some of which have personal relationships with the social media company contacts and some who do not. From their perspective, all too often relevant human rights content is removed and made permanently inaccessible.

Adding to earlier challenges, the rate of content removal has accelerated as a result of replacing human content moderators with algorithms that automate portions of the detection process. The incorporation of machine learning has further increased the pace at which content disappears from public view. While human rights practitioners may scrape,<sup>19</sup>

---

<sup>18</sup> According to Human Rights Watch, “Users [on Facebook between January and March 2020] appealed takedowns for 180,100 pieces of “terrorist propaganda” content, 479,700 pieces of “graphic violence” content, 1.3 million pieces of “hate speech” content, and 232,900 pieces of “organized hate” content. Upon appeal, Facebook restored access to 22,900 pieces of “terrorist propaganda” content, 119,500 pieces of “graphic violence” content, 63,600 pieces of “hate speech” content, and 57,300 pieces of “organized hate” content.” See, Ibid.

<sup>19</sup> “Web scraping is a process in which machine readable data is extracted from the HTML lay-out of websites delivered to a user’s browser and storing that data locally. This process takes data that is meant for display on a user’s device and converts it into a format that can be processed. These often require more development time and include less contextual metadata about each unit, but still make efficient ingestion of large quantities

manually download, or otherwise preserve relevant content before its removal, this approach leads to the siloing of content across different groups, making its identification and location by potential end users extremely difficult.<sup>20</sup> From a legal perspective, even when preserved by external parties, the evidentiary value of such information may be diluted,

as international judges typically require that social media content be provided directly by the companies when used for court purposes to help ensure its authenticity. Over the past couple of years, social media companies have begun sharing information with various national and intergovernmental institutions in order to support international justice processes.

---

of content possible.” Jeff Deutch and Niko Para, “Targeted Mass Archiving of Open Source Information,” in *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation and Accountability*, eds. Sam Dubberley, Alexa Koenig, and Daragh Murray, (New York: Oxford University Press, 2020), 273.

<sup>20</sup> Of course, there can also be security and other advantages to having a series of dispersed and unconnected archives.

# BACKGROUND

## Part II – Typology of Digital Archives

### SOCIAL MEDIA PLATFORMS AS “ACCIDENTAL ARCHIVES”

People around the world turn to the internet to share their experiences and bring attention to injustices, and as a result (and as explained above), social media platforms have become unintended repositories of human rights content—both with and without potential evidentiary value.<sup>21</sup> However, social media platforms were not designed to be historical or evidentiary archives. The companies that operate these platforms must comply with a number of sometimes conflicting legal and human rights obligations and are incentivised by the interests of their advertisers and shareholders, as well as the differing perspectives of diverse civil society organizations. These companies also have user guidelines and content moderation practices that they are expected to uphold evenhandedly—leading to the need to craft carefully designed policy exceptions. The companies decide whether or not content stays up on their platforms and how that content is prioritized for viewing.

In recent years, while human rights groups have been searching social media for content relevant to atrocities, social media companies have increased the speed at which they remove relevant content

through automation—further exacerbating the clash between the two groups. The deployment of machine learning algorithms to flag and remove certain types of posts, such as content deemed violent, extremist, or “terrorist,” has resulted in the removal of large quantities of content that potentially offer valuable documentation of alleged human rights violations, despite human rights advocates’ best efforts to explore options for preserving digital content en masse.<sup>22</sup>

It is, of course, in the interest of social media companies and many of their users—including human rights advocates—to remove or deprioritize dangerous content or posts, which is the very content that may be most useful for human rights documentation. Social media platforms do not always provide human rights practitioners with justifications or reliable notice of takedowns, further hindering those platforms’ ability to proactively preserve relevant content.<sup>23</sup> Over the last few years, social media

---

21 “Removals of Syrian Human Rights Content,” Syrian Archive, accessed January, 2021, <https://syrianarchive.org/en/lost-found/may19-takedowns>

---

22 See e.g., Abdul Rahman Al Jaloud, Hadi Al Khatib, Jeff Deutch, Dia Kayyali, and Jillian C. York, “Caught in the Net: The Impact of ‘Extremist’ Speech Regulations on Human Rights Content,” Electronic Frontier Foundation, May, 2019, <https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content>.

23 Social media companies do not provide warning of takedowns when the underlying content violates their policies and legal mandates related to dangerous organizations and child sexual abuse. In addition, none seem to yet have a comprehen-

companies, advocacy and research organizations, and international courts have explored whether establishing an external repository or designing a new legal framework could ease these tensions.

There are at least four types of digital archives that could be used to inform the development of a mechanism to preserve digital content for human rights cases and/or other purposes.<sup>24</sup> Each offers lessons that might inform a constructive way forward. However, before diving into those models and to provide context, we first summarize the traditional role of archives in society and some of the common principles that inform their structure and function.

## TRADITIONAL ARCHIVES

According to the Society of American Archives, “archives . . . are permanently valuable records—such as letters, reports, accounts, minute books, drafts, final manuscripts, and photographs—of people, businesses and governments.”<sup>25</sup> The definition of records is “information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business.”<sup>26</sup> Archives are the result of active roles and processes, and archivists are those who manage, maintain, and preserve the human records that are generated as a product of individuals’ and societies’ day-to-day living.

In “Theories of the Archive across Disciplines,” former MIT Library Collections Strategist and scholar Marlene Manoff describes the complexity of defining the term *archive*, and how that definition has been both “loosening and exploding”<sup>27</sup> as the value of archives draws increasing attention for a range of social functions. Further contextualizing the diverse role of archives, Emeritus Professor of Archivistics at the University of Amsterdam Eric Ketelaar<sup>28</sup> pinpoints the preservation and construction of memory as key elements of archives. Ketelaar explains that shifts in archival policy have allowed societies to understand why archives are a critical backbone of collective memory, and as such, have a broader goal than joint legal accountability or journalism. Regardless of end use, however, one of the most important characteristics of an archive is that it holds lasting value as a connecting thread through time.

Archives, thus, fulfill a crucial role in society. The information they safeguard, organize, and make accessible allows people to exercise their rights, hold institutions and governments to account, establish historical narratives, protect evidence for later legal processes, and preserve information for future generations.

Archival science is an ever-changing discipline that builds upon set methodologies,<sup>29</sup> while adapting to new technologies. As Ketelaar puts it, in “liberating the file from the one and only context of the record creator” we allow for different perspectives in which “the *subject* of the record” can also become

---

sive system for identifying human rights defenders or organizations.

24 In this report, we use the term “human rights cases” to mean all legal cases that allege violations of human rights, humanitarian and international criminal law.

25 “What Are Archives?” Society of American Archivists, September, 2016, <https://www2.archivists.org/about-archives>.

26 “Technical Committee ISO/TC 46, Subcommittee SC 11. ISO 30300:2011(En), *Information and Documentation — Management Systems for Records — Fundamentals and Vocabulary*. 30300, 2011, 3.1.7.” Online Browsing Platform, accessed May, 2021, <https://www.iso.org/obp/ui/#iso:std:iso:30300:ed-1:vi:en>.

---

27 Marlene Manoff, “Theories of the Archive from Across the Disciplines,” *Libraries and the Academy* 4, no. 1 (2004), 9–25, doi:10.1353/pla.2004.0015.

28 Eric Ketelaar, “Archives as Spaces of Memory,” *Journal of the Society of Archivists* 29, no. 1, (2008): 9–27, <https://doi.org/10.1080/00379810802499678>.

29 See, e.g., Yvonne Ng, “How to Preserve Open Source Information Effectively,” in *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation and Accountability*, eds. Sam Dubberley, Alexa Koenig, and Daragh Murray, (New York: Oxford University Press, 2020), 259–287.

a “party to the record.” As social media platforms take root across disparate geographies, multiple individuals are able to contribute to the documentation of events with their stories, testimonies, and experiences.

Today, societies face the challenge of preserving digital collections with the same rigor, archival authority, access, and meaning attached to analog archives, while using information and communication technologies to reach a broader set of actors and create a more “participatory historical culture”<sup>30</sup> than previously possible. These are significant challenges and are a subject of much discussion in archival science.

Analog archives raise many ethical and philosophical considerations that continue to be relevant in the digital age. For example, critical theory contributes an extensive literature to archival studies. Much of this is beyond the confines of this report, but elements of philosopher Jacques Derrida’s writings are instructive for creating an archive or other repository that honors human dignity and upholds human rights. Lecturer and historian of science Elizabeth Yale explains that for Derrida “violence [is] at the heart of archiving: when memories and stories are recorded in the archive, alternate possibilities, other ways of telling the story, are repressed or suppressed.”<sup>31</sup> By necessity, archiving involves the exclusion of certain documents, and with that, the exclusion of particular voices and understandings.<sup>32</sup> Determining which documents should be kept external to the archive is as important as which

---

30 Digital storytelling started in the 1990s at the Center for Digital Storytelling in California which inspired, among others, the BBC’s Capture Wales project and BBC’s Northern Ireland Story Finders. Ketelaar, “Archives as Spaces of Memory,” 9–27.

31 Elizabeth Yale, “The History of Archives: The State of the Discipline,” *Book History* 18, (2015): 334. doi:10.1353/bh.2015.0007.

32 For a discussion of the ways in which digital technologies and their preservation and use for investigations implicate power, see, e.g., Alexa Koenig and Ulic Egan, “Power and Privilege: Investigating Sexual Violence with Digital Open Source Information,” *Journal of International Criminal Justice*, (2021).

## POTENTIAL USES OF DIGITAL CONTENT



documents are included. It is crucial to consider who is selecting the materials for preservation, whose voices are represented, and whose are obscured. Historically, archiving was a means to consolidate power,<sup>33</sup> and archives were generally created and controlled by powerful individuals, groups, and institutions. Whenever an archive is created, power dynamics are at play—a fact that the broader human rights community should center in its on-going conversations.<sup>34</sup>

Yale asserts that “no archive is innocent,”<sup>35</sup> that regardless of the original intent behind an archive, archives can be harnessed in multiple and unanticipated ways. The same records can be used for terror or justice, depending on who mobilizes them.<sup>36</sup> An example of this is the archive of the Ministry of State Security of the former German Democratic Republic (East Germany), known colloquially as the “Stasi” records. In that context, records that had been gathered by East Germany’s secret police

---

33 Yale. “The History of Archives: The State of the Discipline,” 332.

34 Ibid.

35 Ibid.

36 Ibid., 346.



later helped build the historical record and collective memory of life in East Germany from the 1950s through the 1980s—including revealing the fate of people terrorized during that period.

Similarly, in 2005, delegates from the Guatemalan Procurator for Human Rights discovered the Guatemalan National Police Archives in an abandoned warehouse.<sup>37</sup> Originally developed for book-keeping purposes, these archives are now being used to identify the role of the national police in the Guatemalan Civil War.<sup>38</sup> Today, forensic teams are still archiving the documents to use for historical and legal purposes.

Similar examples have played out in relation to the case against Democratic Kampuchea's (present-day Cambodia's) 1975–1979 genocide during the Khmer Rouge regime. The Yale Cambodian Genocide Project (GCP), started by Yale's History Professor Ben Kiernan in 1994, holds more than 100,000 documents, photographs, and maps to support trials of top Khmer Rouge leaders. The intent of the documentation was “to determine who was primarily responsible for the tragedy.”<sup>39</sup>

Collections of testimonial and documentary evidence have also informed other international trials, including those related to more recent atrocities,

including international crimes in Sudan, Bosnia, Guatemala, East Timor, Iraq<sup>40</sup> and Rwanda.<sup>41</sup>

These examples demonstrate the challenges that may arise when documents created for one purpose (for example record-keeping related to employment conflicts with later uses, such as accountability, but

---

40 The Iraq Memory Project was founded by Kanan Makiya and funded by the U.S. government to collect and preserve documents from Saddam Hussein's Ba'athist government of 1968 to 2003. In 1991, founder Makiya and a BBC filmmaker traveled to Iraq to collect and archive documents that had been seized by Iraqi rebels relating to the Ba'ath party, including the Iraqi government's campaign of ethnic cleansing of Iraqi Kurds. The Iraq Memory Project consists of the following: A Documentation Project, an Oral History Project, a Public Outreach Program, a Research Program, a Liaison and Coordinating Center, and a “Placing the Iraqi Experience” project. The entire collection is available to the public at Stanford's Hoover Institute, including approximately ten million digitized pages and one hundred video files of the Ba'ath Arab Socialist Party of Iraq. A controversy, however, is who should have custody over the archived material. In 2008, the director of the Iraq National Library and Archive in Baghdad and acting Minister of Culture stated that the documents of the Ba'athist government were unlawfully seized from Iraq and should be returned. Additionally, the Society of American Archivists and Association of Canadian Archivists agree that the archived materials should not be held by a private organization. The two organizations issued a joint statement saying that the gathering of those documents was an act of pillage forbidden by 1907 Hague Convention and “should be returned to the government of Iraq to be maintained as part of the official records in the National Library and Archives.” See, “ACA/SAA Joint Statement on Iraqi Records,” Society of American Archivists, April, 2008, <https://www2.archivists.org/statements/acasaa-joint-statement-on-iraqi-records>. For more information about the Iraq Memory Project, see, “About,” Iraq Memory Project, accessed March, 2021, <http://www.iraqmemory.com/en/about>; Renee Montagne, “Iraq's Memory Foundation: Content in Culture,” NPR, March, 2005, <https://www.npr.org/templates/story/story.php?storyId=4554528>; Hugh Eakin, “Iraqi Files in U.S.: Plunder or Rescue?,” The New York Times, July, 2008, <https://www.nytimes.com/2008/07/01/books/01hoov.html>; “Iraq,” Hoover Institution, March, 2021, <https://www.hoover.org/library-archives/collections/iraq>.

41 “Interview: Documenting Year Zero,” POV, July, 2003, <http://archive.pov.org/thefluteplayer/interview-documenting-year-zero/>

---

37 Kate Doyle, “The Guatemalan Police Archives: National Security Archive Electronic Briefing Book No. 170,” The National Security Archive, November, 2005, <https://nsarchive2.gwu.edu/NSAEBB/NSAEBB170/index.htm>.

38 Kirsten Weld, *Paper Cadavers: The Archives of Dictatorship in Guatemala*, (Durham: Duke University Press, 2014), <https://www.dukeupress.edu/paper-cadavers>.

39 Alaina Varvaloucas, “Three decades later, justice for genocide victims,” The Yale Herald, September, 2008, [http://gsp.yale.edu/sites/default/files/varvaloucas\\_yale\\_herald\\_09.26.08.pdf](http://gsp.yale.edu/sites/default/files/varvaloucas_yale_herald_09.26.08.pdf).

the original documentation wasn't conducted in ways that were "fit for [later] purpose." This disjunct sometimes surfaces questions around the ethics, legality and/or utility of using information gathered or created for one purpose, for another. This "dual use" also raises the concern: how to keep any archival repository from becoming a means of inappropriate surveillance and other overreach by government actors and others who might be hostile to the human rights concerns of those whose data is included?

From a logistical perspective, Yale advocates for consistently following archival science principles when creating any repository,<sup>42</sup> or "thinking archivally."<sup>43</sup> This is a process that requires understanding the context and conditions under which each document was created.<sup>44</sup> Archival thinking also "demands that we see archives not only as sources of data to be mined by researchers but also as more than the sum of their parts."<sup>45</sup>

Yvonne Ng of WITNESS has explained the basic components of digital preservation, including how archival principles relate to digital documentation. Citing the Reference Model for an Open Archival Information System, which establishes an international standard for archivists, she notes that "while an archive's preservation strategies must be customized to its circumstances, the nature of its collections, and the needs of its intended users, there are established guidelines that describe [those basic components]."<sup>46</sup> These include thinking about how information is bucketed into one of three "packages," including the submission information package (which is used for transporting information into an archive), archival information packages (the information stored in the archive), and dissemination

information packages (the information shared with users).

In addition, various properties related to the content must be protected and preserved, including the item's authenticity (ensuring that an item remains unchanged), availability (through ongoing existence and retrievability), identity (using a system to make the items identifiable and distinguishable from other items, as with a unique identifier), persistence (the technical integrity and viability of a digital item), renderability (the ability of humans and / or machines to use the digital item), and understandability (a human's ability to interpret or "understand" the digital item). For legal purposes, archivists should also consider the need to maintain chain of custody (logging who has had access to the digital item and when, and what precautions have been taken to avoid alteration), as well as the importance of keeping working copies that can be modified separate from evidentiary copies that cannot, and the feasibility and sufficiency of both long and short-term storage.<sup>47</sup>

Those engaging in the practice of digital preservation must think through every aspect of the process of archiving. In addition to the items listed above, this includes maintaining a sensitivity to context and adherence to diverse principles, and respecting the relative fragility of digital information. They should also consider what might happen if the archive is co-opted by those with malintent, problematically deployed by malicious actors, or used by those with the best of intentions who may find that the archiving of information results in unintended consequences.

