June 2025

Embedding Human Rights in Technical Standards:

Insights from WITNESS's Participation in the C2PA

Executive Summary

Technical standards are not neutral—they shape human rights in the digital age. This paper examines WITNESS's participation in the Coalition for Content Provenance and Authenticity (C2PA) to explore how human rights may be meaningfully embedded in technical standard-setting processes. Drawing from our experience with the C2PA—a coalition of mainly technology and news media companies developing content authenticity standards—we present five key insights: (1) embedding human rights in governance structures, (2) building in civil society participation, (3) conducting comprehensive harm assessments, (4) leveraging non-normative guidance documents, and (5) establishing post-standardization oversight mechanisms. While these insights emerge from a specific context and may not directly transfer to other Standards Development Organizations, they offer practical considerations for broader efforts to make technical standards more just, inclusive, and protective of fundamental rights.

Jacobo Castellanos Rivadeneira
Coordinator, Technology, Threats and Opportunities
jacobo@witness.org

Sam Gregory
Executive Director
sam@witness.org

Acknowledgements
We would like to thank Mallory Knodel, whose expertise, research, and collaborative thinking significantly informed both this report and ongoing port-standardization efforts.

As part of the research underpinning this report and ongoing work, we are also grateful to the following individuals for generously sharing their insights through interviews:

Corinne Cath
Tim Engelhardt
Jessie Lowell (National Network to End Domestic Violence)

Cover photo by Colin Lloyd (@onthesearchforpineapples) on Unsplash

WITNESS
www.witness.org

Introduction and Background

The work of embedding human rights in technical standard-setting builds on years of advocacy and investigation by a broad range of stakeholders that have long recognized that technical standards are not neutral, and that they play a pivotal role in promoting or undermining human rights in the digital age.

The Office of the United Nations High Commissioner for Human Rights (OHCHR) has underscored this relationship in its 2023 report to the Human Rights Council, where it calls for greater transparency, participation, and human rights due diligence in standard-development processes. The report also outlines clear responsibilities for States, companies, and Standards Development Organizations (SDOs) alike.

Civil society organizations and scholars have been at the forefront of efforts aimed at understanding the dynamics and processes that inhibit or facilitate inclusion, diversity, and the protection of human rights. Acting both as external stakeholders and members of SDOs, these groups have documented barriers to inclusive technical standards, such as exclusionary cultures, power asymmetries and governance gaps that make it difficult for marginalized voices to meaningfully influence technical decisions (Cath, 2023).

The technical community has also begun addressing these concerns through formal frameworks. Request for Comments 8280, *Human Rights Considerations for Internet Protocols*, published by the Internet Research Task Force, offers a methodology for evaluating how Internet protocols might impact human rights. This document lays out the groundwork for integrating rights-respecting principles directly into technical deliberations, providing a structure for standards developers to assess human rights implications.

At the policy level, multilateral initiatives have echoed these concerns. The Freedom Online Coalition's 2024 *Joint Statement on Technical Standards and Human Rights* reaffirmed the importance of multi-stakeholder engagement and human-centric design, aligning with the UN's broader push, reflected in instruments like the *UN Guiding Principles on Business and Human Rights,* for inclusive governance of digital technologies.

WITNESS's engagement with the Coalition for Content Provenance and Authenticity (C2PA) emerged within this context—not as a departure from ongoing efforts, but as a practical case in which we have sought to embed human rights considerations in the technical setting of a specifications-development initiative.

This paper is not intended as a case study on the successes or failings of the C2PA. Rather, it reflects on our participation as a way to draw lessons, surface tensions, and explore what it could mean to pursue human rights protections in a technically and politically complex standard-setting environment. Our goal is to contribute to the broader project of making technical standards more just, inclusive, and protective of fundamental rights.

Although not framed explicitly this way, the five insights below are organized to reflect different stages of the standardization process. Insights 1 and 2 highlight the governance conditions proposed for human rights to be protected structurally. Insights 3 and 4 focus on mechanisms for embedding human rights into the design of technical specifications, from harm assessments to non-normative guidance. Insight 5 explores the less examined—but significant—terrain of post-standardization, pointing to implementation and oversight as key spaces where human rights can either be upheld or undermined.

