Coalition for Content Provenance and Authenticity

C2PA Explainer

1.0, 2021-12-15: Release

Table of Contents

1. Introduction . . . 2
2. Goals and Non-goals . . . 3
3. Fundamentals (FAQs) . . . 4
    3.1. What does "Provenance" mean in the C2PA Specifications? . . . 4
    3.2. How is trust in digital assets established? . . . 4
    3.3. What does it mean that provenance data is cryptographically bound to the asset? . . . 5
    3.4. Can C2PA help with assets created from multiple sources? . . . 5
    3.5. What is redaction and how does it work? . . . 5
    3.6. TODO . . . 5
4. Use-case Examples . . . 7
    4.1. Helping consumers check the provenance of the media they are consuming . . . 7
    4.2. Enhancing clarity around provenance and edits for journalistic work . . . 7
    4.3. Offering publishers opportunities to improve their brand value . . . 7
    4.4. Providing quality data for indexer / platform content decisions . . . 7
    4.5. Assisting 'Intelligence' investigators to confirm provenance and integrity of media . . . 8
    4.6. Enhance the evidentiary value of critical footage . . . 8
    4.7. Enforcing disclaimer laws on retouched/edited images . . . 8
5. Stakeholder Feedback . . . 9
6. References . . . 10

Chapter 1. Introduction

The development of this Explainer is ongoing. This Explainer accompanies Version 0.8 of the Technical Specifications.

This Explainer accompanies the C2PA Specifications with the intent of providing further background and clarification on the development of the standard, its goals, mechanisms and guidelines.

This Explainer is not a technical document and is directed towards the general public.

Chapter 2. Goals and Non-goals

The goal of the C2PA specifications is to tackle the extraordinary challenge of trusting media in a context of rapidly evolving technology and the democratization of powerful creation and editing techniques. To this end, the specifications are designed to enable global, opt-in, adoption of digital provenance techniques through the creation of a rich ecosystem of digital provenance enabled applications for a wide range of individuals and organizations while meeting appropriate security and privacy requirements, as well as human rights considerations.

It is important to highlight that C2PA specifications do not provide value judgments about whether a given set of provenance data is 'true', but instead merely whether the provenance information can be verified as associated with the underlying asset, correctly formed, and free from tampering.

Chapter 3. Fundamentals (FAQs)

3.1. What does "Provenance" mean in the C2PA Specifications?

Provenance generally refers to the facts about the history of a piece of digital content assets (image, video, audio recording, document). C2PA enables the authors of provenance data to securely bind statements of provenance data to instances of content using their unique credentials. These provenance statements are called assertions by the C2PA. They may include assertions about who created the content and how, when, and where it was created. They may also include assertions about when and how it was edited throughout its life. The content author, and publisher (if authoring provenance data) always has control over whether to include provenance data as well as what assertions are included, such as whether to include identifying information (in order to allow for anonymous or pseudonymous assets). Included assertions can be removed in later edits without invalidating or removing all of the included provenance data in a process called redaction.

3.2. How is trust in digital assets established?

In the C2PA Specifications, trust decisions are made by the consumer of the asset based on the identity of the actor(s) who signed the provenance data along with the information in the assertions contained in the provenance. This signing takes place at each significant moment in an asset's life (e.g., creation, editing, etc.) through the use of the actor's unique credentials and ensures that the provenance data remains cryptographically bound to the newly created or updated asset.

To enable consumers to make informed decisions about the provenance of an asset, and prevent unknown attackers from impersonating others, it is critical that each application and ecosystem reliably identify the actor to whom a signing credential is issued. This is accomplished through the use of a *certification authority* (CA). CAs perform real-world due diligence to ensure credentials are only issued to actors who are whom they claim to be. In the world wide web, CAs verify that someone requesting a certificate to operate a secure web site owns and/or controls the site's domain name before issuing such a credential. For example, before issuing a certificate for https://c2pa.org/, the CA verified the requestor did in fact control C2PA's domain name before issuing a certificate for that site name.

Unlike the world wide web, C2PA will be used in multiple different application settings, and each setting will have its own requirements. Each application will therefore provide users with one or more *trust lists*, which are lists of certification authorities that issue credentials for that application. For example, in a news and media aggregation application or web site, each newspaper, television network, or other media organization has a globally-recognized identity, and there is only one such organization that operates with that name. In this situation, because brand marks and other visual indicators can and have been reproduced for the purposes of impersonation, consumers want to be certain the media they are consuming actually comes from the source it claims to be. This application would provide at least one trust list maintained by a professional or non-profit organization for journalists and media, which endorses certification authorities that ensure such credentials are only issued to the genuine organization through real-world due diligence.

In another example, an insurance company may employ provenance tracking for images, videos, and other media as

part of underwriting policies and servicing loss claims to be used by its own employees. In this case, only one trust list is applicable: the insurance company's own certification authority operated by its Human Resources department, which may already exist to employee credentials for other purposes. Here, it is not important that every participant have a unique name, as several employees may share the same name, but it is important to be assured all participants are employees of the insurance company, which the Human Resources department is certainly able to confirm and attest.