---

42 Ibid.

43 Yale. "The History of Archives: The State of the Discipline," 345.

44 Weld, *Paper Cadavers: The Archives of Dictatorship in Guatemala*, 13.

45 Ibid.

46 Ng, "How to Preserve Open Source Information Effectively?"

---

47 See, e.g. UC Berkeley Human Rights Center and UN Human Rights Office, "Berkeley Protocol on Digital Open Source Investigations," UN Office of the High Commissioner for Human Rights, 2020, 202.

## DIGITAL ARCHIVES

Digital archives contribute additional qualitative and quantitative challenges to the continuity and security challenges of traditional archives. The first difference is the volume of potential content. Because of the extraordinary scale of potentially relevant data, archivists need tremendous resources and storage space to identify and preserve relevant data. In addition, digital archives must adjust for shifting dynamics between the producers of content, the subjects of content, the holders of content (e.g. tech companies, journalists, state actors), and codes of conduct for processing human rights-related content in ways that respect privacy laws.

In this section, we build off the background provided above to provide a typology of four different types of digital “archives” that have previously been used to aggregate and/or preserve digital content—including but not restricted to social media content. We illustrate those examples with a series of case studies, briefly describing each archive’s history, ownership, structure, financing, and user-accessibility, to explore whether any of these examples may inform the feasibility of creating one or more digital evidence lockers. The case studies are discussed and categorized based on the following criteria:

1. **Who provides the content:** This can be social media companies using their own internal processes or NGOs and other external actors who may scrape or manually download content from platforms. This also includes content creators (for example, those who take video recordings or photographs of human rights content) and uploaders who act as intermediaries between the content creator and repository.
2. **Who holds the content:** This varies based on:
  - a. Whether the social media company holds the information on servers it controls;

- b. Whether the social media company provides the content to an external organization or repository;
- c. Whether an external organization downloads the content and holds the content on their own servers or servers to which they have access; and
- d. What content is included, since the type of content will often dictate where it goes (for example, child sexual exploitation material being held by the National Center for Missing and Exploited Children versus hashes of violent extremist content being shared with the Global Internet Forum to Counter Terrorism).

3. **Legal obligation:** This variable focuses on whether social media companies are legally required to preserve and/or share the content and/or protect the privacy of the content, versus whether participation is voluntary.

4. **Who may access the content:** The scope of access varies significantly between the different models. However, the range of possibilities tends to cluster into the three “buckets” identified below:

- a. Private: Only the social media company and/or law enforcement can access the user-generated content;
- b. Subscribers: Members of the public can request and/or otherwise qualify for access (e.g. by registration, payment, or other process); or
- c. Public: Anyone from the general public can access the user-generated content so long as they have access to the internet.

In some instances, a single repository may provide differential access to each of these groups, with varying requirements for how the information can be used.

## THE DIGITAL ARCHIVE MODELS

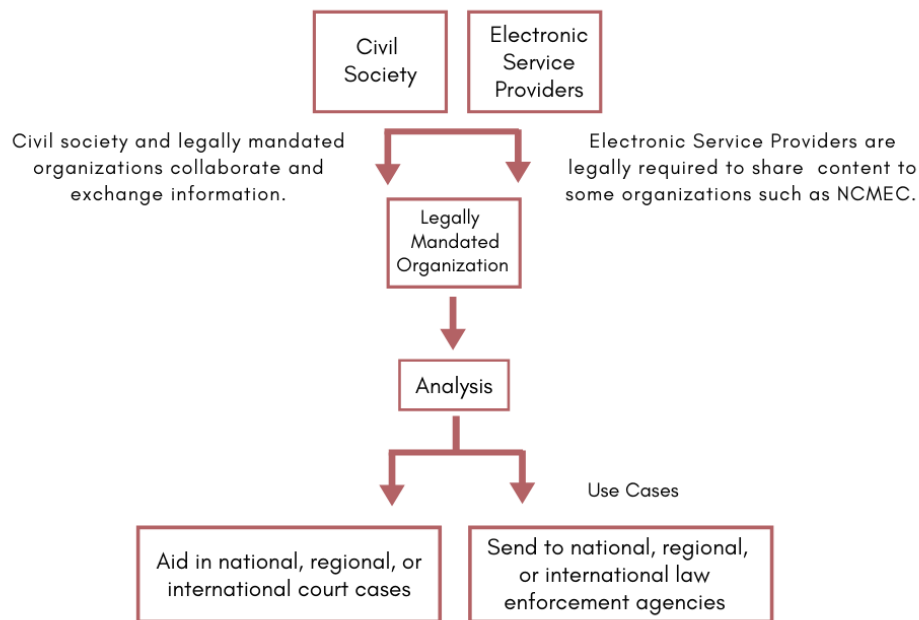
NAME OF MODEL	DESCRIPTION	EXAMPLES
THE LEGAL COMPULSION MODEL	<p>This model consists of entities who have a legal mandate to pursue their investigative and/or human rights duties. Additionally, social media companies are legally required to share content with some of these organizations. For some, the companies preserve content upon service of legal process to fulfill law enforcement requests. In the international circuit this may mean engaging in formal 90-day preservation requests and using the MLAT process to obtain the materials. Alternately, international investigators and others cooperate with law enforcement when their mandates overlap, which allows for more informal information sharing.</p>	<ol style="list-style-type: none"> <li>1. The National Center for Missing &amp; Exploited Children (NCMEC)</li> <li>2. Core International Crimes Project (AP CIC)</li> </ol>
THE VOLUNTARY PARTNERSHIP MODEL	<p>Social media companies voluntarily share content with an external repository. These entities may differ in how they archive and use such content.</p>	<ol style="list-style-type: none"> <li>1. Lumen Database</li> <li>2. The Terrorist Content Analytics Platform (TCAP)</li> </ol>
THE INDEPENDENT COLLECTION MODEL	<p>This model consists of organizations who independently download or scrape content specific to particular crises or themes and with defined geographic and temporal parameters. Relevant content is often collected from online sites, including social media platforms, as well as diverse and dispersed civil society organizations. These repositories may also collect content from non-social media sources, such as other NGOs, or activists located in the sites of conflict. These entities preserve content on their own servers or rented server space. Such archivists may work closely with the social media companies to reinstate accounts when channels related to their mandate close, or to otherwise coordinate their efforts.</p>	<ol style="list-style-type: none"> <li>1. Mnemonic / The Syrian Archive</li> <li>2. Jihadology</li> <li>3. ICRC Archives</li> </ol>
THE HYBRID MODEL	<p>Various providers such as social media companies, investigators, NGOs, IGOs, and civil society organizations share content with a Mechanism under a memorandum of understanding (MOU).</p>	<ol style="list-style-type: none"> <li>1. International, Impartial and Independent Mechanism (IIIM)</li> <li>2. Independent Investigative Mechanism for Myanmar (IIMM)</li> </ol>

After considering these variables, we found that the digital archives we examined tended to fall into one of the following types. Each type varies in terms of how it maintains the integrity of the content, which can have a bearing on the ability to authenticate the content at trial. For clarity, we organized the case studies into four models: the Legal Compulsion Model, the Voluntary Partnership Model, the Independent Collection Model, and the Hybrid

Model. There is considerable overlap between the case studies we present, and as such, these models are neither exclusive nor exhaustive.

Ultimately, the strengths and weaknesses of each of these models for human rights evidentiary purposes can only be assessed on the basis of articulated end goals, and are thus context-specific. Below, we provide examples of each of the models.

## THE LEGAL COMPULSION MODEL



### The National Center for Missing & Exploited Children (NCMEC)

The National Center for Missing & Exploited Children (NCMEC) exemplifies a model where social media companies are required to share content with an external repository under the force of law. The private, non-profit organization was established by the United States Congress in 1984<sup>48</sup> and is mandated to help find missing

children, reduce child sexual exploitation, and prevent child victimization.<sup>49</sup> NCMEC has been heavily supported by the Office of Justice Programs (OJP), a division of the United States Department of Justice. To give a sense of funding and scale, OJP awarded NCMEC \$33 million in the 2019 fiscal year to assist its operations.<sup>50</sup>

48 NCMEC is mandated by 42 U.S.C. §§5771 et seq.; 42 U.S.C. §11606; 22 C.F.R. §94.6.

49 U.S. Department of Justice, “The National Center for Missing and Exploited Children,” Office of Juvenile Justice and Delinquency Prevention, September, 2019, <https://ojjdp.ojp.gov/funding/awards/2019-mu-mu-ko12>.

50 Ibid.

Given its design and public-private partnership structure, NCMEC has access to several databases of content provided by its external partners, including social media companies and the Federal Bureau of Investigation (FBI).<sup>51</sup> The following case study focuses on NCMEC's core responsibilities of managing its CyberTipline and Child Victimization Identification Program (CVIP), both of which predate the Internet and had to be adapted to a social media context. The CyberTipline acts as a clearinghouse for complaints of child sexual exploitation and child pornography. The CyberTipline collects information regarding child sexual abuse imagery and distributes this data to relevant law enforcement agencies. CVIP is a central database of images depicting identified child victims.<sup>52</sup> Electronic communication service providers are legally required to report images of identified child victims to NCMEC and retain related child sexual abuse imagery information for approximately 90 days under the Our Children Act of 2008.<sup>53</sup>

There are two ways in which child sexual abuse material can enter the CyberTipline. First, members of the public who observe someone accessing or disseminating child sexual abuse imagery can report

such occurrence. Second, electronic communication service providers that detect child sexual abuse imagery on their services can report that information to NCMEC.<sup>54</sup> After NCMEC receives the report, an analyst "reviews, augments, and deconflicts" the report.<sup>55</sup> For U.S. cases, NCMEC then sends the information to United States federal or state law enforcement. For cases outside the United States, NCMEC works with Interpol, Europol, and national police.<sup>56</sup>

In 2018, the CyberTipline handled 18.4 million reports, yet the number of individual videos and images that were reported reached as high as 70 million.<sup>57</sup> The majority of the reports came from electronic communication service providers. Many companies rely on software such as Microsoft's PhotoDNA to find and remove images of child exploitation on their platform.<sup>58</sup> Electronic service providers are, however, not mandated to search their platforms for such content, so participation varies significantly. For example, in 2019 Facebook voluntarily submitted over 85 percent of all CyberTipline reports, in part because they are one of few social media companies that actively and voluntarily inspect uploaded content for imagery related to child sexual abuse.<sup>59</sup>

---

51 Elie Bursztein, Travis Bright, Michelle DeLaune, David Eliff, Nick Hsu, Lindsey Olson, John Shehan, Madhukar Thakur, and Kurt Thomas, "Rethinking the Detection of Child Sexual Abuse Imagery on the Internet," *Proceedings of the 2019 World Wide Web Conference (WWW '19)*, May, 2019, <https://elie.net/static/files/rethinking-the-detection-of-child-sexual-abuse-imagery-on-the-internet/rethinking-the-detection-of-child-sexual-abuse-imagery-on-the-internet-paper.pdf>.

52 "CyberTipline: Is a Child Being Exploited Online?," National Center for Missing & Exploited Children, accessed October, 2020, [https://www.missingkids.org/gethelpnow/cyber\\_tipline](https://www.missingkids.org/gethelpnow/cyber_tipline); "Privacy Impact Assessment (PIA) Child Victim Identification Program (CVIP) Innocent Images National Initiative (IINI)," Federal Bureau of Investigation, May, 2003.

53 Legislation Sponsored by Senator Biden, Joseph, "S.1738 - 110th Congress (2007-2008): PROTECT Our Children Act of 2008," Legislation of Library of Congress, October, 2008, <https://www.congress.gov/bill/110th-congress/senate-bill/1738>.

---

54 Bursztein et. al., "Rethinking the Detection of Child Sexual Abuse Imagery on the Internet."

55 Ibid.

56 "National Strategy for Child Exploitation Prevention and Interdiction," U.S. Department of Justice, April, 2016, <https://www.justice.gov/psc/national-strategy-child-exploitation-prevention-and-interdiction>.

57 Caren Harp, "Visiting Our Partners at the National Center for Missing & Exploited Children," Office of Justice Programs, August, 2019.; Gabriel J.X. Dance and Michael H. Keller, "Tech Companies Detect a Surge in Online Videos of Child Sexual Abuse," *The New York Times*, February, 2020, <https://www.nytimes.com/2020/02/07/us/online-child-sexual-abuse.html>. <https://www.ojp.gov/news/ojp-blogs/2019/visiting-our-partners-national-center-missing-exploited-children>.

58 Ibid.

59 Ibid.

The Child Victim Identification Program (CVIP) is a component of the Innocent Images National Initiative, which is part of the FBI's Cyber Crimes Program.<sup>60</sup> CVIP seeks to identify the victims of those who commit sexual exploitation of children, and serves as a central repository for images depicting child victims. Through this program, CVIP assists field offices in their efforts to identify new child pornography victims in CyberTipline reports to reduce duplicate investigative efforts. Any evidence obtained by the field offices is compared with existing datasets via hash values. Since 2002, this program has processed more than 149 million pieces of digital content of alleged child pornography,<sup>61</sup> securing the following data associated with victims (if available): identification number, internet nickname, date of birth, age at the time of the photograph, gender, citizenship, nationality, identifying officer name, and identifying officer contact details. Other data may include physical characteristics such as height, weight, hair color, and eye color.

With the advent of social media platforms, NCMEC has been grappling with several challenges that are overwhelming its capabilities. One of the major obstacles is the fast-growing number of child sexual abuse imagery reports and content. In 2017 alone, 9.6 million reports of child sexual abuse imagery were sent to NCMEC, which constituted around 40 percent of NCMEC's total cases across its history.

This challenge is compounded as new child sexual abuse imagery content is constantly surfacing.

---

60 "Privacy Impact Assessment (PIA) Child Victim Identification Program (CVIP) Innocent Images National Initiative (IINI)," Federal Bureau of Investigation Services, May, 2003, <https://www.fbi.gov/services/information-management/foipa/privacy-impact-assessments/cvip>.

61 Melissa Stroebel and Stacy Jeleniewski, "Global Research Project: A Global Landscape of Hotlines Combating Child Sexual Abuse Material on the Internet and an Assessment of Shared Challenges," National Center for Missing & Exploited Children, 2015, 3, <https://www.missingkids.org/content/dam/missingkids/pdfs/ncmec-analysis/grp.pdf>.

More than 80 percent of this material is reported only once, with a rate of one million reports submitted per month.<sup>62</sup> Scale is a problem, especially given NCMEC's use of relatively outdated technology and the limitations of manual review of collected content. Another complexity that NCMEC faces is offenders' adoption or use of new technologies to mask their identities. Offenders have been covering their digital footprints by connecting to virtual private networks (VPNs), deploying encryption techniques, and using the Dark Web. For example, Facebook reported nearly 60 million photos and videos of child sexual abuse imagery in 2019, most of which was found in its private Messenger App.<sup>63</sup> These hurdles present an overwhelming pressure on NCMEC's manual review capabilities and, in return, law enforcement investigations.<sup>64</sup>

Summary:

- **End-uses:** *Help find missing children, reduce child sexual exploitation, prevent child victimization, and distribute information to law enforcement agencies for potential prosecution.*
- **Who provides the content:** *The FBI and electronic communication service providers, including*

---

62 Bursztein et al., "Rethinking the Detection of Child Sexual Abuse Imagery on the Internet." Also worth grappling with are the differences in how civil society organizations conceptualize "scale" versus social media companies and what that means for pragmatic responses to the preservation of information at risk of removal that has important human rights value.

63 Kate Duffy, "Facebook's Encryption Plans Will Make It Harder to Catch Child Sex Abusers, Governments Warn," Business Insider, October, 2020, <https://www.businessinsider.com/facebook-encryption-harder-catch-criminals-child-abuse-2020-10>. See also Andy Greenberg, "Facebook Says Encrypting Messenger by Default Will Take Years," WIRED, January, 2020, <https://www.wired.com/story/facebook-messenger-end-to-end-encryption-default/>. In its current version Facebook Messenger offers the Secret Conversations feature which allows users to opt into end-to-end encryption.

64 Bursztein et al., "Rethinking the Detection of Child Sexual Abuse Imagery on the Internet."

social media companies, report images of child victims to NCMEC. Members of the public who observe someone accessing or sharing child sexual abuse imagery can report to NCMEC's CyberTipline.