The C2PA as a Case Study

The need for verifiable digital content has become increasingly urgent. As disinformation proliferates online, and as AI-generated media becomes more sophisticated, a growing consensus has emerged around the importance of verifiable provenance—the ability to track where a piece of media comes from, how it was created, and whether it has been altered.

At the center of these efforts is arguably the Coalition for Content Provenance and Authenticity (C2PA), a multi-stakeholder initiative that is defining open technical standards to embed verifiable provenance metadata into digital media files.

Formed through a collaboration between major technology and media companies, including Microsoft, Adobe, Google, BBC, Meta, OpenAI, Sony, and others, the C2PA is building the infrastructure that is intended to help shape how trust in digital content is established across the internet. Its specifications are already being adopted into broader frameworks such as JPEG Trust and the emerging ISO 22144 standard, signaling its potential to become the default approach to content authenticity at scale.

Recognizing the implications of this work for human rights, civil society, and information integrity globally, WITNESS made a strategic decision to engage early in the development of the C2PA. This decision built upon over three decades of experience supporting communities to use video and technology for human rights advocacy, with a particular focus on tools and strategies to ensure the authenticity and credibility of digital media—especially in contexts where trust is fragile and the stakes are high. WITNESS had also contributed to a foundational white paper from the Content Authenticity Initiative that helped lay the groundwork for the Coalition, incorporating human rights perspectives and use cases to inform its design.

This early engagement was informed by extensive prior work in the field. Since 2012, WITNESS has collaborated with the Guardian Project on tools like ProofMode, which helps users capture and verify digital evidence. WITNESS has also published critical research, including *Mal-uses of AI-generated Synthetic Media and Deepfakes* (2018), *Ticks or It Didn't Happen* (2019) and *Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism* (2021). These reports highlight the need for privacy-preserving provenance and outline key considerations for the design of early-stage provenance systems. In addition, WITNESS led in-person workshops in Malaysia, Brazil, and South

Africa to gather global perspectives.

Together, these efforts have shaped WITNESS' understanding of what a human rights-centered, globally relevant content authenticity infrastructure should look like. They have also directly informed our participation in the C2PA, where we work to ensure that human rights values—and diverse, global needs and expectations—are reflected not only in the technical specifications, but also in governance and implementation.

This case study explores our experience engaging with the C2PA—a single purpose, corporate and media-led standards initiative. **As such, the insights shared here may not directly translate to more traditional SDOs. Nonetheless, we hope this contributes to ongoing conversations and encourages the development of tailored strategies for protecting human rights within specific SDOs and specification bodies.**

1. Embed Human Rights in the Governance of Standard-Setting Bodies

Meaningful integration of human rights principles into technical standards requires governance frameworks that explicitly value and create space for human rights considerations, participation from human rights defenders, and rights-preserving processes such as human rights impact assessments. Without these foundational structures, efforts to embed human rights may lack the institutional support necessary for effective implementation and meaningful impact.

How these governance foundations are created or what they ultimately look like will vary across different organizations. The C2PA represents a distinctive case compared to established SDOs, as its creation as a purpose-built coalition allowed early participants like WITNESS to directly shape its governance structure. One of our first efforts was to contribute to the development of the Guiding Principles, which explicitly reference human rights protections, such as protecting privacy, and the need to design with global needs and use cases in mind, including reviewing the Specifications for their potential to be misused or abused. This proved valuable because it added legitimacy and weight to arguments concerning technical decisions with human rights implications. Rather than positioning human rights concerns as subjective preferences or external considerations, we could anchor our input during the specifications-development process in collectively endorsed principles that participants had agreed to uphold.

Additionally, including human rights consideration into the Guiding Principles streamlined deliberation processes by establishing clear boundaries. When a proposed technical feature risked crossing a human rights threshold identified in the principles, we could reference the established framework rather than needing to reopen fundamental discussions about values and priorities. This prevented the frequent renegotiation of human rights commitments and allowed technical conversations to proceed more efficiently while maintaining appropriate safeguards.

