548 Market St
PMB 57274
San Francisco, California 94104

2/1/2024

VIA ELECTRONIC FILING

Director Laurie E. Locascio
U.S. Department of Commerce
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

Re: C2PA’s Response to NIST's Assignments Under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence Request for Information Related (88 Fed. Reg. 88368) (Docket Number: 231218-0309)

Dear Director Locascio,

The **Coalition for Content Provenance & Authenticity (C2PA)** appreciates the opportunity to respond to the National Institute of Standards and Technology (NIST) Assignments Under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence Request for Information (RFI). The Coalition for Content Provenance and Authenticity (C2PA) is an open standards organization created by members of the Content Authenticity Initiative (CAI), Project Origin, and others through the Joint Development Foundation under the Linux Foundation. The C2PA is collectively building an open technical standard to provide provenance and attribution for all forms of digital media. C2PA envisions these open standards will be adopted by content publishers, authors, and creators to build trust in the information ecosystem and ensure interoperability across the internet.

In response to the questions posed in the Federal Register Notice, we offer the following thoughts for your consideration. Our responses are to prompts posed in the request for information which better explain why we believe that the C2PA's open specification should be seriously considered when identifying existing standards, tools, methods, and practices.

C2PA Response

I. Developing Guidelines, Standards, and Best Practices for AI Safety and Security

Established Global Standard

We recommend NIST consider leveraging existing open industry technical standards for transparency in digital content. We would like to highlight the standard developed by the Coalition for Content Provenance and Authenticity (C2PA) to promote the development and deployment of safe, secure, and trustworthy AI systems. The C2PA is a coalition of nearly 100 companies, leading an ecosystem of over 2,000 organizations in the Content Authenticity Initiative (CAI) and Project Origin, committed to promoting the adoption of an open industry standard for content authenticity and provenance. 2023 marked the C2PA's most successful year to date as membership and engagement grew dramatically.

In 2020, the C2PA developed an open specification dedicated to the provenance of digital content. The standard, which is regularly updated based on new technologies, feedback from the community, and its needs, is currently on version 2.0. The C2PA standard is a Joint Development Foundation project under the Linux Foundation. The C2PA standard is also in the early stages of being fast-tracked as an ISO standard under ISO/TC 171/SC 2, the document file formats, EDMS systems, and authenticity of information working group. The C2PA expects ISO to approve the standard for publication by the end of 2024.

Transparency in Digital Content

The C2PA specification is already used to add transparency to digital content across multiple file formats (image, audio, video, and PDFs,). In this response, we would like to emphasize its utility and importance in managing and reducing risk from synthetic and generative technologies as described in section 2 of this RFI. The C2PA's open specification is already being used to: (1) add transparency to generative AI outputs;(2) add transparency to authentic; or non-synthetically generated content; and (3) help content consumers decipher the difference between synthetic and authentically created content. We believe the ability to add provenance information not only to AI-generated content but also authentic content, will remain incredibly important as content consumers seek to verify the authenticity of digital content. This ability will prove helpful to counter the liar's dividend, the dynamic by which bad actors dismiss the authentic as fake in order to avoid accountability.

Further, the C2PA's architecture is built to allow transparency across internet scale. Several unique tenets undergird the C2PA specification that make scale possible. We would like to highlight four critical aspects:

- *Interoperability*: The specification is written so that it can be adopted by any platform, OEM, or device. Any compliant mechanism can read, display, and add information to the provenance chain. This creates an interoperable provenance model for digital content that can move around the internet, with the content, in real-time.
- *Tamper Evidence*: Files signed with the C2PA specification become tamper-evident and therefore, content consumers will know if unauthorized changes have been made.
- *Multiple File Formats:* The specification can be applied to dozens of file formats ranging from image formats such as JPEG and PNG, to audio and video formats such as MP3 and MP4, as well as documents and text files. Also, provenance can be applied to machine learning files for data, models, and applications.
- *Opt-in:* For authentic media captures, the addition of Content Credentials (the cryptographically bound metadata about the file's origin and history) is optional and can be added as creators decide. However, generative AI platforms are increasingly using Content Credentials as a default setting on all generative outputs.

Other critical features such as a provenance chain to track edits, the ability to adapt to newer generative technologies, stability, and scalability are the reasons why the C2PA open standard has grown in popularity and is soon to be a global ISO standard.

Examples of use: Over the past 24 months various implementations of the C2PA open specification have been deployed for consumer and enterprise benefit. Through these implementations, Content Credentials are applied to both synthetic and authentic content around the world. Below is a description of several implementations today:

Image/Media Capture:

- As noted above, authentic images captured on a camera benefit from Content Credentials, because they can be identified as such and differentiated from synthetic creations. Software and hardware cameras now add Content Credentials to images to capture. With regard to software, Truepic's Lens SDK powers applications through its attested software to securely test and capture media, ensuring high integrity of data included in Content Credentials. Concerning hardware, Sony (Alpha), Nikon, and Leica

(M-11P) have released or announced DSLR cameras that can add Content Credentials to images captured by creators. Canon and Starling Labs announced a pilot program to pilot a platform that leverages the C2PA standard and blockchain based hashes to image captures. Also, Qualcomm announced its Snapdragon® 8 Gen 3 Mobile Platform will power any device to securely sign either an authentic original image or generate synthetic media with Content Credentials created on the device.