- **Who holds the content:** *Given its public-private partnership structure, NCMEC has access to several databases of private content provided by external partners.*
- **Who can access the content:** *NCMEC shares content and information with U.S. federal, state, and local law enforcement, as well as Interpol, Europol, and other national police.*
- **Legal obligations:** *NCMEC is legally mandated to help find missing children, reduce child sexual exploitation, prevent child victimization, and send information to national and international law enforcement agencies.*
- **Unique features:** *Evidence obtained by NCMEC's field offices are compared with other datasets via hash values to reduce duplicate investigative efforts. Social media companies, like Facebook, that actively inspect uploaded content for imagery related to child sexual abuse, constitute the majority of the CyberTipline submissions, which aid NCMEC in its operations.*
- **Challenges:** *NCMEC is grappling with the large scale of child sexual abuse imagery, as well as offenders covering their digital footprints and using the DarkWeb for anonymity.*

## Europol's Core International Crimes Analysis Project

The European Union Agency for Law Enforcement Cooperation, commonly known as "Europol," was established in 1991 and supports the EU Member States in preventing and investigating a wide range of crimes with an international dimension, ranging from economic crime to terrorism, cybercrimes,

child sexual exploitation, drug trafficking, and the facilitation of illegal immigration.<sup>65</sup>

Europol has established Analysis Projects (AP's, previously called Focal Points) which bring together teams of specialists, experts, and analysts that coordinate EU Member States' investigations into crime areas from commodity-based, thematic, or regional contexts such as drug trafficking, terrorism, Italian organized crime.<sup>66</sup> Apart from data cross checking and analysis, Analysis Projects also offer specific operation support, expertise, resources, and training to law enforcement authorities.<sup>67</sup>

Among the Analysis Projects is the Core International Crimes (AP CIC)<sup>68</sup> Project, which was established in November 2017 to "support the competent authorities of the EU Member States, and Europol Third Parties core international crimes investigations (genocide, crimes against humanity, and war crimes)."<sup>69</sup> AP CIC emerged out of long standing requests by countries like Germany and the Netherlands to create an international centralized database where various core international crimes information and intelligence can be safely stored, crosschecked, and analyzed in support of current and/or future investigations. Building and enriching the AP CIC database with such data will lead to better streamlining of investigations, build more expertise amongst investigators, assist in the fight against impunity of core international crimes, and prevent duplicative efforts among states and organizations.

Most of the data AP CIC receives are from specialized law enforcement agencies in the EU or countries with whom Europol has established

---

65 "Europol 20 Years," Europol History, accessed January, 2021, <https://www.europol.europa.eu/history/europol-history.html>.

66 "Europol Analysis Projects," Europol, accessed January, 2021, <https://www.europol.europa.eu/crime-areas-trends/europol-analysis-projects>.

67 Ibid.

68 Core international crimes (CIC) refers to the crime of genocide, crimes against humanity, and war crimes. See, *ibid.*

69 Ibid.



operational agreements (Europol Third Operational Parties), including information origination from international organizations and private entities.<sup>70</sup> Information is exchanged through Europol's offline Secure Information Network Application (SIENA), that "is compliant with all of the legal requirements for data protection and confidentiality."<sup>71</sup> SIENA requires that all information shared with Europol has an attached handling code which details how providers of information want Europol to handle their information, as outlined below:

1. Handling Code Ho: This code entails that information may only be used for the purpose of preventing and combating crimes in line with the applicable law. If, for example, Germany shares Ho information that connects to an investigation happening in Poland, then AP CIC can share the data with Poland. This handling code also means that Poland can use that information freely within their law enforcement and judicial systems.
2. Handling Code H1: This code indicates that information provided can be used for investigative purposes, but cannot be used in judicial proceedings. If there is a need to use H1 information in

---

70 AP CIC Member States include Belgium, Cyprus, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Romania, Spain, and Sweden. AP CIC Third Party Operational Partners include Australia, Canada, Denmark, Norway, U.S./ICE, Switzerland, U.K., and Eurojust. International organizations and private entities with whom AP CIC has operational agreements include but are not limited to, CIJA, Open Society Justice Initiative, International Criminal Court, Redress, U.N. IIM, Syrian Archive, eyeWitness, U.N., International Residual Mechanism for Criminal Tribunals, Yazda, and Syria Justice and Accountability Centre.

71 "Secure Informational Exchange Network Application (SIENA)," Europol, accessed April, 2021, <https://www.europol.europa.eu/activities-services/services-support/information-exchange/secure-information-exchange-network-application-siena>.

a judicial proceeding, prior permission from the information provider is required.

3. Handling Code H2: This is Europol's most strict handling code. This code signals that the shared information can be cross checked with existing data within Europol's databases, but in the case of a match or a hit (e.g. a connection with the name of a suspect or a phone number), permission is required from the provider to share the information with the relevant country to assist in their investigations.
4. Handling Code H3: This handling code allows providers of information to mark other restrictions, permissions, or purposes of transmission (e.g., detailing that their information can only be shared with specific countries if there is a match within Europol's database).<sup>72</sup>

Information providers are considered the sole owners of the core international crimes information they share with AP CIC, which may include different types of data relating to persons (suspects, witnesses, victims), armed groups and affiliations (members, organizational structure, chain of command, ranks, logos), crimes committed, locations, etc. AP CIC does not directly receive data from social media companies or from telecommunication providers; however, content provided by Member States, Third Party Operational Parties, or private entities may include social media content or data records. On specific requests, AP CIC may provide countries with automated OSINT or SOCMINT searches conducted by the Europol Internet Referral Unit (EU IRU).

Personal data stored in Europol's databases provided by a Member State or Europol's Third

---

72 If the information shared by a provider is critical to the protection of an individual's life, there are some exceptions to how AP CIC can provide the information to the relevant party. As of April 2021, AP CIC has not encountered such a situation.

Operational Parties is reviewed no later than three years after the start of initial processing.<sup>73</sup> If continued storage of personal data is necessary for Europol's tasks, then the reasons for continued storage are justified and recorded. If data no longer proves relevant to investigations or no decision is taken on its continued storage after three years, the data is automatically erased from Europol's databases.

To ensure the public is well-informed of its activities and to facilitate access to its files, Europol has a public registry on its website that enables public access to public documents including articles, reports, reviews, and threat assessments.<sup>74</sup> This online registry is searchable by year and type of publication. However, no information related to its operational data is accessible to the public. If a document has not been published on Europol's site or is not downloadable through the public interface, Europol allows the public to request the document through a "public access request form."<sup>75</sup>

Although Europol's AP CIC does not receive content directly from social media companies like NCMEC does, this case study offers several lessons. First, it outlines how in the European context a network of information systems can share critical content with law enforcement—including social media content.<sup>76</sup> Second, this case study demonstrates the need for a mandate and resources to allow an organization like AP CIC to proactively research and store open source digital content related to core

international crimes. Third, Europol and AP CIC highlight how digital information can be shared with the public through a public-facing registry. Finally, this case study offers insight into an international centralized database with standardized collection methods facilitated via operational agreements, secure information exchange practices (SIENA), and long-term storage of information for current and future investigations, and the streamlining of investigative and case-building efforts among nations, international organizations, and private entities.<sup>77</sup>

Summary:

- **End-uses:** *Offers operational support, expertise, resources, and information to EU Member States and Europol Third Operational Parties in support of investigations into crimes of genocide, crimes against humanity, and war crimes.*
- **Who provides the content:** *EU Member States, Europol Third Operational Parties, international organizations, and private entities provide information related to alleged crimes.*
- **Who holds the content:** *Provided data is safely stored in Europol's secure databases and can be shared out to EU Member States and Europol Third Operational Parties on the basis of 'hits' or 'matches' and applicable handling codes.*
- **Who can access the content:** *The full repository of content is accessible to only a few Europol staff members. Provided data that 'matches' with an ongoing investigation can be shared with EU Member States and Europol Third Operational Parties on the basis of applicable handling codes. The public can access public-facing articles, reports, and threat assessments through Europol's public registry. The public can also request other documents not available through the website.*

---

73 Personal data sent to AP CIC directly from private entities, such as NGOs, can only be processed to identify a 'match' to an E.U. Member State, Europol Third Operational Partner, or organization with whom Europol has established an operational agreement. 'Match' of data means that the information provided by private entities connects with an existing data (e.g., name of a suspect or a phone number) within Europol's databases. If such a connection or 'match' is not immediately found, AP CIC has to delete the data it received from the private entity.

74 "Publications & Documents," Europol, accessed January, 2021, <https://www.europol.europa.eu/publications-documents>.

75 Ibid.

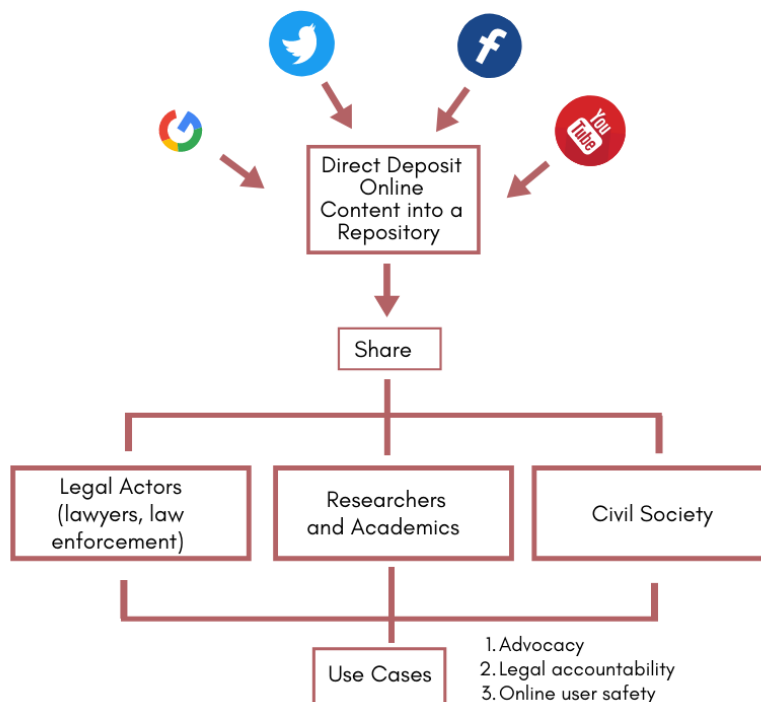
76 Ibid.

---

77 Unless otherwise stated, most of the information outlined within this case study comes from an interview conducted with Jörg Roofthoof, Project Manager of AP CIC, on April 14, 2021.

- **Legal obligations:** AP CICI is operated by Europol, which has its legal basis in EU Regulation 2016/794.
- **Unique features:** Europol can also enter into co-operation agreements with non-EU countries and with international organizations. Different kinds of agreements are used for the exchange of non-personal versus personal data. Information can be stored for three years for longer-term investigative and case-building efforts, with opportunities to archive relevant content for longer periods of time.
- **Challenges:** AP CICI does not conduct OSINT or SOCMINT searches to proactively identify information relevant to core international crimes (the exception being a limited number of cases on request). Instead, it receives such information from law enforcement. The capacity and resource burden on participating organizations can be significant.

## THE VOLUNTARY PARTNERSHIP MODEL



### Lumen Database

An example of the Voluntary Partnership Model—one in which companies *voluntarily* provide information to a centralized, external repository—is the Lumen Database. The database contains cease

and desist letters that demand the removal of online content. This includes notices that ask for the removal of social media links from search engines, as well as notices sent directly to social media sites. These removal requests are collected and analyzed by Lumen to facilitate transparency around who is

sending such requests, to whom, and to what effect. Lumen receives voluntary submissions from companies such as Google, Twitter, YouTube, Wikipedia, Counterfeit Technology, Medium, Stack Exchange, Vimeo, DuckDuckGo, disparate entities within the University of California, and Wordpress, among others.<sup>78</sup> These submissions focus on trademark, defamation, and privacy issues, both domestic and international, and court orders.<sup>79</sup>

This collaborative academic database was established in 2002 to guard against legal threats to online activity.<sup>80</sup> The project was founded by then-Berkman Klein Center Fellow Wendy Selzer and was previously known as “Chilling Effects,”<sup>81</sup> reflecting an intent to protect against the stifling impact of cease and desist notices on online activity.

In 2015, Lumen expanded internationally, announcing a series of pilot partnerships with non-U.S. based centers and institutes. Each partner became a “regional hub,” addressing local takedown requests and drawing upon local expertise to translate notices into local languages. Each partner lends expertise to the larger project, particularly helping to facilitate data-sharing between non-U.S. entities and Lumen’s headquarters. These partners include The NEXA Center for Internet & Society in Turin, Italy; The Instituto de Tecnologia & Sociedade in Rio de Janeiro, Brazil; and The Center for Communication Governance in Delhi, India.

The database aims to foster public awareness and facilitate research pertaining to cease and desist

letters issued for online content.<sup>82</sup> Lumen includes letters ranging from domestic sources to international ones, with subject areas that span allegations of defamation, trademark infringement, and privacy violations.<sup>83</sup> Lumen archives both the complaints themselves and the source of the content removals.<sup>84</sup> According to their website, “the Lumen database collects and analyzes legal complaints and requests for removal of online materials, helping Internet users to know their rights and understand the law.”<sup>85</sup> Through its database, Lumen analyzes the pervasiveness of legal threats to make the source of content removals transparent for Internet users.<sup>86</sup>

According to the Berkman Klein Center and Lumen’s website, the Lumen database is increasing at a rate of 20,000–40,000 takedown notices per week.<sup>87</sup> By the Summer of 2019, the project hosted approximately twelve million notices, which referenced close to four billion URLs.<sup>88</sup> These notices are submitted voluntarily by businesses that receive removal requests *or* send such requests.<sup>89</sup> Given Lumen’s ability to bridge the gap between corporate and academic spaces and interests, Lumen serves as a potential model for stakeholders to look to.

The database has an expansive reach. For example, in 2018, Lumen’s database was visited more than ten million times by users from most countries in the world.<sup>90</sup> Users can browse the database’s

---

78 “About: Lumen,” Lumen, accessed February, 2021, <https://www.lumendatabase.org/pages/about>.

79 “Lumen Notice Information Basics,” Lumen, accessed March, 2021, <https://www.lumendatabase.org/pages/lumen-notice-basics>.

80 Lumen is an independent research project at the Berkman Klein Center for Internet & Society at Harvard University. Lumen is financially supported by a grant from Arcadia, the U.K. charitable fund of Lisbet Rausing and Peter Baldwin. See, “About: Lumen,” Lumen, accessed February, 2021, <https://www.lumendatabase.org/pages/about>.

81 Ibid.

---

82 Ibid.

83 “About: Lumen,” Lumen, accessed February, 2021, <https://www.lumendatabase.org/pages/about>.

84 Ibid.

85 “Home Page: Lumen,” Lumen, accessed October, 2020, <https://www.lumendatabase.org/>.

86 Ibid.

87 “Lumen,” Berkman Klein Center, accessed October, 2020, <https://cyber.harvard.edu/research/lumen>.

88 “About: Lumen,” Lumen, accessed February, 2021, <https://www.lumendatabase.org/pages/about>.

89 Ibid. Lumen’s website states that for more information about these companies’ submissions to Lumen, please refer to each company’s own website and information pages.

90 Ibid.

web interface to find information or access larger swathes of data more easily through Lumen’s application programming interface (API).<sup>91</sup> The API allows individuals and organizations to submit large numbers of notices without having to go through the web interface.<sup>92</sup> Lumen is searchable for free,<sup>93</sup> which allows the public to access the cease and desist letters while also providing a window into the legal processes involved when content is deleted.

While the Lumen database used to provide the full URLs for sites subject to takedown demands, in 2019 the URLs were redacted to only display the website domain names and the number of URLs from each site. In addition, users were no longer able to view file attachments associated with each notice (e.g., copies of court orders) without logging in. In its current form, users can obtain a link to the “unredacted” version with a full URL and files for a limited time period. Further access, which requires login credentials, is determined based on individual requests.<sup>94</sup>

Although the public can browse the database using the API, in order to avoid anyone abusing the system, Lumen limits the number of database queries users can run without an authentication key, stating: “API queries are capped at the first 25 results and 5 requests per day.” To receive a token that allows for unlimited querying, users must request an authentication key from the Lumen Team.<sup>95</sup> All requests for access (except media requests) go directly to that team.

---

91 “Wiki,” GitHub, accessed February, 2021, <https://github.com/berkmancenter/lumendatabase/wiki/Lumen-API-documentation>.

92 The technical documentation for Lumen’s API can also be found on GitHub. See, “Code,” GitHub, accessed October, 2020, <https://github.com/shubhscoder/lumendatabase>.

93 “Researchers,” Lumen, accessed October, 2020, <https://www.lumendatabase.org/pages/researchers>.