Another key action was to advocate for the creation of the Threats and Harms Task Force, which established a dedicated space within the C2PA governance

structure focused on comprehensive threat modeling and societal impact assessment. The Task Force has two core functions: first, to threat model the Technical Specifications to identify potential security vulnerabilities; and second, to harm model the Specifications by examining how they might be misused or abused, assessing their broader societal impact, and developing strategies to prevent, mitigate, or reduce those harms. WITNESS co-chairs this Task Force and leads the work related to the harm assessment. The establishment of this Task Force represents a structural commitment to embedding human rights considerations directly into the technical development process, ensuring that harm prevention is treated as an integral component of standards development rather than an afterthought.

This experience with C2PA governance, while valuable, may not be easily replicated across all standardization contexts. Few standards bodies offer such early, flexible entry points for shaping foundational frameworks. In more established SDOs, governance models may be more rigid, slow to change, or resistant to incorporating rights-based mandates, particularly when they challenge entrenched technical or commercial priorities, or cultural and political dynamics. Even where human rights are acknowledged—whether through explicit human rights language, human rights use cases, safeguards against misuse, or broader considerations for diverse and global users—without mechanisms for enforcement or accountability, these protections risk becoming symbolic rather than substantive.

2. Civil Society Participation Must Be Built In, Not Left to Chance

The development of technical standards requires input from a broad range of stakeholders. Civil society participation can play a critical role in ensuring that diverse perspectives and lived experiences are reflected in the design, governance, and implementation of these standards. However, the presence of civil society remains limited, sporadic, or symbolic—constrained by barriers such as lack of resources, procedural expertise or simply the absence of forums where human rights considerations are recognized as relevant to "technical" decision-making.

For standards to truly reflect public interest and uphold human rights, civil society actors must not only be included early, but supported in ways that enable sustained and substantive engagement.

When the C2PA formed in 2021 to develop an open technical standard for capturing and sharing verifiable digital media provenance, WITNESS joined from the outset—building on years of prior work in this field. This early involvement provided the opportunity to advocate for human rights protections, safeguards against misuse, and design considerations for global and diverse users during the initial stages of both the Coalition's governance structure and its technical specifications.

Our ability to influence the coalition's structures was the result of a specific combination of factors: early intervention, access to adequate—though still insufficient—resources, and specialized expertise in provenance and

authenticity infrastructure informed by global consultations from broader civil society. However, this level of access and participation should not be the exception. Civil society participation must be structurally embedded into the governance of standards bodies and their processes—not left to chance or limited to those with the capacity and resources to self-navigate organizational structures, procedures and technical discussions.

3. Harm Assessments Can Be a Step In the Right Direction

Once appropriate governance structures are in place, the question for standard-setting bodies should not be whether to consider human rights, but how to do so effectively. One meaningful approach is to conduct a harm assessment—a structured process that draws from a broad range of global stakeholders to identify risks associated with the intended use, foreseeable misuse, and potential abuse of a technical standard.

Such an assessment should not only highlight potential harms, but also offer actionable recommendations to prevent, mitigate, or reduce both their likelihood and impact.

As co-chairs of the Threats and Harms Task Force, WITNESS led the development of the Harm Assessment of the C2PA Technical Specifications and their surrounding ecosystem. In consultation with human rights stakeholders and other experts, the methodology adapted elements from Microsoft's Harms Modeling Framework and BSR's Human Rights Due Diligence Assessment, drawing on approaches from value-driven design, human rights due diligence, and security threat modeling. The current version outlines 38 potential harms grouped into four categories:

1. Denial of consequential services
2. Infringement on human rights
3. Erosion of social and democratic structures
4. Physical harm or emotional psychological distress

Each identified harm is accompanied by due diligence actions that inform or shape the Technical Specifications themselves, supporting guidance or documentation, and/or non-technical and multilateral harm response mechanisms.

