C2PA Content Credentials Explained: Addressing Common Questions and Updates

By C2PA Technical Working Group

1. Introduction

1. Overview

The goal of the C2PA Specifications for Content Credentials is to tackle the extraordinary challenge of trusting media in the context of rapidly evolving technology and the democratization of powerful creation and editing techniques. To this end, the C2PA 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, organizations and devices, while meeting appropriate security and privacy requirements, as well as human rights considerations.

Provenance, as C2PA defines it, refers to the facts about the history of a piece of digital content (also known as an asset) in a form such as an image, video, audio recording, or document. At the heart of the C2PA specification is the Content Credential (also known as a C2PA Manifest), a cryptographically bound structure that records an asset's provenance.

It is worth noting that provenance is not always complete. It may happen that an asset is modified in a way that the provenance data is not updated. But later in the lifecycle of the asset, a new active Content Credential can be provided, so even if there is missing provenance information, the asset can still be trusted based on the signer of the active Content Credential.

It is important to highlight that Content Credentials do not provide value judgments about whether a given set of provenance data is 'true', but instead merely whether the provenance information is well-formed and free from tampering, valid and trusted (in that the signer of the Content Credential is associated with a known trust list). In addition, the provenance information can be verified as associated with the underlying asset ("a valid asset").

It is not a cure-all for misinformation, but instead seeks to mitigate against its threats in the digital domain. It complements media literacy, fact-checking, and digital forensics approaches such as deep-fake detection by providing an infrastructure to record all of that information in a tamper-evident structure, representing the provenance of any asset.

2. Technical Design Goals

One of the key design goals of the C2PA, that is deeply engrained in the choices that were made in the design of Content Credentials, is to prefer existing standards and practices over new ones. This is to ensure that Content Credentials can be easily adopted by existing systems and workflows, and are based on established & battle-tested technologies. It is for this reason that while the core technology is based on existing standards, such as X.509 certificates (RFC 5280), CBOR (RFC 8949), and JUMBF (ISO 19566-5), the C2PA specification also includes rich extensibility for other technologies, such as JSON-LD and XML, in order to support existing workflows and integration with other standards.

3. Trust in the C2PA Ecosystem

As described in the C2PA Explainer, the C2PA ecosystem is built on a well-defined Trust Model. This trust model is established through the use of X.509 certificates - the same technology behind SSL/TLS and PDF signatures as described below. The key to trust in the X.509 ecosystem is the chain of trust, which is established through a hierarchy of trusted Certificate Authorities (CAs). This allows for the verification of the authenticity of certificates and the entities they represent. This model has been around for decades and has proven to be effective in establishing trust in digital communications - from the trust lists of the CA/Browser Forum, the Adobe Authorized Trust List, AATL, and the European Union's EUTL.

The C2PA established an official C2PA Trust List as part of their 2.0 specification (Jan 2024, https://c2pa.org/specifications/specifications/2.0/specs/C2PA_Specification.html#_trust_lists). Until recently, that list was an internally maintained & unofficial list, but it is now handled as part of the recently established C2PA Conformance Program that is open to the public. Any organization operating a CA and abiding by well-defined requirements in the C2PA Certificate Policy that match or exceed industry best practices can be included in the C2PA Trust List.

This model of trust lists exists not only for the signer, but also for the trusted time stamp authority (TSA). Accordingly, the C2PA specification also includes a trust list for trusted time stamp authorities (TSA) and that too is covered as part of the C2PA Conformance Program.

4. Specifications vs. Implementations

The C2PA is a standards defining organization (SDO) that defines the specifications for Content Credentials. These specifications provide three sets of requirements:

- File format requirements, concerning the details of how the "binary bits" of the Content Credentials are constructed and embedded into various asset types;
- Claim Generators (the tools that create Content Credentials) implementation requirements;
- Verifiers (the tools that read and verify Content Credentials) implementation requirements.