94 “Lumen Announces New Features of the Database,” Lumen(blog), May, 2019, [https://www.lumendatabase.org/blog\\_entries/803](https://www.lumendatabase.org/blog_entries/803).

95 “Tools for Researchers,” Lumen, accessed March, 2021, <https://lumendatabase.org/pages/researchers>.

Lumen has faced several challenges regarding its practice of making information available about data that was intended to be removed from the Internet. For example, in 2017, a German court ordered Google to stop linking to the Lumen database. The case emerged due to Google’s then-practice where, upon receiving an injunction and taking down a webpage, Google posted an explanatory statement at the website’s address and a hyperlink to Lumen’s website. Clicking this hyperlink re-directed users to a site where the link for the deleted web page still existed, allowing users to read the statements there. The Higher Regional Court of Munich first denied an infringing contribution by Google but then reversed its decision on appeal. A portion of that decision was apparently based on the right to be forgotten.<sup>96</sup> Before the appeal, the claimant asked Google to delete the hyperlink, yet Google refused, citing transparency for users as a priority over the privacy concerns.<sup>97</sup>

---

96 Kieren McCarthy, “When we said don’t link to the article, Google, we mean DON’T LINK TO THE ARTICLE!” The Register, June, 2017, [https://www.theregister.com/2017/06/15/google\\_germany\\_right\\_to\\_be\\_forgotten\\_court\\_case/](https://www.theregister.com/2017/06/15/google_germany_right_to_be_forgotten_court_case/); “Google Wins German Court Case on Links to Sites with Defamatory Content,” DW, accessed May, 2021, <https://www.dw.com/en/google-wins-german-court-case-on-links-to-sites-with-defamatory-content/a-42763265>

97 Eleonora Rosati, “German Court Orders Google to Stop Linking to Lumen Database,” The IPKat, June, 2017, <https://ipkitten.blogspot.com/2017/06/german-court-orders-google-to-stop.html>. Another potential legal challenge to Lumen’s practices comes with Section 512 of the Copyright Act. Section 512 “creates a system for copyright owners and online entities to address online infringement, including limitations on liability for compliant service providers to help foster the growth of internet-based services.” U.S. Copyright Office, “Section 512 Study,” May, 2020.; “Cyberlaw Clinic and Lumen Project Reps Contribute to Section 512 Study,” Berkman Klein Center, April, 2016, <https://www.copyright.gov/policy/section512/>. A former CEO of Copyright Alliance testifying before Congress stated that Lumen’s operations are “repugnant to the purposes of Section 512.” See, “Cyberlaw Clinic and Lumen Project Reps Contribute to Section 512 Study,” Berkman Klein Center, May, 2020, <https://cyber.harvard.edu/node/99449>. This places pressure on

One takeaway is that in planning for future archives, stakeholders must consider how to appropriately protect users' identities when archiving content with human rights value. The net social value of publicizing information that was targeted for removal has also been a topic of debate. Lumen's database of takedown requests is publicly accessible. The information, when inaccurate or misleading, can cause reputational, privacy, or security concerns, especially for those about whom false information was published. Yet the database can also provide critical transparency when wrongdoers attempt to hide their misdeeds from the public by clearing their publicly-searchable reputations.<sup>98</sup>

For the purposes of a human rights-focused archive, the human rights community should consider how catalogs of removed or deleted content may be selected, included and/or accessed in problematic ways. For example, much of Lumen's database can be accessed and used by any member of the public and thus those who may have nefarious purposes. Such practices may raise privacy concerns, especially given the often sensitive or distressing nature of human rights content. For example, one user complained that Lumen was storing a URL from his page that had been de-indexed due to alleged IP concerns. The IP allegations were later proven false

---

Lumen's practices of making public online content available post-takedown. A final issue concerns the protection of users who upload or are featured in archived content. For example, an individual who was suing Google for defamation, malicious falsehood, and breached data protection laws withheld his identity from London's High Court for nine months, reportedly due to fear of being doxxed via the Lumen database. The claimant was concerned that if he was identified through the lawsuit, his personal information would appear in the database. The court ruled that the man must provide his identity to both Google and Lumen for the remedies being sought. See, Gareth Corfield, "Don't Let Google Dox Me on Lumen Database, Nameless Man Begs," *The Register*, August, 2018, [https://www.theregister.com/2018/08/28/nameless\\_man\\_google\\_high\\_court\\_lumen\\_database/](https://www.theregister.com/2018/08/28/nameless_man_google_high_court_lumen_database/).

<sup>98</sup> Ibid.

and his page was reinstated by Google. This reinstatement did not register in Lumen's system and the individual expressed difficulty getting in contact with Lumen's staff to address the issue, as they initially rerouted the individual back to Google and the original sender of the Digital Millennium Copyright Act notice.<sup>99</sup>

One protection that Lumen has implemented to minimize this risk is to work with entities that share notices to display only the information that the party donating the notice elects to share. Lumen also makes a "good faith" effort to redact the information in accordance with local laws. Finally, they have the ability to restrict access to a notice or class of notices to researchers only, or to specific recipients.<sup>100</sup>

Summary:

- **End-uses:** *Lumen archives legal complaints and take down requests so that users can better understand their rights, as well as make removals of online content transparent. Lumen also seeks to protect against the stifling impact of cease and desist notices on online activity.*
- **Who provides the content:** *Lumen receives content from companies that receive requests to remove content from their sites, as well as from those who send such requests.*
- **Who holds the content:** *Lumen archives content in an independent repository.*
- **Who can access the content:** *The public can access and search Lumen's database without fees. The public can also browse larger amounts of data through Lumen's API. To prevent misuse of the API, there is a limit to the number of queries*

---

<sup>99</sup> Peter Kudlacek, "Lumen Database Still Has the URL in Their Database," *Search Console Help*, May, 2019, <https://support.google.com/webmasters/thread/6408158?hl=en>.

<sup>100</sup> Written correspondence with Adam Holland, Project Manager of the Lumen Database.

users can run. For additional access, individuals can request login credentials.

- **Legal obligations:** *Lumen voluntarily preserves and shares archived content.*
- **Unique features:** *Lumen launched a series of partnerships with international institutes to provide expertise on local takedown requests, translate notices into local languages, and facilitate data sharing with non-U.S. entities.*
- **Challenges:** *Lumen has been criticized for making available information that was intended to be removed from the internet. Lumen also raises questions about who should be given access to various information and under what conditions.*

### Terrorist Content Analytics Platform (TCAP)

A second example of the Voluntary Partnership Model is the Terrorist Content Analytics Platform (TCAP). TCAP is a relatively recent initiative hosted by Tech Against Terrorism that was launched with the goal of serving as the first free and secure intelligence-sharing database for online terrorist material. In June 2019, Tech Against Terrorism was awarded a \$1 million grant from Public Safety Canada to support TCAP's creation. The platform has four main goals:

1. To support new apps and smaller tech companies in detecting terrorism-related uses of their services;
2. To facilitate affordable intelligence sharing for smaller internet platforms, and help smaller tech companies expeditiously address terrorist use of their platforms through an alert function;
3. To support secure intelligence sharing between professional researchers and academics, and allow academics and other researchers to improve quantitative analysis of terrorists' use of the internet; and finally

4. To facilitate the coordination of "data-driven solutions" to counter terrorists' use of the internet by making the platform's content available as a training dataset to support the development of automated solutions for detecting extremist content online.<sup>101</sup>

TCAP compiles information from a variety of sources, including aggregators, to create its central repository. Content is uploaded to TCAP based on semi-automated processes. All uploaded content is supposed to undergo a verification process to ensure that the content fits within TCAP's criteria of "terrorist content." According to its website, TCAP stores all content file types, including images, video, audio, text, PDF, and html files.<sup>102</sup> TCAP does not allow any material to be downloaded by users. Additionally, TCAP emphasizes that it will anonymize user data and protect the subject(s) of content on the platform, including faces, names, and other personally identifiable information as compliant with GDPR obligations.<sup>103</sup>

TCAP initially hosted al-Qaeda and ISIS content (limiting the initial information to avoid the "grey areas" of defining terrorist content),<sup>104</sup> but has expanded to include content from designated

---

101 "Tech Against Terrorism in 2019: End of Year Report," Tech Against Terrorism, accessed December, 2020, [https://www.techagainstterrorism.org/wp-content/uploads/2020/04/Tech-Against-Terrorism-2019-End-of-Year-Report.pdf?utm\\_source=Tech+Against+Terrorism&utm\\_campaign=f2fc27f13d-EMAIL\\_CAMPAIGN\\_2019\\_03\\_24\\_07\\_51\\_COPY\\_01&utm\\_medium=email&utm\\_term=o\\_cb464fdb7d-f2fc27f13d-8&utm\\_source=Tech+Against+Terrorism&utm\\_campaign=d7804fffb6-EMAIL\\_CAMPAIGN\\_2019\\_03\\_24\\_07\\_51\\_COPY\\_01&utm\\_medium=email&utm\\_term=o\\_cb464fdb7d-d7804fffb6-141534459](https://www.techagainstterrorism.org/wp-content/uploads/2020/04/Tech-Against-Terrorism-2019-End-of-Year-Report.pdf?utm_source=Tech+Against+Terrorism&utm_campaign=f2fc27f13d-EMAIL_CAMPAIGN_2019_03_24_07_51_COPY_01&utm_medium=email&utm_term=o_cb464fdb7d-f2fc27f13d-8&utm_source=Tech+Against+Terrorism&utm_campaign=d7804fffb6-EMAIL_CAMPAIGN_2019_03_24_07_51_COPY_01&utm_medium=email&utm_term=o_cb464fdb7d-d7804fffb6-141534459).

102 "FAQ," Terrorist Content Analytics Platform, accessed December, 2020, <https://www.terrorismanalytics.org/faq>.

103 Ibid.

104 "Update: Initial Version of the Terrorist Content Analytics Platform to Include Far-Right Terrorist Content," Tech Against Terrorism, July, 2020, <https://www.techagainstterrorism.org/>

violent far-right organizations.<sup>105</sup> When initially interviewed, staff were contemplating using the UN Security Council Sanctions List as a proxy for defining “terrorist” content. Subsequently, they let us know that in order to ground the TCAP in the rule of law, terrorist organizations’ inclusion is now based on designation by democratic nation states and supranational institutions, as further explained in their website’s group inclusion policy.

TCAP plans to use machine learning-based tools to detect and archive terrorist content quickly, and help social media companies analyze that content based on their terms of service.<sup>106</sup> The TCAP team spent October through December of 2019 conducting an online public consultation process to better understand potential user requirements and areas of concern. As of November 2020, the Tech Against Terrorism coalition was expanding its expertise on the use of end-to-end encryption to better grasp how online users and the general public perceive encryption practices.<sup>107</sup>

TCAP’s database is designed to be accessible to “tech companies, academics [and] civil society.”<sup>108</sup> At

the time of our research, staff were working on developing a streamlined process to make voluntary uploading of content to TCAP as easy as possible, as well as to simplify the categorization and retrieval processes.<sup>109</sup>

Summary:

- **End-uses:** *TCAP’s goal is to serve as a free and secure intelligence-sharing database for online terrorist material for academic purposes, and to support tech companies in detecting terrorism-related uses of their services.*
- **Who provides the content:** *TCAP collects terrorist content from open sources and will accept information from external contributors in the future.*
- **Who holds the content:** *Content is stored by TCAP.*
- **Who can access the content:** *Tech companies, academic researchers, and civil society representatives with justification for accessing terrorist content can be granted access to the database. Tech companies include and vary from social media companies to e-commerce and email service providers. Access to and use of TCAP is free for smaller platforms and academics.*
- **Legal obligations:** *TCAP is a voluntary initiative developed by Tech Against Terrorism and supported by Public Safety Canada.*
- **Unique features:** *TCAP is purportedly being developed in compliance with Canadian, U.K., U.S., and E.U. legislation with respect to security and privacy laws, including GDPR. All content uploaded and stored to TCAP undergoes a verification process to ensure*

---

2020/07/02/update-initial-version-of-the-terrorist-content-analytics-platform-to-include-far-right-terrorist-content/.

105 Far-right organizations include The Base, Atomwaffen Division (also known as the National Socialist Order), Proud Boys, and the Russian Imperialist Movement (RIM); see, “January 2021 update,” Terrorist Content Analytics Platform, January, 2021, <https://www.terrorismanalytics.org/blog/tcap-newsletter-january-2021>.

106 “Update: Initial Version of the Terrorist Content Analytics Platform to Include Far-Right Terrorist Content,” Tech Against Terrorism, July, 2020, <https://www.techagainstterrorism.org/2020/07/02/update-initial-version-of-the-terrorist-content-analytics-platform-to-include-far-right-terrorist-content/>; “Group Inclusion Policy,” Terrorist Content Analytics Platform, accessed April, 2021, <https://www.terrorismanalytics.org/group-inclusion-policy>.

107 “November Update,” Tech Against Terrorism, November, 2020, <https://www.techagainstterrorism.org/2020/12/04/november-update/>.

108 “FAQ,” Terrorist Content Analytics Platform, accessed December, 2020, <https://www.terrorismanalytics.org/faq>.

---

109 This overview was informed by public documents and an interview with TCAP’s Research Manager Jacob Berntsson and Product Manager Tom Lancaster on March 25, 2020.



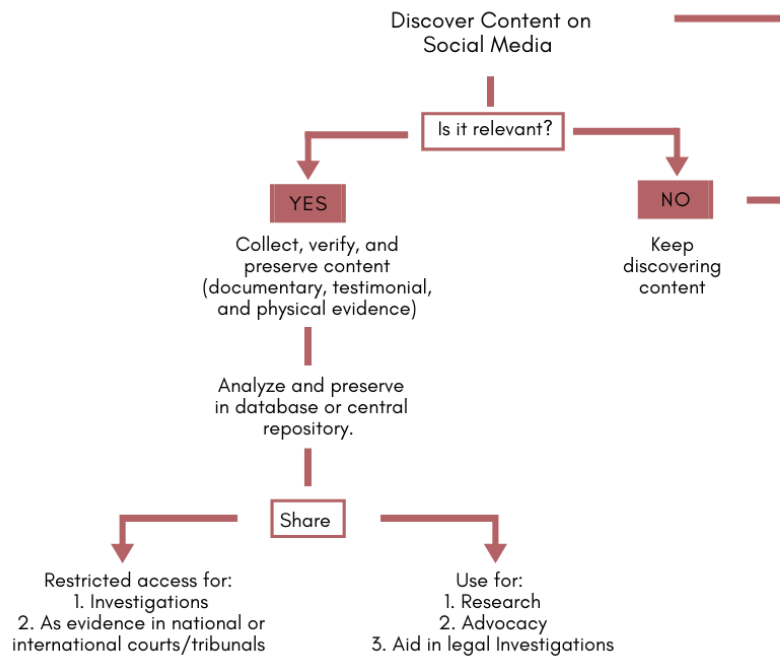
that information meets the criteria for being labelled as terrorist content.

- **Challenges:** *TCAP is addressing how to make their content classification and verification procedures more effective and efficient.*

The Voluntary Partnership Model, as exemplified by both Lumen and TCAP, highlights the variety of content that can be collected with voluntary

participation from tech companies—from analytical products to legal documents and to the underlying content—and made available to a broad public without access fees. Both raise significant questions, however, regarding privacy, potential legal challenges to the storage and publication of user content, long-term funding, and how to best define the parameters of online repositories.

## THE INDEPENDENT COLLECTION MODEL



### Mnemonic/The Syrian Archive

Mnemonic/The Syrian Archive illustrates how non-profit, non-governmental organizations can independently pull and preserve content from social media. The Syrian Archive is a project of Mnemonic, an independently run organization founded in 2014

by Hadi al-Khatib and Jeff Deutch.<sup>110</sup> The Syrian Archive holds visual documentation of human rights abuses committed by all actors involved in

<sup>110</sup> “Home Page(En),” Syrian Archive, accessed October, 2020, <https://syrianarchive.org/en>.

the Syrian conflict. All collected content is stored on private servers, with a subset of that content hosted in a database accessible via the Syrian Archive website. The Syrian Archive treats all content on its servers as the property of those who filmed or shared it. At the time of writing, the Syrian Archive has collected nearly three and a half million pieces of digital content.<sup>111</sup>

The organization is donor-funded; however, they do not accept funding from governments that are implicated in the Syrian conflict. Since its founding, the group has partnered with international human rights organizations such as the United Nations Office of the High Commissioner for Human Rights, Human Rights Watch, and Amnesty International, among others. With regards to their purpose, “The Syrian Archive aims to preserve data as digital memory, to establish a verified database of human rights violations, and to act as an evidence tool for legally implementing justice and accountability as concept and practice in Syria.”<sup>112</sup> The organization’s end goals are historical preservation, generation of evidence for future legal cases, information sourcing for journalists, and raising awareness to support Syrian human rights activists. The organization maintains a systematic process of data collection, preservation, verification, and publication—a process that can serve as a model for a future evidence locker that deploys mass collection of social media content, as well as the authenticity, reliability, and long-term preservation of such content.