Recognizing that harms evolve as technology and adoption contexts change, WITNESS emphasized that the assessment must be an ongoing process. This perspective shaped both the structure of the harm assessment and its methodology. To ensure that the evaluation reflected diverse lived experiences, WITNESS also facilitated regional consultations in Nairobi, Bogotá, São Paulo and Bangkok, as well as thematic online discussions.

Yet despite the depth and scope of this process, key questions remain. Comprehensive harm assessments like this one require significant time, coordination, and resources—an investment that is often out of reach for many standards bodies,

particularly those without strong civil society engagement or funding. Even when such assessments are completed, their actual influence over technical outcomes is uncertain. Within the C2PA, it is still too early to tell whether identified harms and recommended mitigations will be meaningfully prioritized—especially when weighed against commercial pressures or implementer preferences.

4. Non-Normative Documents Matter—Especially for Participation and Rights-Informed Implementations

In many standardization processes, technical specifications are treated as the main—or sometimes the only—product. They are where requirements are formalized and where implementation is defined. In our experience with the C2PA, we believe that non-normative documents—those that offer best practice guidance or context—can play a significant role in shaping how human rights concerns are understood, communicated, and potentially acted upon. These documents may not be binding, but they can influence how implementers interpret the specifications, how the public engages with the standard, and whether broader participation is possible at all.

Along with the Technical Specifications and the Harm Modelling document, the C2PA has published an Implementation Guide, User Experience Guidance for Implementers, Security Considerations, and Guidance for Artificial Intelligence and Machine Learning. Several of these resources include recommendations that, if followed, can significantly strengthen human rights protections. For example, the User Experience Guidance document directly addresses privacy concerns identified in the Harm Modelling assessment by providing specific recommendations for how implementers should design interfaces that give content creators effective control over their personal information when it's included in C2PA Manifests (also known as Content Credentials).

One document that has already shown its potential is the Explainer, which, unlike the previous documents, speaks directly to the general public. WITNESS was heavily involved in developing early versions of this document, recognizing its strategic importance in making the C2PA specifications, objectives, and use cases accessible to diverse audiences. By creating materials that non-technical specialists could understand, the Explainer opened pathways for broader stakeholder participation and invited valuable input from communities who might otherwise be excluded from technical conversations.

Still, while these documents offer important opportunities, their influence remains inherently limited by their non-binding nature.

Implementers are not required to follow them, and there is no formal accountability mechanism to ensure that human rights recommendations are adopted in practice. In contexts where commercial or technical priorities dominate, non-normative documents can be easily sidelined or, worse, used to suggest that human rights concerns have already been addressed, without real implementation.

5. Post-standardization is where rights protections live or die

Human rights protections do not end with the publication of a technical standard—they're tested, and often determined, during implementation. Yet in many SDOs, responsibility for human rights considerations effectively ends once the specification is finalized. Oversight mechanisms are rare, and implementation is seen as out of scope. The Internet Engineering Task Force (IETF), for example, embraces a deliberately hands-off philosophy encapsulated in its well-known maxim: "We are not the Protocol Police."

In contrast, some technical ecosystems require more active oversight to ensure that implementation aligns with the intended purpose and requirements of the standard. The C2PA came to this conclusion following the release of Version 2.0 of its Specifications. It recognized that upholding its Trust Model would require more than verifying technical compliance—it also meant ensuring the integrity of the broader ecosystem. As a result, the C2PA is still currently developing a Conformance Program—a mechanism that will define requirements for implementers to become certified and officially recognized within the C2PA ecosystem.

While the decision to create a Conformance Program raised valid questions—about its necessity, governance, and scope—WITNESS viewed its approval as an opportunity to revisit and reinforce key human rights concerns. In particular, we began advocating for an additional safeguard: an independent mechanism to address harmful implementations that, while technically compliant, may nonetheless be unlawful, fundamentally misaligned with the C2PA's purpose, or in violation of human rights principles. This mechanism, still under conceptual development, would ideally be tasked with recommending appropriate accountability measures, including the potential revocation of certificates.

To guide the development of this proposal, WITNESS is grounding its advocacy in four core principles:

Meaningful Impact: The body's evaluations must be integrated into the Conformance Program and have the power to trigger tangible actions that mitigate, avert, or reduce harm.

