WITNESS

DIGITALMEDUSA

Mrs. Yordanka Ivanova,
Senior Policy Expert, AI Office,
Directorate-General for Communications Networks, Content and Technology,
European Commission

Mrs. Emanuelle du Challard,
Head of Copyright Unit
Directorate-General for Communications Networks, Content and Technology,
European Commission

Subject: EU Commission, users privacy should not be hindered in the name of copyright management

To whom it may concern,

In December 2025, in order to support the AI Act's requirement for general-purpose AI (GPAI) providers to respect copyright opt-outs, the European Commission issued a public consultation with the goal of identifying and setting up standards for the machine-readable "opt-out" protocols that AI developers must respect when training their models. The goal of the initiative and follow up activities was to invite rightsholders, GPAI providers, and other relevant groups to (a) share feedback on the technical feasibility of various Text and Data Mining (TDM) solutions and (b) participate in follow-up workshops.

The main goal of the efforts supported by the EU Intellectual Property Office (EUIPO) is to establish a widely accepted, regularly updated list of state-of-the-art protocols that AI companies must follow to show they are legally respecting copyright protections.

Months after the consultation, the standard developed by the **C2PA** (Coalition for Content Provenance and Authenticity) has emerged as a prime candidate for the "asset-based metadata" protocols the European Commission is evaluating. WITNESS and Digital Medusa both oppose this idea based on the points highlighted below:

The C2PA is not a fit for purpose protocol for for copyright management but for enabling content transparency, including AI generation or manipulation

The C2PA specification is an open technical standard that allows for the creation of a tamper-evident, cryptographic "nutrition label" for digital media. The way it works is through the use of a cryptographically signed "manifest" attached to media files to create a traceable history log. This technology tracks critical information, including the origin of the content, such as whether it was created by a camera or an AI model, and may log every edit made as the file

moves through various software tools. Additionally, it provides clear signaling to indicate whether an image, video, or text was synthetically generated or manipulated.

And while this standard is being widely adopted across several major platforms and tools, the sole purpose for it to exist is to facilitate content authenticity. It is natively integrated into smartphones like Google's Pixel devices, utilized by news organizations such as the BBC to combat misinformation, and employed by social media platforms like TikTok to label AI-generated content on user feeds. Users can verify these credentials directly through dedicated portals, including the OpenAI Verification Tool and the broader Content Credentials Verify Tool.

In a previous statement from WITNESS during the final stages of the General Purpose AI Code of Practice, we have stated the following:

Copyright Protection Must Not Be Based on Metadata Standards Not Intended For This Purpose
The CoP's dedicated copyright section raises significant concerns regarding metadata usage. While implementing appropriate, human rights-respecting copyright policies and ensuring lawful data mining are crucial, the final draft must not pursue methods that compromise privacy or misuse metadata. Specifically, we are worried that the CoP might inadvertently promote the use of metadata standards for rights management, like the Coalition for Content Provenance and Authenticity standards (C2PA), which were not designed for this purpose.

Relying on such standards risks creating gaps in intellectual property protection, undermining content ownership and fair use, and jeopardising privacy through potential Personally Identifiable Information (PII) misuse. C2PA does make explicit commitments to human rights and privacy in their guiding principles, but they have prioritised content provenance and authenticity as use purposes, explicitly excluding rights management from its core design.

The C2PA core specifications do not include assertions for attribution, and repurposing the standard for this function would require adding such assertions to the specs themselves. This reflects a deliberate design choice: keeping attribution out of the core specification allows companies to use provenance for authentication without introducing the privacy and surveillance risks that identifiable attribution data would carry. Where companies want to include identifiable information, extension technologies like the CAWG (Creator Assertions Working Group) already provide a pathway to do so, separately from the core specs. Preserving this separation is a key mechanism for protecting privacy rights: it keeps provenance usable as a tool for authentication without inextricably tying it to systems that enable identification. Pursuing copyright or rights-management functions through the core specification would erode this separation and the privacy protections it is designed to preserve.

