SUBMISSION TO CALL FOR INPUTS PURSUANT TO UNITED NATIONS HUMAN RIGHTS COUNCIL RESOLUTION 47/23 (2021)

SUBMITTED ON: 23 FEBRUARY 2022
BY WITNESS

WITNESS

WITNESS is an international human rights organization that helps people use video and technology to protect and defend their rights. Working across five regions (Asia and the Pacific, Latin America and the Caribbean, the Middle East and North Africa, Sub-Saharan Africa, and the United States) alongside those most excluded or at-risk, our teams identify gaps, design solutions, provide guidance, and co-develop strategies that enable communities to hold the powerful to account and stand up for lasting change. We then scale this work globally on a systems level, sharing what we learn with communities facing similar issues and advocating grassroots perspectives to technology companies and other influential stakeholders to ensure they are translated into policies and solutions.

For more information, visit our website.

I. Introduction

This submission is meant to provide information and share WITNESS's views on the practical application of the Guiding Principles on Business and Human Rights to the activities of technology companies, especially as it relates to these three themes of the consultation:

1. Addressing human rights risks in business models (session one);
2. Human Rights Due Diligence and end-use (session two); and
3. Accountability and remedy (session three).

Based on WITNESS's experience addressing potential harms of emerging technologies at an early stage, most recently as co chairs of the Threats and Harms Task force of the Coalition for Content Provenance and Authenticity (C2PA), we argue for the need to advocate for the application of the Guiding Principles on Business and Human Rights in Standards Development Organizations (SDOs) and in specifications-development working groups.

This submission begins with a description of the Harms, Misuse and Abuse Assessment that we led as members of the C2PA, followed by a commentary on how this may address human rights risk in business models, how it may lead to ensure human rights due diligence, and how it may result in accountability and remedy. We finish this submission by offering recommendations for the international community and civil society organizations to bolster the Guiding Principles on Business and Human Rights within these bodies.

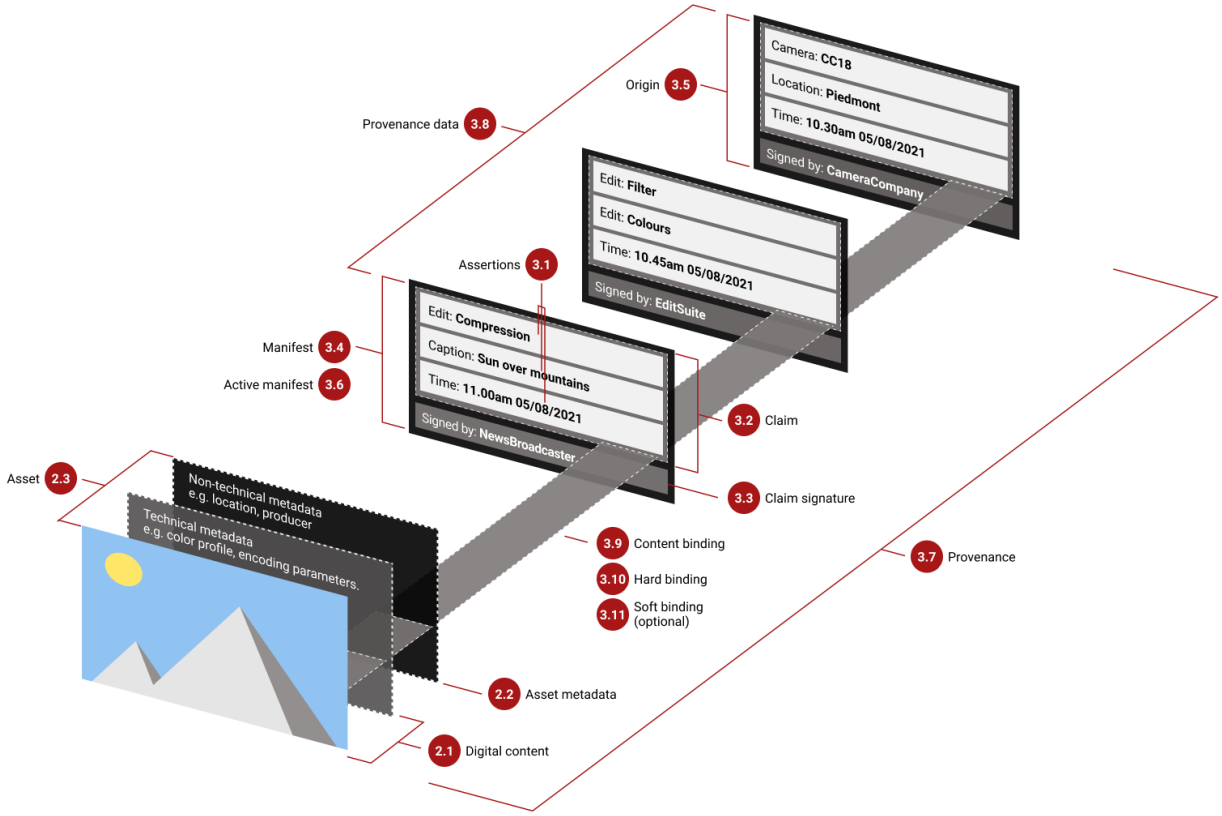
II. Harms, Misuse and Abuse Assessment of the C2PA technical specifications