The Syrian Archive collects content in two ways. First, staff scrape content from social media platforms and other websites daily. Second, the Syrian Archive works with a network of trusted sources, including journalists, lawyers, media groups, and human rights activists, and preserves their material. Before categorizing any collected content as

published from a trusted source, the Syrian Archive employs a strenuous vetting process to ensure they are working with credible individuals and entities that align with their organization’s mission and vision.

Once the content is collected, it is preserved on a secure server with offline backups. From there, staff begin the verification and analysis process. Content is cataloged and organized around metadata (e.g. time, date and location of upload, camera details, etc.).

In order to facilitate the use of such data for accountability purposes, and “given that there are no universally accepted, legally admissible metadata standards,” the Syrian Archive developed their own framework for organizing data in consultation with international investigative bodies.<sup>113</sup> The team has worked with members of the United Nations Office of the High Commissioner for Human Rights, intergovernmental mechanisms, other archives, and human rights and research organisations.<sup>114</sup>

The Syrian Archive strives to create a credible model that includes an authentication process for new sources. They are efficient at preserving content and have standard operating procedures for verification and categorization. The model they use to collect, verify, and preserve digital content related to Syria has been successfully replicated, albeit at a smaller scale, with their newer Yemeni Archive<sup>115</sup> and Sudanese Archive.<sup>116</sup> They are also launching an archive for Chile with the Chilean organization Testigo en Línea (“Online Witness”).<sup>117</sup> Support for and uptake of the Syrian Archive’s methodology is

---

111 Ibid.

112 “About(En),” Syrian Archive, accessed October, 2020, <https://syrianarchive.org/en/about>.

---

113 “Methods and Tools,” Syrian Archive, accessed January, 2021, <https://syrianarchive.org/en/about/methods-tools>.

114 Ibid.

115 “Home Page(En),” Yemeni Archive, accessed October, 2020, <https://yemeniarchive.org/en>.

116 “Home Page(En),” Sudanese Archive, accessed October, 2020, <https://sudanearchive.org/en>.

117 “Recursos, Recopilación y denuncia sobre violaciones a los DD.HH. en Chile,” Testigo en Línea, accessed October, 2020, <https://testigoenlinea.cl/recursos.html>.

an indication that civil society and human rights activists alike believe in the accessibility, transparency, and verification processes of this digital archive model, and can effectively cooperate towards diverse human rights ends.

Ultimately, the Syrian Archive offers insight into a non-profit and non-governmental organization that acts as both the primary collector and preserver of digital content specific to a particular crisis. Their process demonstrates flexibility in receiving content from non-social media sources, including non-governmental organizations and activists, as well as the flexibility to communicate with social media platforms in order to have accounts reinstated when taken down—although anecdotally that process has suffered from a lack of systematicity with regards to platform responsiveness.

Though the Syrian Archive’s model has aspects that should be considered “best practice,” it may be difficult to scale due to the need for in-house tech capacity to appropriately manage content. Much of its success relies on a robust grassroots movement, and it is much more efficient at preserving content than verifying and publishing, given the necessary resources and time constraints intrinsic to the latter. This is highlighted by the banner at the bottom of their homepage, which cites over 3.5 million pieces of digital content preserved, with only eight thousand of those pieces verified.<sup>118</sup> They also suffer from funding limitations, underscoring the need for sufficient and reliable long-term resourcing. Despite this, if the archival focus is on collection and aggregation, and doesn’t require verification and publication, this is an especially interesting model to consider and/or adopt.

#### Summary

- **End-uses:** *Historical preservation, creation of a digital memory about the Syrian conflict, and evidence provision for legal accountability.*

---

118 “Home Page(En),” Syrian Archive, accessed October, 2020, <https://syrianarchive.org/en>.

- **Who provides the content:** *Syrian Archive staff scrape content from social media platforms and other websites daily. The Archive also receives content from a trusted network of journalists, lawyers, media groups, and human rights activists.*
- **Who holds the content:** *Content is stored in private servers with offline backups, with a subset of content hosted in a database on Syrian Archive’s website.*
- **Who can access the content:** *The public can freely access all other content that has been preserved and verified on the Syrian Archive website.*
- **Legal obligations:** *The Syrian Archive is a non-profit, non-governmental organization and thus subject to national, regional and international laws.*
- **Unique features:** *Syrian Archive is a Syrian led project that treats all content on its servers as the property of those who filmed or shared it. The archive also maintains a systematic process of data collection, preservation, verification, and publication.*
- **Challenges:** *Given their strenuous verification process, the Syrian Archive has only been able to authenticate a small portion of all preserved content. The Archive also requires relatively advanced technological proficiency, and suffers from funding and resource limitations.*

#### Jihadology

Jihadology is a non-profit, non-governmental academic project that, similar to the Syrian Archive, downloads and stores digital content. Jihadology initially began collecting online content from password-protected forums, and between 2012-2016 primarily collected information from social media platforms. Today, Jihadology mainly downloads

content from encrypted messaging applications like Telegram or decentralized servers.<sup>119</sup>

Jihadology.net was founded by Aaron Zelin, a researcher who focuses on Sunni Arab jihadi groups in North Africa and Syria, as well as online Jihadism, Jihadi governance and the trend of foreign fighting. The website claims to hold the largest collection of primary source online jihadi material and propaganda—including videos, posters, and speeches. The website is run by founder and volunteer Zelin, with most of the content accessible to those who register with some form of legitimate institutional email address. The majority of the content is related to jihadi actors irregardless of group affiliations. Some content contains violent and graphic materials, such as executions.<sup>120</sup>

When first established, the audiovisual and textual information uploaded by Zelin could be accessed by the public without any formal request process. The website is hosted by the U.S. tech group Automatic (which owns WordPress), and does not receive any money through advertising.<sup>121</sup> Jihadology.net's only source of revenue, which amounts to a sporadic few hundred dollars at a time, comes from translation services offered by Zelin for the material that is uploaded and preserved on the site. The archive is fairly rudimentary and functions as a basic catalog, with content coded by the date and title of the extremist groups to which the content belongs (e.g.

ISIS, al-Shabab, al-Qaeda), media groups, and relevant country identification.

The website has arguably been helpful in facilitating academic and journalistic research that relies on jihadi materials, as well as research related to the use of online platforms by extremist groups like ISIS. Journalists, researchers, government analysts, and academics have all used the material for research and intelligence gathering.<sup>122</sup> The website's founder has previously run a Jihadology podcast that was posted to his website, published to a range of podcast platforms, and reposted by the well-known national security blog Lawfare. In the podcast, Zelin analyzes a portion of the content uploaded to his website and has guests discuss their research related to the jihadi movement. However, there has been some concern that Jihadology.net gives the public access to sensitive material that includes extremist content, violence and/or killings—a subset of which may be material social media platforms voluntarily or are legally required to remove.

In addition to the distress that viewing such content can cause, a key concern by security experts is that the website could facilitate extremism. There has been at least one documented case of an extremist using the content. For example, Omar Mateen, the perpetrator of the Pulse nightclub shooting in Orlando, Florida, that killed 49 individuals, was known to have scrolled through Jihadology before his attack.<sup>123</sup> As a result of these concerns, in 2017 the British government and intelligence officials called for the website to either remove extremist content or secure the content from potential misuse through password protections. Founder Zelin has also received multiple requests from the UK

---

119 Aaron Zelin, "The Case of Jihadology and the Securitization of Academia," *Terrorism and Political Violence* 33, no.2 (2021), <https://doi.org/10.1080/09546553.2021.1880191> For more of Zelin's work, see *Your Sons are at Your Service: Tunisia's Missionaries of Jihad* (New York: Columbia University Press, 2020).

120 Graeme Wood, "Don't Shut Down the Internet's Biggest Jihadist Archive," *The Atlantic*, December, 2018, <https://www.theatlantic.com/ideas/archive/2018/12/dont-shut-down-internets-biggest-jihadi-archive/577630>.

121 David Bond, "Library of Jihadi Material to Restrict Access to Most Sensitive Material," *The Financial Times*, April, 2019, <https://www.ft.com/content/55338012-5ae0-11e9-9dde-7aedca0a081a>.

---

122 Wood, "Don't Shut Down the Internet's Biggest Jihadist Archive."

123 "Archiving Terrorist Propaganda." WNYC Studios, March, 2019, <https://www.wnycstudios.org/podcasts/otm/segments/archiving-terrorist-propaganda-jihadology>.

government to shut down the website completely.<sup>124</sup> This transpired into a larger debate between civil society groups, governments, journalists, and other activists as to whether potentially dangerous and violent content should ever be accessible to a general public.<sup>125</sup>

To relieve some of these concerns, Tech Against Terrorism<sup>126</sup>—a UN-backed organization that works with tech companies to mitigate the spread of extremist content—partnered with Jihadology to restructure the website such that all content from primary sources are password-protected. These re-

strictions require that users who are interested in accessing the most sensitive content, such as execution videos, prove they are affiliated with an academic institution, government institution, humanitarian organization, or news organization via an email address. “Whitelisted” email domains with a legitimate institutional affiliation, such as *.edu*, *.gov*, and *.mil*, are automatically granted access.<sup>127</sup> Those planning and crafting a future evidence locker who are contemplating a similar whitelisting process should discuss and consider whether this is the most viable way forward or whether there are other options for allowing access to sensitive material. In the case of Jihadology, third-party researchers who do not have “whitelisted” email domains can obtain permission to access the password-protected material on a case-by-case basis, by emailing Aaron Zelin. The public is still able to access the analysis portion of Jihadology’s website—just not the underlying data unless they have registered with a whitelisted email domain.

Overhauling the website to include this user registration system cost about 100,000 pounds and was funded by the Global Internet Forum to Counter Terrorism. Jihadology remains the sole administrator of the website and no user registration information is forwarded to any person or entity, including government agencies.

Despite the many challenges of preserving and providing access to potentially sensitive content, there are also numerous benefits to making such content available.<sup>128</sup> Ultimately, Jihadology presents

---

124 Shirin Jaafari, “British Parliament Wants to Shut down Extremist Content Online — at What Cost?,” *The World from PRX*, December, 2018, <https://www.pri.org/stories/2018-12-14/british-parliament-wants-shut-down-extremist-content-online-what-cost>.

125 There have been a few studies gauging how often individuals or groups link to the Jihadology website. One study analyzed mentions of the word ‘Rumiyah’ (the Islamic State official English language magazine) between November 1, 2016 to October 31, 2017 on Twitter and found that “Jihadology does not appear in one the top 20 domains that have outlinks... [And] only 5 [out of 892 distinct links] linked directly to Jihadology.” See, Zelin. “The Case of Jihadology and the Securitization of Academia,” 234.

A second study analyzing “636 pro-IS Telegram channels and groups that contain English-language content collected between June 1, 2017 and October 24, 2018... Jihadology does not appear in the top 20 domains that were outlinked to.” The researchers told Zelin that Jihadology ranked 131st with 13 Jihadology files shared out of the 101, 773 pieces of content disseminated on Telegram. See, Zelin. “The Case of Jihadology and the Securitization of Academia,” 235. A third study of Arabic language Telegram accounts between January-May 2019 shows that Jihadology is neither in the top 25 domains of overall links from a specific domain, nor is it amongst the top 33 links shared in the study.” See, Zelin. “The Case of Jihadology and the Securitization of Academia,” 234.

126 Tech Against Terrorism’s funding consists of contributions from companies and government. In 2017 the project was supported by the Government of Switzerland, the Republic of Korea, Facebook, Microsoft, Google, and Telefonica. See, “Frequently Asked Questions (FAQ),” Tech Against Terrorism, accessed January, 2021.

---

127 Bond. “Library of Jihadi Material to Restrict Access to Most Sensitive Material.” According to Jihadology founder Zelin, “as of June 30, 2020 there [were] 1,019 whitelisted domains from institutions all over the world and 2,631 registered users.... [Additionally, there have been] 294 registrations [which] have been rejected due to... attempts” such as “trying to trick [the website] into thinking they have a legitimate domain when in fact it is actually a fake.” See, Zelin. “The Case of Jihadology and the Securitization of Academia,” 237.

128 *Ibid.*, 225-241.

stakeholders interested in developing an archive for human rights purposes a glance into the variety of content feasible for independent preservation, potential end-uses of such a registry, and options for allowing access by various members of the public. Additionally, this case study raises questions about the efficacy and limitations of built-in security features that may be appropriate for a future evidence locker(s).

Summary:

- **End-uses:** *Jihadology hosts content for research and other academic purposes.*
- **Who provides the content:** *Jihadology downloads content from social media sites focused on online jihadism and jihadi governance.*
- **Who holds the content:** *Jihadology holds and publishes the content on its website.*
- **Who can access the content:** *Secondary sources and features are accessible to the public. All other material is password-protected. Only those with legitimate institutional emails, for example academic, government, or news organization email domains, are granted access to password-protected content. Others may email Aaron Zelin to request access on a case-by-case basis.*
- **Legal obligations:** *Jihadology is a non-profit, non-governmental organization that voluntarily collects and preserves content.*
- **Unique features:** *The website partners with Tech Against Terrorism to mitigate the spread of extremist content and safeguard sensitive material.*
- **Challenges:** *The website was controversial in its initial format (pre-password protection), with some critics arguing that broad access to extremist content could facilitate extremism. This case study also raises insights into the possibilities and limitations of disparate security practices.*

## ICRC Archives

The final example of the Independent Collection Model are the archives held by the International Committee of the Red Cross (“ICRC” or “Committee”)—an independent organization dedicated to protecting those affected by armed conflict and other violence. The ICRC is subject to international humanitarian law (IHL) and related regulations. The Committee abides by the principles of humanity, impartiality, neutrality, independence, unity, universality and voluntary service, which guide access to and operation of its archives.

The ICRC Archives consist of preserved “ICRC documents dating from [1863] to present day” that have been made “available for research.”<sup>129</sup> The historical records “comprise 6,700 linear metres of textual records and a collection of photographs, films and other audio archives.”<sup>130</sup> The ICRC considers its Archives as “living;” they are mandated to preserve the history of IHL, the work of the International Red Cross and Red Crescent Movement, the history of warfare, the history of captivity in war, the history of war victims, and the preservation of the memories of those assisted by the organization.<sup>131</sup>

Given its role in the collection and preservation of documents, photographs, videos, and auditory material over 158 years, the institution and its archives have undergone substantial changes—from being focused around thematic and geographic filing plans, to indexing documents for computerized

---

129 “ICRC Archives.” International Committee of the Red Cross, accessed March, 2021, <https://www.icrc.org/en/archives>.

130 Ibid.

131 Valeria McKnight Hashemi, “A balancing act: The revised rules of access to the ICRC Archives reflect multiple stakes and challenges,” *International Review of the Red Cross*, April, 2018, 376 and 378, <https://international-review.icrc.org/articles/balancing-act-revised-rules-access-icrc-archives-reflect-multiple-stakes-and-challenges>.

access, and finally adopting an electronic filing system in 2010.<sup>132</sup>

To date, the ICRC preserves all types of information, including personal data, and has developed a comprehensive framework of rules for that information's protection.<sup>133</sup> Given the sensitive nature of content pertaining to individuals and families in sites of conflict, the ICRC Archives offer key insights regarding potential access. Beginning in 1925, and still practiced today, all materials considered especially sensitive are "stored in a safe at the ICRC headquarters."<sup>134</sup> The ICRC Archives' most recent revisions to its access rules occurred in 2017. The ICRC classifies its documents on a scaling system from strictly confidential, to confidential, internal, and public.<sup>135</sup> Its 2017 guidelines outline the following:<sup>136</sup>

1. Content preserved in the ICRC General Archives can now be consulted by the public 50 years after their preservation date;
2. Materials containing personal data can be consulted by the public 70 years after their preservation date;
3. The ICRC retains the right to extend the protection period for documents containing information whose disclosure may violate the protection of personal data or jeopardize individuals' safety;
4. Information about those who are the subject of ICRC protection or concern is made available to

family members via the Central Tracing Agency, which archives and transmits information related to missing persons, prisoners of war, and civilian internees;

5. The ICRC will grant special access to certain materials in the Archives for research purposes prior to official release based on specified conditions and with protections for personal data;
6. The ICRC prohibits any use of the Archives that threatens the dignity and/or physical or mental integrity of individuals;
7. The ICRC is not legally obligated to share its archives with the public, but is committed to publicizing its preserved material after the designated protection periods given the historical value such content may offer to communities.