Furthermore, utilising metadata standards for rights management could create a discriminatory two-tiered system. Only those with the technical capacity to leverage these standards would have their works fully protected, leaving others vulnerable. This creates an uneven playing field where content protection is determined by technical compatibility, not inherent rights.

Copyright protection should not hinder users privacy

Recently, WITNESS has issued a report named *C2PA Content Credentials and the Surveillance Risk: Adversarial Scenarios and Governance Gaps in the Content Provenance Ecosystem*, where we explore how provenance and authenticity systems could amplify surveillance risks if implemented without the right safeguards.

And while C2PA is a relevant and groundbreaking standard when it comes to providing the means for AI outputs and systems to be more transparent and inform users when they are being confronted with AI-generated and manipulated content, the system still needs to improve on the end of ensuring no personal data from individuals can be exposed.

The C2PA's reliance on cryptographically signed metadata poses significant safety risks by potentially stripping anonymity from content creators. By binding unique device footprints or personally identifiable information (PII) to media, the technology creates a "paper trail" that can expose filmmakers and whistleblowers. Furthermore, this transparency mechanism increases the threat of government subpoena, as authoritarian regimes or hostile actors could compel platforms and manufacturers to link cryptographic signatures to the real-world identities of users, thereby endangering human rights defenders.

Added to that, the use of such a protocol to tackle copyright infringement could also result in behavioural profiling of users. The C2PA's durable metadata structure facilitates behavioral profiling by enabling adversaries to track content creators across various online platforms. Because signing keys and editing patterns may serve as unique digital fingerprints, bad actors can analyze these metadata chains to map the movements, professional connections, and behavioral habits of individuals like human rights defenders, even when their real identities remain undisclosed.

Lastly, when it comes to the possibility of states mandating the use of the C2PA standards, we would warn the commission about the possibility of provenance tools being transformed into mandatory surveillance apparatuses that track content creators across platforms rather than simply explaining how media was produced. This shift creates a dangerous power imbalance where government-mandated "C2PA-certified" labels marginalize grassroots or citizen footage that lacks expensive, compliant infrastructure, effectively allowing authorities to weaponize "untrusted" designations to censor authentic voices and centralize control over what is considered truthful.

Mandating government-issued identifiers — national ID numbers, biometric identifiers, press credential numbers — in Content Credentials as a condition of content creation or distribution is incompatible with press freedom and freedom of expression. Requiring credentialing as a condition of platform distribution in contexts where credentialing is not genuinely voluntary for all creators converts an authenticity tool into a licensing mechanism. Both conditions should trigger mandatory assessment and public findings.

Conclusion

We urge the Commission to pursue other avenues for dealing with copyright management and reject the adoption of C2PA standards for this exclusive purpose. While we recognize the importance of copyright protection in the age of AI, attempting to retro-fit provenance and authenticity protocols for this purpose is fundamentally misaligned with their intended design and functionality.

As outlined, C2PA's architecture was built to promote transparency in media creation, not to manage the complexities of intellectual property rights. Imposing this system creates substantial risks: it threatens the privacy of content creators through potential PII exposure, facilitates dangerous behavioral profiling, and risks creating a discriminatory two-tiered system that marginalizes grassroots voices who lack the capacity for expensive, compliant infrastructure.

We, therefore, strongly encourage the Commission to pursue alternative, purpose-built solutions that protect intellectual property without sacrificing the safety, privacy, and freedom of expression of content creators worldwide.

Signatories

Contacts

Bruna Martins dos Santos, Policy and Advocacy Manager
WITNESS | bruna@witness.org

Jacobo Castellanos, AI & Trust Manager
WITNESS | jacobo@witness.org

Farzaneh Badiei, Founder
Digital Medusa | farzaneh@digitalmedusa.org