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, privacy-protecting and opt-in 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's goal to create technical specifications is the latest, most significant move towards more widespread efforts driven by governments, platforms and public demand, as opposed to the opt-in and niche authenticity infrastructure that has existed till now. As of January of this year, version 1.0 of the Technical Specifications have been published. These specifications are open, so they may be readily included into any tool or device that creates or processes digital media. Within a C2PA-enabled ecosystem, it means that provenance information can be tracked from the moment that an image, video or audio (and later potentially documents) 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. Finally, a C2PA-enabled publishing tool would allow viewers to trace the source and history of this asset to determine whether they believe in it or not.



B. Harms Modelling in the C2PA

Although this could be a powerful tool for journalists, activists and others to help them discern truth from falsehood, it also opens up doors that could lead to potential harms to a broad range of individuals and communities, especially those that are already most at risk. In an effort to avert and mitigate these potential harms, WITNESS advocated for including clear global human rights and journalism use cases from the start, and for highlighting global human rights concerns in the Guiding Principles for C2PA designs and specifications:

- 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 of the framework;
- 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 globally.

Based on these principles, WITNESS, as members of the C2PA, was able to lead a harms modelling exercise which focuses on analysing 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. The process of harms modelling 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 modelling 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 modelling 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 use and unintended use. 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.

The Guiding Principles, the inclusion of human rights use cases from the beginning, as well as the harms modelling exercise have informed the continuous design of the specifications to address some concerns that have already been identified. They have for example served to establish that the specifications should be opt in, and have highlighted the need for features and user experience guidance that enables content creators to retain effective control of their data, including mechanisms for redacting sensitive information.

III. Commentary on advocating for the application of the Guiding Principles on Business and Human Rights in Standards Development Organizations (SDOs) and in specifications-development working groups

In developing strategies for the practical application of the Guiding Principles on Business and Human Rights to the activities of technology companies, we believe that it is necessary to consider the following:

1. The fast-evolving nature of technologies requires that a global range of human rights organizations and relevant civil society potentially impacted by these technologies participate in early stages of their design and development in order to understand, prepare and shape their potential impact. Standards-setting bodies are one significant area of early-stage development where there has been a distinctly low level of participation from these sectors of societies.
2. Tools, services, business models and industries may be determined by the standards that define their ecosystem. It is thus imperative that SDOs and specifications-development working groups be informed by human rights organizations and other civil society actors. It is also important that SDOs be held accountable for any action that may directly or indirectly lead to harms or threats to human rights, or disproportionate risks to vulnerable groups globally.

By various accounts, the detail and effort put on this Harms, Misuse and Abuse assessment is uncommon within SDOs. There are several lessons to highlight that could help bolster these Guiding Principles on Business and Human Rights to the activities of technology companies:

First, the assessment, in combination with the Guiding Principles for C2PA designs and specifications, establishes the intent of the coalition to assess the specifications for their capacity to cause unintended harms, threats to human rights, or disproportionate risks to vulnerable groups globally. This also establishes a first step towards accountability.

Second, the assessment has already resulted in a long list of potential harms for which we may already prepare. This preliminary report (as it will be ongoing, even after the publication of version 1.0) already includes existing and potential mitigation measures to each of the potential harms identified. As mentioned, some of these existing mitigation measures include having the C2PA specifications be opt-in, and that they include features and user experience guidance that enables content creators to retain effective control of their data, including mechanisms for redacting sensitive information.

Third, it recognizes that the assessment cannot be exhaustive, so it establishes the need for continuously monitoring the impact of the specifications, and for developing mechanisms to reflect an evolving landscape and addressing unidentified and unmitigated threats and harms.

Fourth, it informs other areas and activities for which the C2PA is responsible that could lead to averting and mitigating potential harms, including drafting guidance for implementers, user experience guidance, security considerations and an ‘explainer’ aimed for the general public.

Fifth, the assessment notes privacy and accessibility concerns that may perhaps not be addressed at the specifications level, but by the companies that are implementing them. This can help delineate areas of responsibility to establish accountability and to address human rights risks.

IV. Recommendations

Considering the discussion above, WITNESS offers the following recommendations:

- Recognize the need for a global range of human rights organizations and relevant civil society potentially impacted by these technologies to participate in early stages of the design and development of technologies, including in Standards Development Organizations (SDOs).
- Commission an investigation to understand the applicability of the Guiding Principles on Business and Human Rights in Standards Development Organizations and in specifications-development working groups, recognizing and building on top of the work that has already been done by human rights activists and organizations;
- Call on states and Standards Development Organizations to incorporate harms modelling exercises and to bolster the Guiding Principles on Business and Human Rights in the design and development of technical specifications and in their governance.