While this means that the C2PA specifications define how Content Credentials are structured and how they shall be generated and verified, it does not define how these specifications are implemented in practice. The C2PA

specifications are designed to be flexible and extensible, allowing for a wide range of implementations across different platforms and devices. For example, a device with limited resources may choose to implement a subset of the C2PA specifications, while a more powerful device may choose to implement the full set of features. This flexibility is a key design goal of the C2PA specifications, allowing for a wide range of use cases and adoption scenarios.

However, if a claim generator or verifier does not implement the C2PA specifications correctly, it will not be able to generate or verify Content Credentials as intended. This is why the C2PA has established its Conformance Program, which provides a way for claim generators and verifiers to demonstrate that they conform to the C2PA specifications. The Conformance Program is open to the public and allows for third-party testing and validation of implementations against the C2PA specifications.

2. Specific Claims

1. Fundamental Problems

1. Use of X.509 Certificates

1. General

The use of X.509 certificates is not a fundamental problem, but rather a well-established and widely accepted practice in digital security for over 30 years now.

Content Credentials use X.509 certificates in the same way that PDF has done, since introduced in PDF 1.3 (2000) and which was later standardized by the EU initially as ETSI TS 103 172 and later updated as ETSI EN 319 142. Additionally, both ETSI and the ISO have standardized equivalent uses of X.509 certificates in other container formats in standards known as CAdES (ETSI EN 319 122), XAdES (ETSI EN 319 132), ASiC (ETSI EN 319 162) and JAdES (ETSI TS 119 182-1).

This same approach is what is used by web browsers to verify the authenticity of TLS/SSL certificates, which are also X.509 certificates.

Where the C2PA does differ is that it has established it own C2PA Trust List, rather than relying on existing lists such as the CA/Browser Forum's CA/Browser Forum or Adobe's AATL. This is because the C2PA is focused on a different problem space (digital content provenance), where the importance of ensuring that only trusted entities can issue Content Credentials, is key. This trust list is open to the public via our Conformance Program.

2.1.1.2. Conformance Program Overview

The C2PA Conformance Program provides assurance that products producing and consuming Content Credentials adhere to the Content Credentials specification, and fulfill a set of security requirements to ensure they are producing and validating Content Credential data correctly.

This Conformance Program is a risk-based, transparent and unbiased governance process intended to hold Generator Products, Validator Products and Certification Authorities accountable to the Content Credentials specification, the

Certificate Policy, and the Implementation Security Requirements. Conforming products are placed on a publicly accessible list. This conveys confidence in the implementation and its security to the public, and guarantees interoperability across products in the Content Credentials ecosystem.

Makers of Generator Products or Validator Products, and Certification Authorities wishing to participate in the program must execute a legal agreement with the C2PA, adhere to the agreement requirements, and provide evidence of conformance in order to be considered conformant.

2.1.1.2.1. Security Levels

Since v1.0, the C2PA has published their C2PA Security Considerations. This document summarizes recognized security threats to the C2PA technology and was used as a foundation to define the C2PA Generator Product Security Requirements. In particular, the C2PA Generator Product Security Requirements document translates the recognized security threats into security objectives and defines security requirements that satisfy these objectives up to different levels. These levels can be utilized by relying parties to develop confidence that the conformant C2PA Generator Product has been used as intended (e.g., not compromised).

Similarly to the C2PA specification, the C2PA Generator Product Security Requirements do not define exactly how the products are to be implemented. Instead, they require certain properties and characteristics across the elements of the Content Credentials. These include, among others, the security of digital contents and assertions at generation and in transit, security of the environment generating the Content Credentials, and security of the signing key.

The definition of Generator Product Security Requirements highlights the importance of implementation security. The implementation security matters should be expressed and recorded and can be considered by relying parties when analysing the provenance of digital content. The security level is recorded within the C2PA certificate granted to the signer by the Certificate Authority after analysis of both static (i.e., architectural analysis) and dynamic (i.e., implementation security state at the time of certificate issuance) evidence of meeting the security requirements.

