SUBMISSION TO CALL FOR INPUT

"The relationship between human rights and technical standard-setting processes for new and emerging digital technologies (2023)" - Report of the High Commissioner for Human Rights

SUBMITTED ON: 3 March 2023

BY WITNESS

I.   Introduction

This submission shares WITNESS's views on the relationship between human rights and technical standard-setting processes for new and emerging digital technologies. This document is shaped by our three decades of experience helping communities advocate for human rights change to create trustworthy information, to protect themselves against the misuse of their content, and to challenge misinformation that targets at-risk groups and individuals.

Of relevance to this submission is WITNESS's experience as members of the Coalition for Content Provenance and Authenticity (C2PA or the 'Coalition'), a technical standard-setting body that is building specifications for capturing verifiable provenance of media content. As part of this Coalition, WITNESS co-chairs the 'Threats and Harms Taskforce', and has led the development of a Harms, Misuse and Abuse Assessment that identifies threats to human rights, as well as existing and potential mitigation strategies and actions.

This submission covers our work related to the "C2PA Harms, Misuse and Abuse Assessment" and provides insights that can be of value for future discussions within standard-setting bodies committed to protecting and promoting human rights globally.

II. Overview of the 'Harms, Misuse and Abuse Assessment' of the C2PA technical specifications

A. Introduction to provenance and authenticity infrastructure and the C2PA

For three decades, WITNESS has been helping communities advocating for human rights change to create trustworthy information, to protect themselves against the misuse of their content, and to challenge misinformation that targets at-risk groups and individuals. One of the strategies we have identified to catalyze this is to build more robust ways to track whether images, video, and audio have been manipulated, mis-contextualized or edited, and if so, when and by whom. These digital and physical structures and tools that allow for the source and history of media to be tracked collectively make up what is now known as provenance and authenticity infrastructure. WITNESS has been part of this growing area of exploration to advocate for systems that empower critical voices and reflect human rights and privacy concerns.

The C2PA is a project that brings together the efforts of the Content Authenticity Initiative (CAI) and Project Origin. Founded in late 2019 by Adobe in collaboration with the New York Times and Twitter, the CAI is building a system to provide provenance and history for digital media. Project Origin, founded in 2019 by BBC, CBC Radio Canada, Microsoft and the New York Times, focuses on tackling disinformation in digital news by defining an end-to-end process for publishing, distribution and attaching signals to a content to demonstrate its integrity. The C2PA binds the efforts of these two groups and focuses exclusively on the development of open, global technical standards to channel the content provenance efforts of the CAI and Project Origin.

The C2PA's goal to create technical specifications is the latest, most significant move towards more widespread efforts on provenance and authenticity infrastructure driven by governments, platforms and public demand. As of January of 2022, version 1.0 of the Technical Specifications are publicly open and they may be included into any tool or device that creates or processes digital media. Within a C2PA-enabled ecosystem, the provenance information of an image, video or audio (and later potentially documents) can be tracked from the moment that is captured, all the way until the content is published and consumed.

> **Example use-case**
>
> *A human rights defender captures footage of a war crime using a C2PA-enabled camera. The provenance information would offer verifiable signals to suggest that this is a raw, unedited video. Then, with a C2PA-enabled editing software, sensitive information such as the faces of individuals that appear in the video may be blurred or redacted, leaving a trace of what was done to the media file and what was not edited out. Finally, a C2PA-enabled publishing tool would allow viewers to trace the source and history of this asset.*

B. Harms Modeling in the C2PA

Provenance and authenticity technologies can be a powerful tool for journalists, activists, human rights defenders and others helping to bring to the truth out. However, these technologies can also lead to potential harms to a broad range of individuals and communities, especially those that are already most at risk. To avert and mitigate these potential harms, WITNESS advocated for including diverse human rights and journalism use cases from the beginning of the C2PA design process, and highlighted to the Coalition global human rights concerns in relation to the Guiding Principles for C2PA design and specifications. WITNESS emphasized that:

- C2PA specifications MUST respect the common privacy concerns of each of the target users named earlier;
- C2PA specifications MUST be reviewed with a critical eye toward potential abuse and misuse; and
- C2PA specifications MUST be reviewed for the ability to be abused and cause unintended harms, threats to human rights, or disproportionate risks to vulnerable groups across different countries, not only the Global North.

Based on these principles, WITNESS led a harms modeling exercise which focused on analyzing how a socio-technical system might negatively impact users, stakeholders, broader society, or otherwise create or re-enforce structures of injustice, threats to human rights, or disproportionate risks to vulnerable groups globally. Harms modeling systematically requires combining knowledge about a system architecture and its user affordances, with historical and contextual evidence about the impact of similar existing systems on different social groups. This combined information frames the ability to anticipate harm.