The independent collection, analysis, and preservation of content by the ICRC is particularly useful in thinking about preservation practices through the lens of privacy. More specifically, the organization has been debating the right to be forgotten and the right to request rectification of one's personal data in open files and archives.<sup>137</sup> This translates to practices of not sharing "information forming part of bilateral and confidential dialogue with authorities and parties to armed conflict in the context of legal proceedings."<sup>138</sup> Confidential ICRC material cannot be used in legal proceedings; ICRC staff cannot be compelled to testify.<sup>139</sup> This case study offers an important perspective for stakeholders seeking to prioritize and safeguard privacy, historical memory, and an authoritative source of knowledge and

---

132 Ibid., 380.

133 Ibid., 377. For the ICRC "personal data" refers to identifiers such as name, audiovisual material, an identification number, location data, online identifiers, or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of a data subject. It also includes data identifying or capable of identifying human remains.

134 Hashemi. "A balancing act: The revised rules of access to the ICRC Archives reflect multiple stakes and challenges," 380. For a thorough analysis into its public access practices beginning in 1925 to 2016, see pages 380 to 386.

135 Ibid., 388.

136 Ibid., 392-393.

---

137 "The ICRC published its Handbook on Data Protection in Humanitarian Action in 2017," guidelines wish it uses to inform its preservation practices. See Ibid., 389.

138 Ibid., 387.

139 Ibid. The ICRC does not have control over the use of content which is made public.

learning for material emerging from armed conflicts and contexts of violence.<sup>140</sup>

Summary:

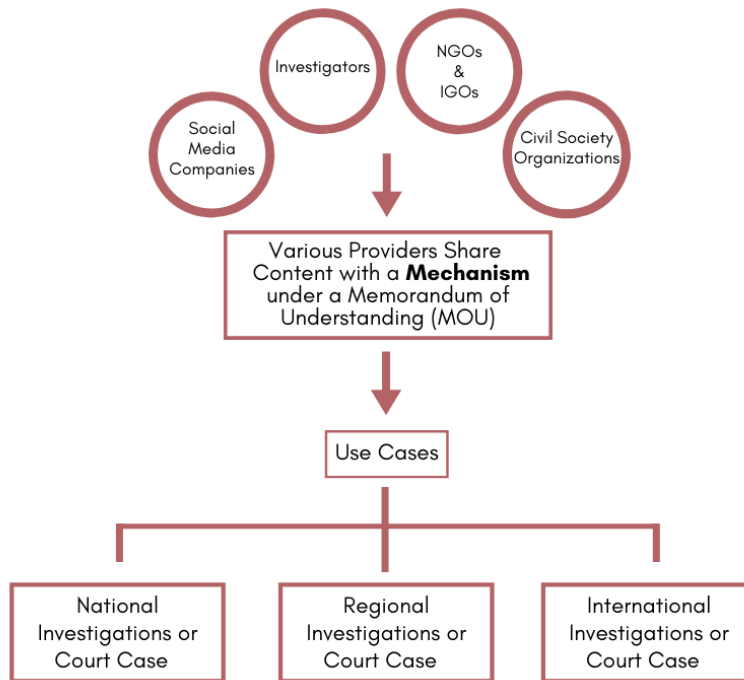
- **End-uses:** *The Archives preserve the history of IHL, the work of the ICRC, the history of warfare, the history of captivity in war, the history of war victims, and information about those assisted by the organization.*
- **Who provides the content:** *The ICRC independently collects its own content, including text-based records, photographs, films and audio files.*
- **Who holds the content:** *The ICRC holds the most sensitive material in a safe at its headquarters in Geneva, along with some of its content in the ICRC Library, and in an online repository.*
- **Who can access the content:** *Materials in the Archives are classified as strictly confidential, confidential, internal, or public, with restricted access for most material. After a 50-70 year protection period, materials are declassified and accessible by the public. Those seeking authorization to access content for research purposes in advance of the end of the protection period may submit access requests.*
- **Legal obligations:** *The ICRC Archives are not legally obligated to collect, preserve, or share content with any external entity.*
- **Unique features:** *The Archives are especially careful about privacy, and are also grappling with how to respond to the right to be forgotten and the right to request rectification of one's personal files. Individuals who are the subject of collected documents and their family members may request access to their files through the Central Tracing Agency.*
- **Challenges:** *Most material is declassified after 50-70 years, which limits the use of that content in the decades immediately following conflict.*

---

140 Hashemi. "A balancing act: The revised rules of access to the ICRC Archives reflect multiple stakes and challenges," 376.



## THE HYBRID MODEL



### International, Impartial and Independent Mechanism (IIIM)-Syria

The International, Impartial and Independent Mechanism to Assist in the Investigation and Prosecution of Persons Responsible for the Most Serious Crimes under International Law Committed in the Syrian Arab Republic since March 2011 (hereinafter “the Mechanism,” “IIIM” or “IIIM-Syria”) illustrates the “Hybrid” archive model, which differs from the repositories discussed above in that it is an intergovernmental organization that aggregates content held by the social media companies and diverse civil society organizations to store and package that information as evidence for a variety of potential end users across a range of jurisdictions. The IIIM is legally mandated to carry out investigative and preservation duties. The IIIM—established in December 2016 by the UN General Assembly<sup>141</sup>—is

responsible for investigating the most serious crimes committed in Syria since March 2011.<sup>142</sup> They are required to “to collect, consolidate, preserve and analyse evidence of violations” and also “to prepare files to facilitate and expedite fair and independent criminal proceedings in national, regional or international courts, in accordance with international law.”<sup>143</sup> The Mechanism previously received voluntary contributions from supporting UN member states and now receives regular funding through the United Nations’ budget.<sup>144</sup>

The Mechanism is not a tribunal, court, or prosecutor’s office, but an independent body that collects,

---

overview of current mechanisms for aggregating digital and other evidence of atrocities).

142 “Frequently asked questions,” The International, Impartial and Independent Mechanism, accessed January, 2021, <https://iiim.un.org/faq/>.

143 Ibid.

144 “Frequently asked questions,” The International, Impartial and Independent Mechanism, accessed January, 2021, <https://iiim.un.org/faq/>.

---

141 See, Beth Van Schaack, *Imagining Justice for Syria*, (New York: Oxford University Press, 2020), 339-396 (providing an

analyzes, preserves, and shares information with legal actors for investigation and prosecution purposes.<sup>145</sup> The Mechanism does not solely rely on content from social media platforms, but serves as an important example of a repository that aggregates social media content with other content, and whose collection and preservation processes feed into international criminal law processes.

This case study is relatively unique given its role in collaborating with civil society organizations. The Mechanism, for example, receives content from organizations such as the “Independent International Commission of Inquiry on the Syrian Arab Republic, the Joint Investigative Mechanism, States, international or regional organizations, entities of the United Nations system, non-governmental organizations, foundations and individuals.”<sup>146</sup> The Mechanism also collects its own content, including forensic material, witness testimony, and documentary information.<sup>147</sup> Because Mechanism personnel are “barred from entering Syria by the Syrian government, it has focused on the collation of digital documentation,” including the “signing of a Protocol of Cooperation with Syrian civil society groups that collect digital evidence.”<sup>148</sup>

Stakeholders interested in planning and establishing an evidence locker for human rights-related purposes can gain valuable lessons from this body. As of August 2020, the Mechanism had “created a central repository of information and evidence that holds more than 2 million records.”<sup>149</sup> To assist in

the processing of content, the Mechanism adopted two software programs to enhance audio and video review capabilities, in addition to “recognizing and managing entities, such as locations and incidents.”<sup>150</sup> Other tools that help organize and analyze content within the central repository use object, glyph, and Arabic character recognition, and identify duplicate videos.<sup>151</sup> To aid analysis of collected materials, the Mechanism implements user-friendly analytical tools that “encompass intelligence-oriented reports, legal research, litigation-oriented briefs and visual products such as organization charts and timelines.”<sup>152</sup>

A key feature of the Mechanism is its adoption of relatively flexible frameworks for potentially sharing material and cooperating with other organizations, including civil society actors. To facilitate formal and informal cooperation, they have relied on “verbal agreements, exchanges of letters, memorandums of understanding, protocols and bilingual frameworks.”<sup>153</sup> These diverse engagements can be emulated by others who seek to archive and share content for various end uses, although notable is its structure as an intergovernmental organization that may be free from the constraints of many national laws.

In tension with broader social accountability and transparency, the Mechanism is not expected to “publicly report on its substantive work,” and only shares a report to the “General Assembly twice a

---

145 Ibid.

146 Ibid.

147 Van Schaack, *Imagining Justice for Syria*, 366-372.

148 Rebecca Hamilton, “Social Media Platforms in International Criminal Investigations,” *Case Western Reserve Journal of International Law* 52, no.1, (2020), 217, <https://scholarlycommons.law.case.edu/cgi/viewcontent.cgi?article=2571&context=jil>.

149 United Nations. “International, Impartial and Independent Mechanism to Assist in the Investigation and Prosecution of Persons Responsible for the Most Serious Crimes under International Law Committed in the Syrian Arab Republic since March 2011,” General Assembly Seventy-fifth Session, August,

---

2020, 3, <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/210/85/PDF/N2021085.pdf?OpenElement>.

150 Ibid., 6.

151 Ibid. See also Benetech’s Justice AI tool, which is designed to deduplicate and cluster videos of atrocity.

152 United Nations. “International, Impartial and Independent Mechanism to Assist in the Investigation and Prosecution of Persons Responsible for the Most Serious Crimes under International Law Committed in the Syrian Arab Republic since March 2011,” 3, <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/210/85/PDF/N2021085.pdf?OpenElement>.

153 Ibid.

year on the implementation of its mandate.”<sup>154</sup> That is, the central repository is not made available to the public. Thus, while a potentially interesting model for evidentiary purposes, it may be unsuitable for other end uses, including social science research or developing historical narratives of conflict.

Summary:

- **End-uses:** *Collect, analyze, and preserve evidence to facilitate and expedite independent criminal proceedings in national, regional, or international courts.*
- **Who provides the content:** *The IIMM receives content from some social media companies as well as investigators, civil society organizations, and others. The Mechanism also collects its own content, including forensic material, witness testimony, and documentary information.*<sup>155</sup>
- **Who holds the content:** *The Mechanism.*
- **Who can access the content:** *The Mechanism only reports to the General Assembly. Information can be shared with carefully vetted users for legal investigation and prosecutorial purposes.*
- **Legal obligations:** *The Mechanism is legally mandated to collect, consolidate, preserve, and analyze evidence of violations for criminal proceedings.*
- **Unique features:** *The Mechanism uses diverse digital tools to organize and analyze content in its repository. The Mechanism also cooperates with intergovernmental, international and national organizations.*
- **Challenges:** *The general public does not have access to the Mechanism’s central repository, which makes it an unsuitable model for social science*

*research and those developing historical narratives of conflict.*

## Independent Investigative Mechanism for Myanmar (IIMM)

The final Hybrid case study is the Independent Investigative Mechanism for Myanmar (“IIMM”). The United Nations Human Rights Council established the IIMM in September 2018 with a mandate to collect, analyze, preserve, and share the most serious international crimes and violations in Myanmar since 2011. The IIMM prepares files to facilitate independent criminal proceedings in national, regional, and international courts or tribunals with potential jurisdiction over the crimes in accordance with international law and standards.<sup>156</sup>

The IIMM collaborates with the Independent Fact-Finding Mission on Myanmar for evidence collection. In addition to interviews with victims and witnesses, IIMM staff collect content from open sources such as the internet, news, social media, and public reports.<sup>157</sup> The IIMM also collects other types of evidence such as audio-visual, digital, electronic, geospatial, and forensic material.<sup>158</sup> To ensure admissibility of content for criminal proceedings in a court or tribunal, the IIMM follows relevant international law and jurisprudence. Once material has been collected, the Mechanism stores the material for accountability efforts in future cases. Similar to the IIMM-Syria, the IIMM reviews and analyzes all material it gathers and preserves that material in a secured online database. To facilitate chain of custody, reliability and probative value, the IIMM

---

<sup>154</sup> “Frequently asked questions,” The International, Impartial and Independent Mechanism, accessed January, 2021. <https://iiim.un.org/faq/>.

<sup>155</sup> Ibid.

---

<sup>156</sup> “Homepage,” The International, Impartial and Independent Mechanism, accessed January, 2021, <https://iiim.un.org>.

<sup>157</sup> “Evidence Collection and Case Building,” Independent Investigative Mechanism for Myanmar, accessed January, 2021, <https://iiim.un.org/evidence-collection-and-case-building/>.

<sup>158</sup> Ibid.

abides by protocols “consistent with the UN Charter, UN rules, regulations, policies and good practices, relevant international law and jurisprudence.”<sup>159</sup>

With respect to sharing collected information, the IIMM obtains assurance from external entities that they will “respect the scope of consent of the sources of information, victims, witnesses, governments and non-governmental providers of information.”<sup>160</sup> The IIMM shares information for criminal proceedings and related purposes to advance justice and deter further crimes when “the safety and privacy of victims and witnesses are assured.”<sup>161</sup>

Similar to the IIM-Syria, the IIMM consists of “impartial and objective” discoverers, collectors, preservers, and providers of information. Both mechanisms are considered intergovernmental organizations given that they are situated within the United Nations, which is comprised of various countries, and neither accept government or other external instructions or interference with their duties. Both institutions are particularly useful for thinking through which national and international guidelines and protocols can be followed to ensure appropriate chain of custody, protect victims and witnesses, and still share material with external parties for legal purposes—although through models that remain free of the constraint of most national legal frameworks.

Summary:

- **End-uses:** *Collect, analyze, preserve and package evidence to facilitate and expedite independent criminal proceedings in national, regional, or international courts.*
- **Who provides the content:** *The IIMM collects content from diverse sources on the internet, news and human rights organizations, social media, and public reports. The mechanism also obtains information through interviews with victims and witnesses, and by collecting forensic and geospatial material.*<sup>162</sup>
- **Who holds the content:** *The IIMM reviews, analyzes, and preserves material in a secure database and abides by international protocols to maintain appropriate chain of custody.*
- **Who can access the content:** *The IIMM only shares content with actors pursuing criminal proceedings and external entities that respect the scope of consent.*
- **Legal obligations:** *The IIMM is legally mandated to collect, preserve, and analyze evidence of violations for criminal proceedings.*
- **Unique features:** *The IIMM spotlights useful national and international protocols that can be followed to ensure appropriate chain of custody, protect victims and witnesses, and share material with external parties while maintaining impartiality.*
- **Challenges:** *As an intergovernmental organization, this model is specific to that organization. Each intergovernmental organization has a different mandate provided by its creating body and thus is not easily replicated in disparate contexts.*

---

<sup>159</sup> Ibid.

<sup>160</sup> “Information Sharing,” Independent Investigative Mechanism for Myanmar, accessed January, 2021, <https://iimm.un.org/information-sharing/>.

<sup>161</sup> “Homepage,” Independent Investigative Mechanism for Myanmar, accessed January, 2021, <https://iimm.un.org>.

---

<sup>162</sup> “Evidence collection and case building,” Independent Investigative Mechanism for Myanmar, accessed January, 2021, <https://iimm.un.org/evidence-collection-and-case-building/>.

# BACKGROUND

## Part III – Legal, Technical, and Operational Challenges

Any model that is adopted, adapted, or designed to hold social media content with human rights value will face significant legal, technical, and operational challenges. A digital locker and the stakeholders involved in its creation, use, and maintenance will likely have to comply with applicable national and international laws, some of which may be unclear, inconsistent, or in the process of amendment.

### DEFINING TERMS AND SCOPE

While a thorough overview of relevant law is outside the scope of this report, this section highlights three key issues that must be addressed for any sort of mechanism to work. First, there will need to be a delineated end-user group with clarity on who and who does not fall within that group. This includes the type of organization (intergovernmental, non-governmental, academic, etc.) as well as the scope of its work (human rights, journalism, criminal investigation, social science research, etc.) Second, the relevant social media content must be clearly defined, as mentioned earlier. Third, stakeholders will need to establish what is meant by “preservation” of and “access” to that content, since the various archives above have conceptualized and operationalized both concepts quite differently.

### End-User Group

The end-user group refers to the individuals and organizations that can request preservation of content, as well as anyone who is permitted to access the content in the locker. An establishing body will have to determine and define the categories of people who might be permitted to request preservation and/or access, and what content falls within the scope of interest. In the legal evidence context, limiting access to judicial bodies would minimize the difficulty of reconciling any future mechanism with current law. Similarly, limiting privileges to request and/or access the content to those vested with investigative authority by an intergovernmental organization, and who have a clear mandate with a well defined scope, will not be as difficult legally or operationally as granting privilege to human rights non-governmental organizations, researchers or academics, or members of the public, whose parameters are less well defined. Another concern is the credibility of organizations, particularly NGOs, which lack regulatory oversight of their investigations.