For additional details on the Conformance Program, its governance, and the security levels and requirements, please consult the public documentation.

2. Manifests

1. General

A C2PA Manifest consists of three components: Assertions, the Claim and the Claim Signature. The assertions represent the statements (of fact) made by the signer or other actors that wish to have that information attributed to them. The claim is a structure that connects the assertions to the signer and the claim signature is a digital signature that provides the tamper-evident binding for the Manifest as a whole.

The C2PA specification lists 20 standard assertions that can be used in a Manifest, but it is not limited to those. The C2PA specification also allows for custom assertions to be defined, when it is necessary store information that is not provided for directly by the specification. Two of these standard assertions are required to be present in every C2PA Manifest: an actions assertions and a hard binding assertion. The actions assertion describes the actions that the signer is asserting were performed on the asset, such as "created", "edited" or "published". The hard binding

assertion is a cryptographic binding that ties the Manifest to the asset itself, ensuring that any changes to the asset will invalidate the Manifest.

While most of the assertions are unique to the C2PA, there are some assertions that represent ways to store inside of the C2PA Manifest information that could also be stored in other metadata formats, such as IPTC Photo Metadata, XMP or EXIF. However, by bringing that information inside the C2PA Manifest, it can serve not only as a single source for this information (without requiring the use of multiple metadata formats) but also enables that information to be directly attributed to one or more actors - something not possible with that information today.

The majority of these assertions, as well as the claim and the claim signature are all serialized as CBOR, RFC 7049, but in order to support existing workflows and other standards, the C2PA specification also includes rich extensibility for other technologies such as JSON-LD, XML and arbitrary binary data. This allows for the inclusion of additional information that may be relevant to the asset in a specific industry workflow such as DDEX or ODRL.

2. Ingredients

Provenance represents the history of an asset, including the actions that were performed on it, the actors that performed those actions, and any other relevant information. This is not only about the creation of the asset, or about the current state of the asset, but also about the history of the asset and its transformations over time. Each of these points in time can have a Content Credential associated with it, providing a verifiable record of the asset's history. These "points in time" are called ingredients in the C2PA specification, and they are represented as one of the standard types of assertions.

There are three types of ingredients: parents, components, and inputs, each serving a different possible relationship between itself and the current asset. A parent ingredient is a direct predecessor of the current asset, such as a source image that was used to create the current asset. A component ingredient is an asset that is part of the current asset, such as a video clip that is part of a larger video. An input ingredient represents information used to help create/edit the current asset, such as the prompt provided to a Generative AI system.

As ingredients are added to a new Content Credential, they are validated against the requirements of the C2PA specification, and the validation results are recorded in the Manifest. This allows for the verification of the integrity of the asset as well as its history.

3. Hard bindings

A hard binding is type of assertion that represents a cryptographic binding that ties the Manifest to the asset itself, ensuring that any changes to the asset will invalidate the Manifest. This is a key feature of the C2PA specification, as it provides the tamper-evident binding between the Manifest and the asset - which is why these assertions are required in every C2PA Manifest.

Because the C2PA specification supports a wide range of asset types, many of which are themselves complex formats that may be quite large and/or be compose of multiple files, it was necessary to define multiple ways to bind the Manifest to the asset. The C2PA specification defines five types of hard binding assertions - some of which are generic (e.g. simple data/byte-range hashing) and others that are specific to certain container types (e.g., BMFF hashing and collection hashing). By providing these different types of hard binding assertions, the C2PA specification allows

implementers the ability to choose the best type for the asset they are working with, as well as the implementation's operating environment. For example, some hardware implementations are unable to keep the entire asset in memory, so they may choose to use a general box hash instead a data hash.

4. Timestamping

Adding a trusted time-stamp to a C2PA Manifest serves to establish, that at given point in time, this exact asset and its associated C2PA Manifest existed. The most common way to accomplish this is to contact (via the Internet) a trusted Time Stamp Authority (TSA) and store their response directly in the certificate. Performing these actions at the time of signing is recognized as the "T" (time-stamped) level of signatures by the European Union's Advanced Electronic Signature (AdES) standards.