True Independence: It must function independently from implementers, Certification Authorities, and the C2PA itself.

Global Representation: Its membership should reflect a diversity of geographies, perspectives, and lived experiences.

Transparency and Accountability: Its operations must be guided by a clear public charter, with processes that are open to scrutiny, review, and appeal.

One of the central questions these four principles implicitly seek to address is who should be responsible for oversight after a standard has been published. Assigning that responsibility solely to the standards body—particularly when it is industry-led—risks reinforcing the same power dynamics and incentives that often

marginalize human rights considerations. However, this does not mean that SDOs or specification bodies should be excluded entirely. On the contrary, they can play a vital role in enabling independent oversight by providing visibility, allocating resources, and creating space for such mechanisms to operate with impact. That said, oversight should be proportionate and context-dependent; in some cases, formal structures may not be necessary and could even prove counterproductive.

Note: This analysis reflects early-stage observations of the post-standardization process in the C2PA. As its Conformance Program continues to evolve, the descriptions and assessments contained herein may no longer accurately reflect current implementation or oversight practices.

Final Reflections for SDOs and civil society organizations

The insights shared in this paper raise practical challenges, particularly around the pace of development of standards. Integrating human rights considerations—through inclusive governance, harm assessments, global consultations, or oversight mechanisms—can slow down standardization. These processes take time, and they require taking on tasks that are often seen as external to technical work. But slowing down does not mean stalling—it may be what is needed to make standards more robust and ultimately more aligned with the diverse realities they impact.

Questions may also arise about the resources needed to support this kind of engagement, but there are untapped resources and energy that could support this work. Civil society organizations have consistently demonstrated a willingness to participate meaningfully in standard-setting processes, and to bring relevant expertise, networks, and global perspectives. What often limits this participation is not interest, but the lack of clear entry points, support, or impact. When rights-holders and their advocates see that their input is valued and consequential, they may be more likely to invest the time and resources needed for sustained engagement, especially during early stages when influence potential is greatest. Still, meaningful participation and execution will require investment from SDOs, including to facilitate a priori consultations that can help shape standardization objectives.

It is also worth considering the benefits that human rights integration can bring to SDOs and the standards they develop. Standards that proactively address human rights concerns may be more likely to achieve broad adoption. They can reduce legal and reputational risks for implementers, particularly as regulatory frameworks increasingly emphasize digital rights. Human rights-informed standards will also be more inclusive by design, expanding their potential user base and market reach. Additionally, the diverse perspectives that civil society brings can identify technical blind spots and edge cases that homogeneous development teams might miss, ultimately resulting in more robust and resilient specifications. In competitive

markets, standards that demonstrably protect user rights can become differentiating factors, while those that ignore human rights considerations may face public backlash, regulatory scrutiny, or implementation challenges that undermine their long-term viability.

Looking ahead, the challenge is not only to recognize the importance of human rights in standard-setting, but to translate that recognition into tailored strategies within specific SDOs and specification bodies.

WITNESS's experience with the C2PA is just one example. We believe that meaningful progress will come from building on these collective efforts: by sharing methodologies, co-developing practical tools, and working together to identify context-specific entry points for action. Collaboration between civil society, standards bodies, and other stakeholders is essential to ensure that technical standards serve the public interest and uphold fundamental rights.

With the support of

Digital Democracy Initiative

References

BSR. (n.d.). Human rights due diligence of products and services. https://www.bsr.org/en/reports/human-rights-due-diligence-of-products-and-services

Castellanos, J. (2024, August 8). Fortifying the truth in the age of synthetic media. WITNESS. https://blog.witness.org/2024/08/fortifying-the-truth-in-the-age-of-synthetic-media/

Castellanos, J. (2023, December 13). Generative AI in Latin America. WITNESS. https://blog.witness.org/2023/12/generative-ai-latin-america/