### Human Rights Content

Determining what constitutes “human rights content” will be difficult given the lack of any single authoritative definition. However, there have been efforts to define and regulate similarly slippery

categories of information by international bodies and social media companies, processes that may provide guidance on how to approach this problem—for example, terrorist content or hate speech. One of the biggest challenges in curating or regulating any content is defining the parameters of that content. For example, if you outlaw “sexually explicit” content, there must be a clear definition of what qualifies as “sexually explicit.” If you require the preservation of “human rights content,” there must be a clear definition of what falls within that category and what does not.

*Types of violations and crimes:* Stakeholders must determine what types of violations and crime qualify. For example, will “human rights” content only include material related to violations of codified human rights laws or will it include information related to violations of international criminal, humanitarian, and human rights law and their national equivalents?

*Degree of relevance and probative value:* The group must determine how relevant the content must be to proving a violation. In addition to the types of violations, the level of relevance and probative value will need to be determined, since information can be used in a number of different ways.<sup>163</sup> Therefore, the group will need to determine whether the content in question only applies to direct evidence of violations, or whether it also encompasses indirect or circumstantial evidence, lead information, intelligence and contextual information.

---

163 As Professor Jay D. Aronson of Carnegie Mellon University points out, online content can be beneficial to human rights investigations in a number of different ways: “Video[s] have potential value at every stage of a human rights investigation, whether that investigation is designed to feed into advocacy or legal proceedings. Most commonly, video[s] [generate] leads that can be used to start an investigation. [They] can also provide evidence to establish that a crime or violation occurred, or [they] can be used to support a particular factual finding, such as whether a particular weapon was used in a conflict or whether pollution from a particular mining site is polluting a water source. Sometimes, [videos] can also link a particular person, group, government, or company to the violation in question.”

Defining human rights content will be a challenge, but there are some analogous situations that could serve as guidance. For example, there are no universally accepted definitions of “hate speech.” As UN Special Rapporteur David Kaye notes, hate speech “is a shorthand phrase that conventional international law does not define” and one which has “double ambiguity.”<sup>164</sup> Because of this lack of a universal definition, states and companies have faced tremendous difficulty in addressing and mitigating its proliferation online, and ultimately capturing the totality of digital forms that hate speech manifests.<sup>165</sup>

Despite this double ambiguity, however, international bodies, social media companies, and other organizations have had to come up with processes for defining such speech. Processes range from establishing a set of variables for assessing content; developing a list of terms that act as proxies or triggers for identification; adopting definitions established by authoritative bodies like the United Nations; and constructing individualized, context-based assessments that come from the communities that are impacted.<sup>166</sup> In the present case, human rights content could include content that shows direct crimes or violations of human rights; content that is indirect evidence of those crimes or violations; content that contains lead information for investigating crimes or violations; intelligence that helps decision-makers regarding policy about the crimes or violations; or contextual information that provides investigators and researchers a broader understanding of the crimes or violations. Ultimately, whatever definition is adopted will have to be operational, meaning capable of detection by human and/or algorithmic monitors.

---

164 David Kaye, “Promotion and Protection of Human Rights: Human Rights Questions, Including Alternative Approaches for Improving the Effective Enjoyment of Human Rights and Fundamental Freedoms,” United Nations General Assembly, October, 2019, [https://www.ohchr.org/Documents/Issues/Opinion/A\\_74\\_486.pdf](https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf).

165 Ibid.

166 See, e.g., Hate Speech Methodology Report (Human Rights Center, forthcoming 2021). We have conducted extensive research into different processes that have been used to create operational definitions for difficult-to-define terms.

## Preservation and Access

Finally, given the above challenges, stakeholders will need to be able to articulate what is meant by “preservation” and “access.” Will both metadata and content be retained? How and where will the information be held, by whom, for how long, and in what format? Is the information anonymized prior to its preservation and/or access by external parties? What is meant by access? Can external parties view, embed and/or download the data into a separate repository? When “accessing” the archive, what information is available (for example, just the metadata or the content itself or something else)?

## LEGAL COMPLIANCE

Under current legal frameworks, social media companies have no obligation to participate in the development of a digital evidence locker or other type of human rights archive, or to participate once the mechanism is created. Their current participation is voluntary. Absent the modification of existing law or creation of new law, the mechanism must be designed in a way that provides social media companies with legal cover (ensuring that they can comply with applicable laws) and incentivizes them to cooperate.<sup>167</sup> At present, social media companies must be compelled to preserve and share user data. Key legal challenges include compliance with data protection and privacy laws, which are inconsistent and in many cases still being developed.

---

<sup>167</sup> For a proposal to address U.S.-based legal barriers to preservation and access of social media evidence of alleged human rights abuses, see Olivia Mooney, Kate Pundyk, Nathaniel Raymond and David Simon, “Social Media Evidence of Alleged Gross Human Rights Abuses: Improving Preservation and Access Through Policy Reform,” *Mass Atrocities in the Digital Era Initiative Working Paper*, 2021.

## Data Protection and Privacy Laws

One of the most complicated challenges a mechanism will face is compliance with privacy and data protection laws in the United States and Europe,<sup>168</sup> as well as under international law—both because of how complicated the law is but also because privacy is at least as great an interest for human rights as accountability, and thus whatever approach is created has to balance both interests as well as other human rights interests, such as freedom of expression and access to information. Stakeholders must examine the applicability and impact of any relevant national and international laws. In particular, stakeholders should thoroughly assess the laws and regulations that control the processing, transfer, retention, and destruction of personal data. This section spotlights some of the most significant of these laws.

*Electronic Communications Privacy Act (ECPA):* Enacted in 1986, the ECPA covers wireless communications, including email and other digitally transmitted conversations, and prohibits the disclosure of customer records to third parties.<sup>169</sup> If the third party is a government, service providers can share only customer records in response to a valid judicial order or in the case of an emergency threatening immediate danger of death or serious physical injury.<sup>170</sup>

*Stored Communications Act (SCA):* Section 2701 et seq. of the ECPA addresses the privacy and disclosure of emails and other electronic communications.

---

<sup>168</sup> Most global social media service providers are U.S.-based. Therefore, several U.S. laws apply.

<sup>169</sup> “Electronic Communications Privacy Act (ECPA),” See Electronic Privacy Information Center, accessed January, 2021, <https://epic.org/privacy/ecpa/>.

<sup>170</sup> *Ibid.* The ECPA allows law enforcement agencies to “forgo even the minimal burden of a subpoena or a court order and claim there is an emergency that necessitates the records to be turned over.”

According to the SCA,<sup>171</sup> providers must disclose account records and other information to the U.S. government upon judicial order. The SCA limits the ability of certain technology providers to disclose user information. It also limits third parties' ability to access electronic communications without sufficient authorization. Because of the SCA's restrictions on disclosure, technology providers and litigants often invoke the SCA when seeking to quash civil subpoenas for electronic communications. The SCA also provides a detailed framework governing law enforcement requests for electronic communications.

*California Privacy Protection Act (CPPA):* Many of the major social media companies are headquartered in California and must also comply with the CPPA. The California Public Records Act (CPRA), which will go into effect in 2023, may also be relevant. For example, the CPRA restricts third party contractors that have received personal data from a business from combining that data with data it receives from other businesses, although there may be exceptions for some business purposes.

*U.S. Constitution:* The Fourth Amendment protects the right to privacy by limiting government intrusion on individuals' privacy interests. Relevant privacy interests may include those of the content uploader, those depicted in or otherwise implicated by the content, etc. Under the current legal framework, U.S. law enforcement agencies can compel social media service providers to preserve user content and metadata past the standard retention period (which varies with each company) with a preservation request for 90 days with possibility to extend.<sup>172</sup> Once preserved, these agencies can sub-

mit a judicially authorized warrant or subpoena to the company in order to compel the information. Other Constitutional provisions that may be implicated by creation of an evidence locker are the First Amendment (freedom of speech) and Fifth Amendment (due process).<sup>173</sup>

*Other U.S. Based Laws, Regulations and Policies:* The Federal Trade Commission plays a role in establishing privacy protections. Frameworks for information sharing across national lines—such as the Mutual Legal Assistance Treaty and the CLOUD Act—also contain provisions that protect citizens from overbroad or arbitrary government requests, without exception for atrocity crimes. Finally, social media companies have data disclosure and privacy policies that will need to be reconciled with any next steps. The companies' privacy policies, terms of use, and terms of service help govern relations between users and platforms, and are subject to applicable laws and regulations reflecting privacy protection principles.<sup>174</sup>

*European Union General Data Protection Regulation:* In Europe, the GDPR regulates the processing and transfer of personal data, with a focus on particularly sensitive data such as personally identifying information (PII).<sup>175</sup> The GDPR applies to EU data subjects (people), not just to data or companies lo-

---

171 "Electronic Communications Privacy Act (ECPA)," Electronic Privacy Information Center (Epic), accessed October, 2020, <https://epic.org/privacy/ecpa/>.

172 Orin Kerr, "The Fourth Amendment and Email Preservation Letters," *The Washington Post*, October, 2016, <https://www>

---

[.washingtonpost.com/news/volokh-conspiracy/wp/2016/10/28/the-fourth-amendment-and-email-preservation-letters/](https://www.washingtonpost.com/news/volokh-conspiracy/wp/2016/10/28/the-fourth-amendment-and-email-preservation-letters/).

173 Related international human rights provisions that may be of particular interest to human rights advocates include the right to Freedom of Expression and the right to Access to Information. See, e.g., David Kaye, *Speech Police: The Global Struggle to Govern the Internet*, (New York: Columbia Global Reports, 2019).

174 Ibid.

175 Article 4(1) of EU Regulation 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 (hereinafter "EU GDPR").



cated in the EU—meaning that the GDPR applies to data generated by EU entities, even if those entities are outside of the EU. PII is specific to an individual and under the EU GDPR, “natural persons should have control over their own personal data.”<sup>176</sup> A transfer of control, or shared control, may be appropriate and/or permitted when the data subject uses services that require the collection of their personal information. Data altruism processes, whereby individuals can “opt in” to allow their PII to be used for specific purposes, may allow for the transfer of personal data for human rights purposes in some contexts. Of course, it’s unlikely that human rights violators or war crimes perpetrators (for example) would ever consent to such uses.

*Article 8 of the European Convention on Human Rights:* Article 8 confirms that individuals have a right to respect for their “private and family life... home and correspondence,” with exceptions limited to what is “necessary in a democratic society,” in pursuit of a legitimate aim, and in accordance with the law.

*Organisation for Economic Co-operation and Development (OECD) Privacy Guidelines:* In 2013, the OECD released privacy guidelines that provide a valuable framework for compliance with privacy regulations.

## Power to Compel Disclosure

Many tech companies have taken the position that they can only share data with law enforcement or other external entities if such sharing is lawful and they are compelled to do so by law. Tech companies routinely reiterate their commitment to user privacy by stating that they carefully review all disclosure

requests and only share information in accordance with applicable law.<sup>177</sup>

*Mutual Legal Assistance Treaties (MLAT):* Governments can make preservation requests through a few different mechanisms, and can receive information through Mutual Legal Assistance Treaties (MLATs). MLATs provide a legal means by which foreign law enforcement agencies can request evidence in the United States.<sup>178</sup> The MLAT process only applies to law enforcement and, in practice, can take an extremely long time to process.

*Clarifying Lawful Overseas Use of Data (CLOUD) Act:* The CLOUD Act was signed into law in 2018 to serve as an expedient alternate to MLAT for regulating access to data across borders, with a goal of “improv[ing] privacy and civil liberties protections.”<sup>179</sup> The CLOUD Act allows governments to “request data of non-U.S. persons from U.S.-based providers,” and allows the U.S. government to track and review how partner governments use the requested

---

<sup>177</sup> See, “How Google Handles Government Requests for User Information,” Google Privacy and Terms, accessed October, 2020, <https://policies.google.com/terms/information-requests> for Google’s statement on government requests. See, “Guidelines for Law Enforcement on Preservation Requests,” Twitter Rules and Policies, accessed October, 2020, <https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support> for Twitter’s policies on law enforcement requests. See, “Information for Law Enforcement Authorities,” Facebook Safety Center, accessed October, 2020, <https://www.facebook.com/safetyv2>. for more information on Facebook’s policies regarding law enforcement preservation requests.

<sup>178</sup> See, e.g., Alexa Koenig, Keith Hiatt, and Khaled Alrabe, “Access Denied? The International Criminal Court, Transnational Discovery, and The American Servicemembers Protection Act,” *Berkeley Journal of International Law*, May, 2018, <https://lawcat.berkeley.edu/record/1128475?ln=en>.

<sup>179</sup> Jennifer Daskal and Peter Swire, “Why the CLOUD Act is Good for Privacy and Human Rights,” *Lawfare*(blog), March, 2018, <https://www.lawfareblog.com/why-cloud-act-good-privacy-and-human-rights>.

---

<sup>176</sup> Article 4(1), EU GDPR, April 2016.

data.<sup>180</sup> A bi-lateral, contractual approach to information sharing, the CLOUD Act includes substantive and procedural privacy protections. The Act bars foreign governments from targeting data of U.S. citizens and residents—governments interested in such data must obtain a warrant based on probable cause.<sup>181</sup> Furthermore, requests must be based on credible facts and pertain to a specific person, account, address, personal device, and/or identifier.<sup>182</sup> All requests must be reviewed by judicial authorities, not be kept beyond a reasonably necessary period, not infringe on freedom of speech, and finally, the foreign governments must agree to compliance reviews to safeguard against abuse.<sup>183</sup> Ultimately, the CLOUD Act “lifts [previous] legislative bar[s] on disclosure,” and reinforces privacy protections by allowing the U.S. to enter into strict executive agreements with partner governments.<sup>184</sup>

### Duty to Preserve Evidence

Some stakeholders have asked whether the duty to preserve evidence could apply to human rights content on social media sites. The duty to preserve evidence is a legal concept that arises when litigation is reasonably anticipated—for example, when a lawsuit is served or a governmental investigation initiated. Federal Rule of Civil Procedure (FRCP) 37(e) states that a party must preserve documents and electronically stored information (ESI) when it reasonably anticipates litigation. Professor Aronson has argued that the increased need for digital documentation of human rights violations necessitates an acknowledgment of the “duty to preserve” online content in order to expand the historical record of

global human rights abuses.<sup>185</sup> Similarly, Professor Noha Aboueldahab of Georgetown University states that “documentation keeps the issue of justice in Syria alive.”<sup>186</sup> While this is an interesting idea, it is hard to see how this could serve as the basis for an evidence locker, since a primary objective of such a locker would be to preserve human rights content before a formal investigation is initiated. Thus, the group may want to consider whether there is a legally feasible way to expand existing interpretations of the duty to preserve.

### AUTOMATED DETECTION OF GRAPHIC CONTENT

In addition to the legal challenges highlighted above, there are several operational considerations that impact the feasibility of a mechanism for social media preservation, including how to identify relevant information before it is deleted. As discussed above, content moderation plays a significant role in what digital information is available to the public. Historically, social media platforms have employed human monitoring to review content that may violate their terms of service (for example, hate speech or violent extremist speech). Today, tech companies often use algorithms to monitor and flag online content. Automation helps the companies address challenges of scale and speed, minimize some of the impact on human reviewers, reduce costs, and comply with increasingly fragmented and stringent legislation. However, algorithms are far from perfect—and are often either too narrow or overbroad

---

180 Ibid.

181 Ibid.

182 Ibid.

183 Ibid.

184 Ibid.

---

185 Jay D. Aronson, “Preserving Human Rights Media for Justice, Accountability, and Historical Clarification,” *Genocide Studies and Prevention: An International Journal* 4, no. 1, May 2017.