In some provenance workflows, however, the C2PA Manifest needs to be created offline, because it is not possible to obtain a trusted time-stamp from a TSA at the time of signing. However, the lack of a time-stamp means that signing certificates will expire after a certain period of time, thus leading to an invalid C2PA Manifest at that point. To prevent that expiration, a trusted time-stamp can be added at a later point in time (provided the certificate has not yet expired) for that C2PA Manifest and (in the case of the active manifest) its associated asset. This is accomplished through the use of a time-stamp assertion. By providing for the addition of a time-stamp at any future point (or at many future points), it allows one to keep a C2PA Manifest valid for a longer period of time, even if the original signing certificate has expired. This approach is recognized as the "A" (archival) level of signatures by the EU's AdES standards.

5. Certificate Status

In addition to time-stamps, the C2PA specification also recommends the inclusion of certificate status (sometimes called revocation status) information to be present in the C2PA Manifest. This is to ensure that the certificate used to sign the Manifest is valid at time of signing and had not been revoked. The C2PA specification recommends that the Claim Generator retrieve that information via OCSP. However, since OCSP requires an online connection to the CA, it is not always possible to retrieve that information at the time of signing. Therefore, just as with time-stamps, the C2PA specification allows for the inclusion of certificate status at a later point in time. This is accomplished through the use of a certificate status assertion. Using this approach is recognized as the "LT" (long term) level of signatures by the EU's AdES standards.

When combined with the time-stamp assertion, one can achieve the "LTA" (long term with archival) level of signatures recognized by the EU's AdES standards as a long-term signature.

6. Externally hosted manifests

In order to support asset types that cannot support embedding Content Credentials (e.g., text formats, Camera RAW, etc), for use cases where the C2PA Manifest may be too large or for situations where the asset cannot be modified, the C2PA specification allows for the C2PA Manifest to be stored externally to the asset. The C2PA Manifest is stored identically outside as it is inside (i.e., a JUMBF superbox) and is validated in the same way as an embedded Manifest. This means that any modifications to the asset will invalidate the Manifest, regardless of whether it is embedded or stored externally.

The C2PA specification defines a number of different ways in which a validator (or other manifest consumer) can discover the sidecar Manifest, including the use of a sidecar file with a specific file extension and http link headers. For formats that support embedded XMP, a URL can be provided to a external C2PA Manifest. Because the use of URLs can indeed open up attack scenarios (e.g., "phone home" situations), the C2PA specification recommends (but **does not** require) validators to look for sidecar manifests.

2. Durable Content Credentials

As defined in the C2PA specification, a Durable Content Credential "is a Content Credential for which there exists one or more soft bindings that enable its discovery in a manifest repository".

This means that the Content Credential (or a copy thereof) has been stored in a manifest repository (e.g., a hosted storage system of manifests) from which it can be retrieved through the use of the soft binding resolution API, based on either provided or dynamically generated soft bindings. This optional approach to manifest retrieval is provided because the Content Credential may be routinely removed or corrupted by legacy or non-Content Credential capable platforms during distribution. This is common, for example, on social media platforms that display asset renditions (e.g., altering the resolution, form factor or quality of the digital content) that do not have the appropriate C2PA Manifests declaring those modifications. Whilst these renditions may not create user perceptible change, they nevertheless change the underlying binary representation of the digital content.

Soft bindings provide a means for identifying the active manifest, and associated C2PA Manifest Store, that has become 'decoupled' from its associated asset in these circumstances. The C2PA specification defines two different types of soft bindings: watermarks and fingerprints.