Once the credential is issued by a certification authority, the identity it has confirmed and placed in the credential cannot be altered by anyone else, including and especially the credential's owner. This allows the consumer or user to rely upon the identity presented in a signed asset.

3.3. What does it mean that provenance data is cryptographically bound to the asset?

The provenance data and the asset are the two parts of the same puzzle - a unique puzzle. The possibility of any other pieces ever matching, either by coincidence or by purposeful creation, is so low that it would be practically impossible. This is known as a hard binding.

In other words, any alteration to either the asset or the provenance, however insignificant, would alter the mathematical algorithm - the shape of the piece of the puzzle – in such a way that they would no longer match.

For more technical information on this see "Hard binding" in the glossary and the non-normative guidance.

3.4. Can C2PA help with assets created from multiple sources?

When one asset is created from a series of other assets, those sources are referred to as the ingredients. Each ingredient that is used in the final (composed) asset can be identified as part of the final provenance, including the addition of the provenance of each individual ingredient.

3.5. What is redaction and how does it work?

Redaction is the process of permanently removing information - in the world of C2PA it specifically refers to removing assertions.

For example, if a human rights organization wishes to remove from an image assertions about the photographer, it can do so via the redaction process. In addition, they could, at the same time, add more information about the time and location of where the image was taken. The act of redaction (and possibly adding addition information) becomes part of the provenance of the asset.

3.6. TODO

NOTE

Some of the additional areas that information will be provide about include: - Synthetic content - Privacy - Guarantees C2PA is providing the consumer - Who can be / Process of becoming a signer (signing entities) - Discussion of how provenance does not have to include everything from the origin to current due to the Trust Model.

Chapter 4. Use-case Examples

The following is a non-exhaustive list of potential and general use-cases of the C2PA Specifications. Some of these are taken from, or built upon, the use-cases developed within the Project Origin Alliance and the Content Authenticity Initiative (CAI) frameworks. Each use-case wil l be described using some generic personas to help make the flow clear.

For technical use-case examples, see [Non-normative guidelines].

4.1. Helping consumers check the provenance of the media they are consuming

Alice sends a video to a friend, Bob. The video includes text with alarming and controversial allegations. Bob immediately seeks confirmation of its validity, starting with its provenance.

The video that Alice sent contains C2PA provenance. With a C2PA-enabled application, Bob is able to establish that this video has been validated as being published by an organisation he can trust and is held in public high regard.

4.2. Enhancing clarity around provenance and edits for journalistic work

A photojournalist uses a C2PA-enabled capture device during a newsworthy event they are covering. The assets are then brought into a C2PA-enabled editing application, and after editing it, they are sent to a photo editor. The editor makes additional edits also using a C2PA-enabled application. The finalized asset is moved into the content management system of a news organization, which is also C2PA-enabled, before posting the asset to social media.

4.3. Offering publishers opportunities to improve their brand value

A news publisher is concerned about standards of public comprehension and brand value of its publications which it makes available online through a number of social media platforms. To improve audience confidence about their content, it wishes to provide a means for the audience to verify the content that originated through its output.

For content that is consumed without any C2PA provenance, the publisher hopes that the consumer will take extra steps to verify the provenance and authenticity of the asset, instead of immediately attributing it to the site on which it is published.

4.4. Providing quality data for indexer / platform content decisions

A news video is posted to a social media platform. By utilizing the C2PA-enabled provenance in the video, the social

media platform is able to verify that it came from the same source that posted it.

4.5. Assisting 'Intelligence' investigators to confirm provenance and integrity of media

An individual in a news/other context using open-source intelligence techniques (OSINT) can use the presence of C2PA provenance in assets to better confirm the history and integrity of media. Additionally, an individual may use a [decoupled binding database] to re-correlate relevant media to its C2PA provenance.

4.6. Enhance the evidentiary value of critical footage

A human rights defender manages to capture footage containing C2PA-enabled provenance of police violence during a protest. The human rights defender sends the footage to a human rights organization that verifies that the asset meets video-as-evidence criteria. The human rights organization redacts information about the defender using a C2PA-enabled editing application in order to protect their identity. The C2PA-verified asset is then used to improve the chances of that footage being admissible in court proceedings.

4.7. Enforcing disclaimer laws on retouched/edited images

To prevent dangerous stereotypes of ideal bodies, a government enacts a law that requires advertisers and social media influencers to specify that their image has been edited if any aspect of a body’s size, shape or skin has been altered. By having their C2PA-enabled editing application add information about each action performed, they can easily comply and the government can easily confirm.

Chapter 5. Stakeholder Feedback

NOTE Implementers and other stakeholders may already have publicly stated positions on this work. They will be listed here with links to evidence as appropriate.

Chapter 6. References

NOTE Any additional documents, articles, books, etc. that would be useful to read to get a better understanding of the work of the C2PA will go here.