186 Noha Aboueldahab, “Writing Atrocities: Syrian Civil Society and Transitional Justice,” *Brookings Doha Center Analysis Paper*, no. 21, (2018), [https://www.brookings.edu/wp-content/uploads/2018/04/transitional-justice-english\\_web.pdf](https://www.brookings.edu/wp-content/uploads/2018/04/transitional-justice-english_web.pdf).

in terms of the content they catch. As Dia Kayyali, associate director for advocacy at Mnemonic/Syrian Archive and formerly of WITNESS, warns, algorithmic filters cannot distinguish between radical propaganda and a video documenting atrocities.<sup>187</sup> This is because algorithms are still bad at understanding context and intent.<sup>188</sup>

Integration of Contextual AI, a human-centric approach to AI, is limited.<sup>189</sup> A publication entitled *Caught in the Net: The Impact of 'Extremist' Speech Regulations on Human Rights Content*, jointly produced by EFF, Syrian Archive, and WITNESS, highlights that, “Google’s transparency report shows that YouTube removed 33 million videos in 2018, amounting to roughly 90,000 per day.<sup>190</sup> Of those flagged for potential violation of terms of service, 73 percent were removed through automated processes before the videos were even available for viewing [by humans].”<sup>191</sup> Between July and September 2020 alone, Facebook’s detection and reporting tools purportedly helped identify 98.2 percent of violating content before users reported it. In the January–May period, Facebook’s automated detection tools

identified 99.2 percent of violating content.<sup>192</sup> The company’s detection and reporting tools help the company identify potentially violating content, which is then allegedly reviewed.<sup>193</sup>

Social media platforms’ use of algorithms for mass detection and deletion exacerbate the challenges of preserving data for use as evidence. Although the transparency reports offer valuable quantitative information, the algorithms operate as a black box to outsiders, who cannot review the datasets used for training those algorithms, or qualitatively assess any quality control mechanisms that are applied. Legal investigators, lawyers, and human rights researchers may never see large volumes of content if it is automatically removed. While technology providers are required to report some imagery such as content involving child exploitation to a federally designated clearinghouse and retain related information for 90 days, this timeframe is insufficient for international criminal and human rights cases, which may be brought years after an atrocity occurs.

---

187 “YouTube Community Guidelines enforcement,” Google Transparency Report, accessed October, 2020, <https://transparencyreport.google.com/youtube-policy/removals>.

188 Of course, humans are also often imperfect at detecting context and intent. One issue is how to maximize the relative responsibilities and functioning of digital and human review to maximize the accuracy of interpretation.

189 Oliver Brdiczka, “Contextual AI: The next frontier of artificial intelligence,” Digiday, accessed January, 2021, <https://digiday.com/sponsored/adobesbl-contextual-ai-the-next-frontier-of-artificial-intelligence/>.

190 Abdul Rahman Al Jaloud and Hadi Al Khatib, “Caught in the Net: The Impact of ‘Extremist’ Speech Regulations on Human Rights Content,” Electronic Frontier Foundation, May, 2019, <https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#fnref10>.

191 “YouTube Community Guidelines Enforcement,” Google, accessed January, 2021, <https://transparencyreport.google.com/youtube-policy/removals>.

---

192 “Community Standards Enforcement Report,” Facebook, accessed January, 2021, <https://transparency.facebook.com/community-standards-enforcemen>.

193 Ibid.

# DISCUSSION

Given the important role that social media content may play for human rights investigators and legal authorities, some form of digital evidence locker—whether designed as a freestanding repository, integrated into an existing mechanism, or housed within the companies themselves—is likely critical for strengthening human rights accountability. However, the potential and actual value of social media content for accountability needs further exploration and articulation—as it does for other end uses, such as social science research, advocacy, and creation of an historical record.

Several efforts are currently being contemplated to create a mechanism for the preservation of social media content that has potential evidentiary value for human rights investigations, along with other end uses. These range from suggestions that social media companies work with existing frameworks and processes but expand who can request preservation of social media content on that company’s servers in anticipation of later legal requests; to creation of a permanent mechanism that could ingest social media content along with other kinds of information with evidentiary or academic value;<sup>194</sup> working with

---

194 See e.g., Federica D’Alessandra, Stephen Rapp and Sareta Ashraph, “Anchoring Accountability for Mass Atrocities: Providing the Support Necessary to Fulfil International Investigative Mandates,” *OpinioJuris*, September, 2020.; see also John Bowers, Elaine Sedenberg, and Jonathan Zittrain, “Platform

governments, and especially the U.S. government, to create a system for brokering requests for social media content from researchers, international investigators, researchers, and others (shifting decision-making around preservation and provision to an independent body);<sup>195</sup> and creating a series of independent, conflict-specific archives that are ideally located close to and accessible from both recent and historic sites of atrocity,<sup>196</sup> as well as developed and controlled by those most impacted.<sup>197</sup>

It’s unlikely that any one model can meet the very different end purposes that have been outlined by the diverse range of human rights organizations participating in current debates. Thus, it’s important as a next step for human rights organizations to identify the following:

**First, what are the various potential end uses for human rights content posted to social media?** At least three are regularly being discussed by academics

---

Accountability Through Digital ‘Poison Cabinets,” Knight First Amendment Institute at Columbia University, April, 2021.

195 Olivia Mooney, Kate Pundyk, Nathaniel Raymond and David Simon, “Social Media Evidence of Alleged Gross Human Rights Abuses: Improving Preservation and Access Through Policy Reform.”

196 See e.g., Wille, “‘Video Unavailable’: Social Media Platforms Remove Evidence of War Crimes.”

197 See, e.g., Zakiya Collier, “Call to Action: Archiving State-Sanctions Violence Against Black People,” *Medium*, June, 2020.

and practitioners: evidence, social science research, and community based interests, including advocacy, broader transitional justice efforts, and development of an historical record. Of the three, evidence is likely the smallest use case. Of course, content may overlap and be relevant for all three purposes,



or limited to a single use case, as depicted in the diagrams below.

Even though stakeholders are all using the term “evidence locker,” they are using that term in different ways. Each use case—evidence for courts, social science research, and broader advocacy or transitional justice purposes—should and probably will need to be treated very differently, given varying operational needs and legal contexts. For example, while anonymized data may suffice for social science research, it will often be inadequate for legal evidentiary purposes, where what is most useful is often who was at the site of an atrocity, how higher-ups may be implicated, and how those individuals are networked.<sup>198</sup> Similarly, with regards to social media content used as potential evidence in court,

---

198 Alexa Koenig, Eric Stover, Peggy O’Donnell, Camille Crittenden, “Beyond Reasonable Doubt: Using Scientific Evidence to Advance Prosecutions at the International Criminal Court,” Human Rights Center, October, 2012, [https://www.law.berkeley.edu/wp-content/uploads/2018/02/Bellagio\\_report\\_2018\\_7.pdf](https://www.law.berkeley.edu/wp-content/uploads/2018/02/Bellagio_report_2018_7.pdf); Alexa Koenig, “The New Forensics: Using Open Source Information to Investigate Grave Crimes,” Human Rights Cen-

most valuable will be evidentiary “packages” that include hashed content and any metadata, which come straight from the social media companies and/or that otherwise retain clear chains of custody—a level of detail and effort that’s not as necessary for research or advocacy. Each of those potential scenarios is subject to different legal, ethical and operational constraints. Stakeholders should disentangle and could “reverse engineer” from anticipated use cases to design effective processes for the preservation and potential sharing of information.

**Second, what do stakeholders want to have happen with social media content? For example, do content creators want the content they generate used for accountability? For research? For historical memory?** Answers to these questions will likely vary, of course, from person to person and organization to organization. In addition, sometimes what content creators and other social media users want is irrelevant (for example, if social media content is evidence of a crime and is sought by law enforcement, no one is going to seek the preferences of perpetrators, such as those who post videos of egregious crimes and human rights abuses online when determining how and when it’s used). However, as potential use cases are identified, it would be helpful to identify the reasons why people are posting photos, videos and text online and what they want to result from that activity, to ensure that whatever is created has survivors’ and victims’ interests at its center.

**Third, how should “human rights content” be defined and who should make that determination?** To make this question actionable, civil society organizations should consider which content is most critical to preserve. Such determinations are necessary for clearly communicating with social media

---

ter, 2017, [https://www.law.berkeley.edu/wp-content/uploads/2018/02/Bellagio\\_report\\_2018\\_7.pdf](https://www.law.berkeley.edu/wp-content/uploads/2018/02/Bellagio_report_2018_7.pdf).

companies with regards to what kinds of content is at issue.

**Fourth, which scenarios should be prioritized and used when piloting one or more options?** It may make sense to “stress test” a pilot locker or series of lockers with a subset of content that falls within the “human rights evidence” bucket and learn from that pilot testing to minimize any unanticipated risks endemic to the process design. For example, should a test run prioritize a particular conflict? Prioritization could also be based on regions or other geographical markers, particular types of atrocities (such as chemical weapons attacks or attacks on hospitals, both of which lend themselves to automated detection), etc.

**Fifth, what financial, political and other resources are needed to create those models, and where will those resources come from?** Will the evidence locker depend on voluntary or mandatory contributions of social media content? Will social media companies provide financial or other support? Will governments? Will foundations? How sustainable is that funding? Can an endowment be established along with safeguards on who determines how much is allocated to evidence locker staff and support? How does the potential funding source alter who will engage with the locker and under what conditions, and the relative independence of the locker? Who should have custody over any evidence locker that is developed and why?

**Sixth, should discussions around preservation and access be disaggregated or synced? If preservation is prioritized, what models might best facilitate later requests for access?** Preservation is likely a simpler challenge to solve than access, the latter of which is subject to a much greater array of legal limitations with regards to who can have access to social media content and under what conditions—in part due to the increased risks to society and individuals that can accrue from the sharing of sensitive data. Will laws need to be modified or established to enable access? Ultimately, it may make sense to first focus on preservation to safeguard content at risk of deletion, but keep potential access models in mind when doing so.

Ultimately, issues related to discovering, preserving, and assessing the privacy of those potentially affected by the content should be central to determining the model or models best suited for preserving human rights content and identifying which stakeholder(s) should be responsible for assessing and responding to any vulnerabilities.

Finally, interested parties include companies, human rights organizations, government representatives, and those who post information to online places with the hope that those posts will generate justice. Ongoing conversations between stakeholders should first and foremost be grounded to achieve the human rights interests of all parties involved, provide a process through which human rights groups can request the preservation of content, and finally, be based on further research regarding the ethical and privacy implications and challenges in creating a digital evidence locker unit for human rights violations, and especially for atrocity crimes.

# RECOMMENDATIONS

In order to craft a way forward that responds to practical, ethical and legal challenges—and best protects and respects the human rights interests of disparate stakeholders—we make the following recommendations:

## **RECOMMENDATION ONE:**

**Funders should support a series of stakeholder-organized workshops** designed to ensure that any mechanism or framework is carefully thought out, founded in empirical research, inclusive in its process, grounded in the interests and concerns of those closest to underlying atrocities, protective and respectful of human rights, and carefully designed to meet clearly articulated aims. In the legal context, these conversations should prioritize preservation over access given the ephemerality of online evidence, and given existing pathways for accessing evidence for legal processes.

## **RECOMMENDATION TWO:**

**Stakeholders should hold separate but connected conversations about the need to preserve human rights content for three different end uses: 1) evidence for courts, 2) social science research, and 3) historical record-development (including human rights documentation and the development of community archives).** It's unlikely that any single evidence locker realistically can (or should) be created to meet all three needs, given the very different legal and operational constraints specific to each. If a

single mechanism or legal framework *is* established, access to archival content will likely need to differ depending on the identity of the requesting party, the desired end use, and other variables identified in this report.

## **RECOMMENDATION THREE:**

**Interested stakeholders should start by identifying a concrete and relatively clear use case to explore next steps.** The use case could focus on a particular type of violation, such as the illegal use of chemical weapons or targeting of hospitals; or geography, for example, by focusing on the genocide in Myanmar or alleged war crimes in Syria. Eventually, stakeholders should also develop a working definition of “human rights content.” Several resources mentioned in this report may help, including processes used to define similarly broad terms, such as hate speech and terrorism. Stakeholders will need to make that definition operational, reflecting the practical reality that identification of relevant content may need to be automated or a combination of manual and technical processes.

## **RECOMMENDATION FOUR:**

**Stakeholders should identify and/or define what is meant by “preservation” and “access,” and develop standards for how preservation and access might be operationalized legally, ethically, and technically across a variety of use cases.** This may include articulating what should be preserved and by whom,

how long relevant information should be held, who should be permitted access to what information, and under what conditions.

**RECOMMENDATION FIVE:**

**Stakeholders should identify how international and national privacy laws, including those mentioned in this report, limit and provide opportunities for the creation of one or more potential evidence lockers.**

The human rights community needs to better understand the various laws to which social media companies are subject and how those laws vary based on geography and use case--and how such a locker can guard against governmental and nongovernmental abuse of data. Such legal research can help shape an understanding of what is legally, fiscally, and operationally possible.

**RECOMMENDATION SIX:**

**Interested stakeholders should identify the economic model(s) that could sustainably support an evidence locker.** This may include creating a permanent or term endowment<sup>199</sup> (if a freestanding repository), budgeting by the social media companies (if they preserve the content on their servers), or ongoing fundraising (via philanthropy or crowdsourcing). Stakeholders should also consider whether funding should be accepted from corporations or governments, and if so, under what conditions.

**RECOMMENDATION SEVEN:**

**Social media companies should provide a mechanism by which human rights organizations (to be defined) can request the preservation of content.** Given the temporal lag of international human rights cases compared to domestic criminal cases, such discussions should be grounded in the length of time legally and operationally permissible for social media companies to hold deleted content,

ideally starting with a time period of three to five years with the possibility of renewal. Stakeholders should determine how to define eligible human rights organizations, considering any legal, ethical, or operational constraints. Definitions and processes should be designed with transparency and nondiscrimination in mind, while safeguarding against security and privacy risks.

**RECOMMENDATION EIGHT:**

In planning for and crafting a future evidence locker(s), interested stakeholders should evaluate whether the discovery of content will be undertaken by staff affiliated with the evidence locker, external stakeholders, or both.

**RECOMMENDATION NINE:**

**Social media companies should make transparent how they employ algorithms for mass monitoring, flagging, and removing user-generated content,** including information about training datasets and their algorithms' error rates, to inform the design of realistic processes for the preservation of human rights content.

**RECOMMENDATION TEN:**

**Civil society organizations should be able to articulate whether they are proposing a "supply driven" or "demand driven" model.** They should be able to explain why the particular content is needed and for what end. Helpful information, if available, would be whether there are cases that are failing because of a lack of preservation and access to social media content. What research is unable to happen because of the lack of preservation and access? What histories are potentially being lost?

---

<sup>199</sup> This refers to an endowment created for a determined length of time.



## CONCLUSION

As social media companies increasingly automate the detection and removal of content, including content with human rights evidentiary value, civil society is debating the creation of a centralized mechanism or series of mechanisms for preserving such content for a range of purposes. However, before such a system can be designed, a number of threshold issues must be addressed, including how to define “human rights content,” identifying the various relevant use cases, looking at existing repositories to see what options may be available for designing a system (or systems) for responding to those use cases, researching the legal issues implicated by the various models, and identifying the stakeholders who will be most impacted and thus should be integrated into the design and consultation process.

While the challenges to creating a human rights evidence locker or series of lockers are significant, existing repositories suggest those challenges can be surmounted—either through a system that involves social media companies’ voluntary participation or through the passage of legislation that imbues next steps with legal weight.

Given the tremendous power of digital content to strengthen the evidentiary foundations of cases, and the significant need for increased accountability for grave violations of human rights, humanitarian and international criminal law, as well as for social science research and for history, these are efforts worth undertaking. Ultimately, they hold tremendous possibility for strengthening the role of those on the frontlines in the quest for justice.


# ACKNOWLEDGMENTS

This report was authored by Alexa Koenig, Shakiba Mashayekhi, Diana Chavez-Varela, Lindsay Freeman, Kayla Brown, Zuzanna Buszman, Rachael Cornejo, Amalya Dubrovsky, Therese Franklin, Sofia Jordan, Sang-Min Kim, Lucy Meyer, Pearlé Nwaezeigwe, Sri Ramesh, Maitreyi Sistla, Eric Sype, and Ji Su Yoo. The authors also thank the following individuals, who provided critical feedback on earlier

drafts: Jacob Berntsson, Jeff Deutch, Jacqueline Geis, Sam Gregory, Adam Holland, Dia Kayyali, Aaron Kearney, Andrea Lampros, Raquel Vazquez Llorente, Kelly Matheson, Libby McAvoy, Olivia Mooney, Yvonne Ng, Brandie Nonnecke, Nathaniel Raymond, Jörg Roofthoof, David Simon, Miranda Sissons, Eric Stover, Belkis Wille, and Aaron Zelin.



HUMAN RIGHTS CENTER  
UC BERKELEY SCHOOL OF LAW  
2224 PIEDMONT AVENUE  
BERKELEY, CA 94720  
510.642.0965

[HRC@BERKELEY.EDU](mailto:HRC@BERKELEY.EDU)  
[HUMANRIGHTS.BERKELEY.EDU](mailto:HUMANRIGHTS.BERKELEY.EDU)  
[MEDIUM.COM/HUMANRIGHTSCENTER](https://medium.com/humanrightscenter)  
 [@HRCBERKELEY](https://twitter.com/HRCBERKELEY)