1. Watermarking

The C2PA specifications describe a standardized way to apply and look-up watermarks to promote interoperability, and is independent and agnostic to the watermarking technology used. Both proprietary and open-source watermarking algorithms may be used, and examples of both exist on the C2PA Soft Binding Algorithm List, which is maintained by the C2PA within their github repository. The purpose of this list is to assign unique names to each algorithm so that they may be described within a C2PA Manifest in an interoperable way, using the soft binding assertion. Anyone may submit algorithms for inclusion on this list, including non-C2PA members, by filing a pull request (PR) on that github repository. This enables a rich and open ecosystem of watermarking algorithms to be used. Submissions are reviewed promptly by the relevant technical working group of the C2PA to guard against spam entries to the list prior to acceptance.

Invisible (i.e., substantially human imperceptible) watermarks embed a unique identifier within the asset's digital content. This modifies the content and that modification is recorded using the relevant action and soft binding assertions. The purpose of the watermark is to discover the active manifest, which necessitates the storage of the manifest (but not the asset) within a manifest repository. Being agnostic to the watermarking technology choice, the C2PA does not proscribe any centralized repository or related look-up service. Manifest repository operators are distinct from the C2PA and may implement their own policies in relation to user privacy and management of their entries. Rather, the C2PA describes an interoperable way to query look-up services offered by repositories, using the

Soft Binding Resolution API specification. Depending on the implementation choice, the API may be invoked either by sending an asset to the service, or by locally extracting the soft binding (e.g. watermark) and sending only the identifier to the service. The location of the service for a given soft binding technology may be described either within the Soft Binding Algorithm List or may be discovered using a decentralized look-up process. The latter is an open area of discussion within the C2PA community, including the deployment of some Proofs of Concept (PoCs) by various C2PA members that demonstrate how such a lookup facility could be implemented.

2.2.2. Fingerprinting

Fingerprinting is the general process of determining a set of inherent properties computable from digital content that identifies the content or near duplicates of it. Just as with watermarking there are many different algorithms that can be used for fingerprinting, that vary based on supported media type, computational complexity, and other factors. The C2PA Soft Binding Algorithm List provides a list of algorithms that can be used for fingerprinting.

If a given asset has neither an invisible watermark nor an embedded C2PA Manifest, a fingerprint may be used as a fallback search key to discover the C2PA Manifest Store. The C2PA Soft Binding Resolution API specification supports discovery based on fingerprints as well. As algorithms used for fingerprinting typically offer near - rather than exact - matching, it is recommended that the Content Credentials retrieved using a fingerprint as a search key are presented to the user for manual review.

3. C2PA as an organization

The C2PA is an Linux Foundation Joint Development Foundation (JDF) project, which means that it is a collaborative effort of over 300 organizations, institutions and individuals. It also has established liaisons with many other SDOs including the ISO, PDF Association, ONVIF and others. The C2PA is governed by a Steering Committee (SC) composed of 11 members.

The development of the C2PA specifications are done in the organization's Technical Working Group (TWG), which is open to all members of the C2PA. The TWG is responsible for the development of the specifications, as well as the conformance program and other related activities. The TWG is supported by a number of sub-groups, called Task Forces, each focused on a specific area of the C2PA specifications or related activities. Some of the active Task Forces include Conformance, Threats and Harms, User Experience, AI/ML, Live Video and Audio. Each one has a pair of co-chairs, who are responsible for the operations of the Task Force and report to the TWG.

All work in the TWG (and its TFs) takes place either synchronously via video calls or asynchronously via the C2PA's Slack instance and GitHub repositories. Every member has access to all of these channels and can participate in the discussions, provide feedback and contribute to the development of the specifications. All decision making is done via standard consensus-based processes, as used in many other SDOs, such as the ISO and W3C.

4. Independent Reviews

While the C2PA itself has not requested any independent reviews of its specifications, several independent reviews have been conducted by various organizations and individuals. These reviews have focused on different aspects of

the C2PA specifications, including security, privacy, and usability. Most of the reviews we have received have been sent directly to us, but we are aware of at least two public reviews that have been published online:

- Reducing Risks Posed by Synthetic Content An Overview of Technical Approaches to Digital Content Transparency | NIST
- NSA and Five Eyes partner CSI report on Content Credentials