Harms modeling considers the ramifications of a technological system both from the perspective of the technology developers as well as users and non-user stakeholders. In other words, harms modeling considers what kinds of harms may result from the configuration of a system as well as what kinds of harms may result from both its intended and unintended uses. It is necessary to combine all of these considerations to achieve a broader perspective on potential harms, which is why our process captured external feedback from people with a broad range of lived, practical and technical experiences, all coming from different parts of the world, and acting across areas that included civic media, human rights, misinformation and disinformation, activism, technology advocacy and accountability, and digital rights.

III.   **Learnings from WITNESS's participation in the C2PA that are relevant to the relationship between human rights and technical standard-setting processes for new and emerging digital technologies**

A. **General comments and considerations**

Standards Development Organizations (SDOs) are an important stakeholder in the early-stage development of new and emergent technologies, and have the ability to shape them and the larger technical ecosystem–including tools, services, business models and industries. Regrettably, by numerous accounts, including our own experience, civil society and human rights organizations barely take an active role in these standards-setting bodies. Relatedly, consultation with civil society tends to lack diversity and global representation.

Considering their role in shaping technological ecosystems, SDOs should be responsible for incorporating human rights considerations in technical standards–especially when these technologies may directly or indirectly lead to harms or threats to human rights, or bring disproportionate risks to vulnerable groups globally. As highlighted by WITNESS in previous submissions to the OHCHR, this is also in line with the *Guiding Principles on Business and Human Rights* regarding Standards Development Organizations (SDOs), and in specifications-development working groups (Council Resolution [47/23](#), 2021).

One of the obstacles that may inhibit participation from civil society and human rights organizations is the notion that standards-setting processes are strictly technical–as opposed to technical processes imbued by interests and worldviews. This oftentimes results in spaces that value technically trained specialists and undermine experience from other stakeholders, including the same people that may be most affected by the standards being designed. The fact

that SDOs are not held to the same degree of oversight as companies offering products and services further erodes any motivation to include a broader range of stakeholders.

Another obstacle is that the expertise and experience required for designing and developing standards are often not the same as the ones for carrying out a harms modeling process. This means that to promote effective participation from civil society and human rights organizations it is necessary to develop targeted and thoughtful mechanisms that can bridge different lived, technical and professional experiences with the intention of upholding human rights concerns.

B. Learnings from WITNESS's participation in the C2PA

According to informal accounts from peers and colleagues, the detail and effort put on this [Harms, Misuse and Abuse Assessment](#) is uncommon within SDOs. WITNESS was able to carry out this process because we had the technical expertise, experience and resources to become part of the Coalition and to actively participate in it. This is important to note because it reflects the reality of a system that can otherwise be exclusionary, from a technical and operational point of view–for instance nonprofits are exempt from paying a membership fee in C2PA.

There are two aspects of the harms assessment process that are replicable and can inform future efforts to promote wider participation from civil society and human rights organizations. First, from the very early phases of the coalition, human rights use cases were presented by WITNESS to inform the development of the technical specifications. This included journalists, human rights defenders, inadvertent frontline witnesses, and others capturing evidence of human rights violations in situations that required considerations to uphold their privacy, safety and human rights, while also adding signals of trust to the content they create.

Second, WITNESS advocated for the [Guiding Principles](#) of the C2PA to include the need for the specifications to be reviewed for their potential to be abused and cause unintended harms, threats to human rights, or disproportionate risks to vulnerable groups globally. This laid the foundations for the creation of the 'Threats and Harms Taskforce' that WITNESS co-chairs and which was charged with developing the 'Harms, Misuse and Abuse Assessment'.

In addition to the procedural implementation that enabled the development of harm modeling within the C2PA, the methodology itself can be relevant for SDOs and other stakeholders seeking to include human rights considerations in the design of standards and technical specifications. For instance, the assessment was developed in consultation with a broad range

of global stakeholders that were not participating directly within the C2PA. If carried out intentionally, an indirect form of participation can be a solid alternative in situations where relevant stakeholders, especially those that may be in situations of disproportionate risk, have different levels of commitment and availability. Moreover, harms modeling is part of technology development processes that are often iterative and agile. Hence, assessing the potential risks and harms should not be a one off step that ends with the first version of the technical specifications. Quite the opposite, a commitment to incorporating human rights considerations into the development of technical specifications for new and emergent technologies requires a continued engagement as specifications are updated and technology evolves.

IV. Recommendations

1. Recognize that communities that may be impacted by new and emergent technologies, human rights groups and civil society organizations should play a key role in the design and development of new and emergent technologies, from early stages, including in Standards Development Organizations (SDOs).

2. Recognize SDO's as entities that have the power to influence and shape the development of new and emergent technologies, and as such, should be responsible for incorporating human rights considerations in technical standards–especially when these technologies may directly or indirectly lead to harms or threats to human rights, or bring disproportionate risks to vulnerable groups globally.

3. Call on states and SDOs to incorporate agile and continued harms modeling exercises and processes to bolster the implementation of the *Guiding Principles on Business and Human Rights* when designing, developing and implementing technical specifications.