Generative AI:

- Some of the world's largest and most notable generative AI platforms have leveraged the C2PA specification to add Content Credentials. Adobe's Firefly and Microsoft's Image Creator by Designer (formerly Bing Image Creator) and DALLE-3 on Azure OpenAI, have led the way as the first generative AI platforms to mark all outputs with Content Credentials. To date, the platforms have created billions of images with Content Credentials. Hugging Face, in partnership with Truepic, launched two spaces on its platform adding Content Credentials and watermarks to open-source model outputs. Further, both Stability AI and Open AI have committed to marking outputs with Content Credentials aligning with the C2PA open standard.

Voice Cloning:

- Synthetic audio is becoming increasingly sophisticated. In early January, deep fake advertisements recently appeared that used video clips of the pop star and synthesized versions of her voice to look and sound as if she were doing a giveaway offer. In the run-up to New Hampshire's 2024 presidential primary, a deep-fake robocall of President Joe Biden telling voters not to vote in an apparent effort to suppress voter turnout demonstrates how sophisticated the technology has become and how it can be used to subvert our democratic process. Respeecher, one of the world's premiere voice cloning companies that make voice cloning tools for creatives and content creators realized how their tools could be used to sow misinformation recently described how they have integrated the CAI's open-source C2PA tool into their marketplace. As synthetic audio files are created on their servers, it is automatically cryptographically signed as being a product of the Respeecher marketplace and then when it's downloaded, it contains metadata with Content Credentials stating that it was converted into a different voice by Respeecher. Truepic also developed new functionality to add secure Content Credentials to audio capture on its SDK to power the highest integrity audio capture.

Flexibility: Compatible with Watermarking

With wide-scale adoption of the open specification, Content Credentials will be able to flow from website to website, and device to platform maintaining provenance information. However, in the meantime, while adoption is growing, the combination of watermarks and provenance on a single piece of digital content can be examined. Combining these two techniques can potentially increase the resilience of the transparency markings on the content. The watermark can serve as a backup reference in case the Content Credentials are lost, especially when content is shared with an incompatible service. The watermark can allow for the retrieval of a restored, signed version of the image from before the data was lost. For this reason, several watermarking companies like Digimarc and Steg.ai have joined the C2PA, increased collaboration, and begun creating products combining the two approaches. Digimarc released a browser plug-in, and Truepic and Steg.ai combined technologies on the Hugging Face platform, "Watermarked Content Credentials." All of these iterations illustrate the compatibility of C2PA cryptographic provenance and watermarking for interoperable provenance information with enhanced resiliency.

NIST Identification of Spec

The most beneficial approach and way to scale transparent digital content online is through the widespread adoption of Content Credentials by technology builders, content creators, and platforms. Establishing consumer awareness is also critical to increase adoption. Furthermore, NIST's identification of and recognition of the C2PA's open specification will help increase adoption and scale the benefits. Countless platforms, CDNs, and OEMs will look to NIST's guidance on implementing identified standards for the purposes of transparency.

The C2PA has demonstrated its commitment to supporting the USG and NIST's efforts to address issues in its Generative AI framework. We are proud members of the NIST Generative AI Public Working Group and have also signed the Cooperative Research & Development Agreement to join NIST's advisory committee on Gen AI. Further, the C2PA has consistently engaged and provided briefs on the C2PA standard and best practices for disclosure of media provenance to various USG stakeholders including the White House's Office of Science and Technology Policy, State Department, Department of Commerce, Department of Defence, DARPA, and other agencies. Further, the C2PA regularly briefs the US Senate and House of Representatives publicly and privately and has testified in various hearings.

Harm Mitigation

The C2PA developed its own threats and harms task force led by the non-profit Witness to identify unintended consequences to scaling of the standard. The specification is regularly updated to address and mitigate downside risks in implementation and/or potential harm. One key feature is the optionality of adding Content Credentials to content and the ability to redact any identifying information. The specification is designed in such a manner to ensure that populations under threat, or those who want to remain anonymous can still do so. The 2.0 version of the C2PA specification has further split information based on the originating entity. The updated specification recognizes assertions that are made by systems and mechanisms as the primary digital signature. All created assertions will ladder to a trust list ensuring that those mechanisms with the ability to sign have been vetted and approved.

If there is additional information that a person or other entity is asserting to and adding themselves (e.g. authorship), then that information will be identified separately as gathered assertions. This division helps mitigate potential misuse or harm by forcing content labels to carry identity (such as name, organization etc..) information. These are examples of how the C2PA regularly takes steps to ensure scaled transparency in the origin and history of digital content as safely as possible.

Future work to drive research and education

As consumer awareness of the C2PA spec increases, it will be important to assess to what extent content consumers understand content credentials, including what they do and do not convey. We recommend NIST lead scientific studies in this area for both media provenance and other disclosure methods, which could then inform future media literacy efforts. We also recommend that, in its guidance to the White House and USG agencies, NIST consider how the C2PA standard and Content Credentials can be used to add transparency to government operations and communications.

The C2PA and its members recognize that raising awareness and educating the public on digital content provenance is challenging. Future efforts to identify the best and most feasible ways to educate society, business, government, and other facets of society on its utility will be paramount. Equally important will be education and awareness campaigns on what digital content provenance does and does not do to mitigate misuse and misunderstanding of the technology and approach. The C2PA envisions various levels of educational programs will be needed, targeting the many facets of the digital ecosystem.

We look forward to continuing engagement with NIST and thank you for your consideration.

Sincerely,

Andrew Jenks
Chairman
Coalition for Content Provenance and Authenticity (C2PA)