Cath, C. [Corinne]. (2023, April). Loud men talking loudly: Exclusionary cultures of internet governance [Primer]. Critical Infrastructure Lab. https://criticalinfralab.net/wp-content/uploads/2023/06/LoudMen-CorinneCath-CriticalInfraLab.pdf

Coalition for Content Provenance and Authenticity. (2024, April 9). C2PA technical specification version 2.2. https://c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html

Coalition for Content Provenance and Authenticity. (n.d.). Guiding principles for C2PA designs and specifications. https://c2pa.org/principles/

Coalition for Content Provenance and Authenticity. (n.d.). Harms modelling (C2PA specification v2.0). https://c2pa.org/specifications/specifications/2.0/security/Harms_Modelling.html

Freedom Online Coalition. (2024, October). Joint Statement on Technical Standards and Human Rights in the Context of Digital Technologies [Joint statement]. Freedom Online Coalition. https://freedomonlinecoalition.com/wp-content/uploads/2024/10/FOC-Joint-Statement-on-Technical-Standards-and-Human-Rights-in-Digital-Technologies.pdf

Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. Journalism, 0(0), 1–22. https://doi.org/10.1177/14648849211060644

Gregory, S. (2018, June). Mal-uses of AI-generated synthetic media and deepfakes: Pragmatic solutions discovery convening – Summary of discussions and next-step recommendations. WITNESS. https://www.researchgate.net/publication/341464776_Mal-uses_of_AI-generated_Synthetic_Media_and_Deepfakes_Pragmatic_Solutions_Discovery_Convening_June_2018_Summary_of_Discussions_and_Next_Step

Grover, G., ten Oever, N., Cath, C., & Sahib, S. (2021, January). RFC 8962: Guidance for registration data access protocol (RDAP) server operators on the use of data [RFC]. Internet Engineering Task Force. https://www.rfc-editor.org/rfc/rfc8962.html

Ivens, G., & Gregory, S. (2019). Ticks or it didn't happen [Report]. WITNESS. https://library.witness.org/product/ticks-or-it-didnt-happen/

International Organization for Standardization. (2025). ISO/CD 22144: Authenticity of information — Content credentials (Committee Draft). https://www.iso.org/standard/90726.html?browse=tc

JPEG Trust. (n.d.). JPEG Trust: A framework for secure and trustworthy imaging. Joint Photographic Experts Group (JPEG). https://jpeg.org/jpegtrust/index.html

Microsoft. (n.d.). Harms modeling [Internal architecture guide]. Microsoft Azure. https://learn.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/harms-modeling/

Parsons, A. (2020, August 3). CAI achieves milestone: White paper sets the standard for content attribution [Blog post]. Content Authenticity Initiative. https://contentauthenticity.org/blog/cai-achieves-milestone-white-paper-sets-the-standard-for-content-attribution

ten Oever, N., & Cath, C. (2017, October). RFC 8280: Research into human rights protocol considerations. Internet Engineering Task Force. https://datatracker.ietf.org/doc/rfc8280/

United Nations Office of the High Commissioner for Human Rights. (2023, September 18). Human rights and technical standard-setting processes for new and emerging digital technologies (Report No. A/HRC/53/42). United Nations. https://digitallibrary.un.org/record/4031373?v=pdf

United Nations Office of the High Commissioner for Human Rights. (2011). Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework. United Nations. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

Vazquez Llorente, R., Castellanos, J., & Agunwa, N. (2023, May 9). Generative AI in Africa. WITNESS. https://blog.witness.org/2023/05/generative-ai-africa/

WITNESS Asia-Pacific. (2020, March). Deepfakes: Prepare now (Perspectives from South and Southeast Asia) [Workshop summary]. WITNESS Media Lab. https://lab.witness.org/asia-deepfakes-prepare-now/

WITNESS. (2019, July 25). Deepfakes: Prepare now (Perspectives from Brazil) [Workshop summary]. WITNESS Media Lab. https://lab.witness.org/brazil-deepfakes-prepare-now/

WITNESS. (2020, February 19). What we learned from the Pretoria deepfakes workshop [Blog post]. WITNESS. https://blog.witness.org/2020/02/report-pretoria-deepfakes-workshop/