# 1. Introduction

# Overview

# With the increasing velocity of digital content and the increasing availability of powerful creation and editing techniques, establishing the provenance of media is critical to ensure transparency, understanding, and ultimately, trust.We are witnessing extraordinary challenges to trust in media. As social platforms amplify the reach and influence of certain content via ever more complex and opaque algorithms, mis-attributed and mis-contextualized content spreads quickly. Whether inadvertent misinformation or deliberate deception via disinformation, inauthentic content is on the rise.

Currently, creators who wish to include metadata about their work (for example, authorship) cannot do so in a secure, tamper-evident and standardized way across platforms. Without this attribution information, publishers and consumers lack critical context for determining the authenticity of media.

Provenance empowers content creators and editors, regardless of their geographic location or degree of access to technology, to disclose information about who created or changed an asset, what was changed and how it was changed. Content with provenance provides indicators of authenticity so that consumers can have awareness of who has altered content and what exactly has been changed. This ability to provide provenance for creators, publishers and consumers is essential to facilitating trust online.

To address this issue at scale for publishers, creators and consumers, the Coalition for Content Provenance and Authenticity (C2PA) has developed this technical specification for providing content provenance and authenticity.

# Scope

This specification describes the technical aspects of the C2PA architecture; a model for storing and accessing cryptographically verifiable information whose trustworthiness can be assessed based on a defined [trust model](#_bookmark78). Included in this document is information about how to create and process a C2PA Manifest and its components, including the use of digital signature technology for enabling tamper-evidence as well as establishing trust.

It is important to note that the specification does not negatively impact content accessibility for consumers.

# Technical Overview

The C2PA information comprises a series of statements that cover areas such as asset creation, authorship, edit actions, capture device details, bindings to content and many other subjects. These statements, called [Assertions](#_bookmark12), make up the provenance of a given asset and represent a series of trust signals that can be used by a human to improve their view of trustworthiness concerning the asset. Assertions are wrapped up with additional information into a [digitally signed](#_bookmark77) entity called a [Claim](#_bookmark13).

These assertions, claims, credentials and signatures are all bound together into a verifiable unit called a [C2PA](#_bookmark14) [Manifest](#_bookmark14) by a hardware or software component called a Claim Generator. The set of C2PA Manifests, as stored in the asset’s C2PA Manifest Store, represent its provenance data.

**C2PA Manifest**

**Assertions**

**Claim**

**Claim Signature**

*Figure 1. A C2PA Manifest and its constituent parts*

# Establishing Trust

The basis of making trust decisions in C2PA, our [Trust Model](#_bookmark78), is the identity of the actor associated with the cryptographic signing key used to sign the claim in the Active Manifest. The identity of a signatory is not necessarily a human actor, and the identity presented may be a pseudonym, completely anonymous, or pertain to a service or trusted hardware device with its own identity, including an application running inside such a service or trusted hardware. C2PA Manifests can be validated indefinitely regardless of whether the cryptographic credentials used to sign its contents are later expired or revoked.

# 2. Glossary

# Introductory terms

## Actor

A human or non-human (hardware or software) that is participating in the C2PA ecosystem. For example: a camera (capture device), image editing software, cloud service or the person using such tools. An organization or group of *actors* may also be considered an *actor* in the C2PA ecosystem.

## Signer

An *actor* (human or non-human) whose credential’s private key is used to sign the *claim*. The *signer* is identified by the subject of the credential.

## Claim generator

The non-human (hardware or software) *actor* that generates the *claim* about an *asset* as well as the *claim signature*, thus leading to the *asset*'s associated *manifest*.

## Manifest consumer

An *actor* who consumes an *asset* with an associated *manifest* for the purpose of obtaining the *provenance data* from the *manifest*.

## Validator

A *manifest consumer* whose role is to perform the actions during [*Validation*](#_bookmark88).

## Action

An operation performed by an *actor* on an *asset.* For example, "create", "embed", or "apply filter".

# Assets and Content

## Digital content

The portion of an *asset* that represents the actual content, such as the pixels of an image, along with any additional technical metadata required to understand the content (e.g., a colour profile or encoding parameters).

## Asset metadata

The portion of an *asset* that represents non-technical information about the *asset* and its *digital content*, as may be stored via standards such as Exif or XMP.

## Asset

A file or stream of data containing *digital content*, *asset metadata* and optionally, a *C2PA Manifest*. For the purposes of this definition, we will extend the typical definition of "file" to include cloud- native and dynamically generated data.

## Derived asset

A *derived asset* is an *asset* that is created by starting from an existing *asset* and performing *actions* to it that modify its *digital content* and *asset metadata*. An audio stream that has been shortened or a document where pages have been added.

## Asset rendition

A representation of an *asset* (either as a part of an *asset* or a completely new *asset*) where the *digital content* has had a 'non-editorial transformation' *action* (e.g., re-encoding or scaling) applied but where the *asset metadata* has not been modified. A video file that is re-encoded for reduced screen resolution or network bandwidth.

# Core Aspects of C2PA

## Assertion

A data structure which represents a statement asserted by an *actor* concerning the *asset*. This data is a part of the

*C2PA Manifest*.

## Claim

A digitally signed and tamper-evident data structure that references a set of *assertions* by one or more *actors*, concerning an *asset*, and the information necessary to represent the *content binding*. If any *assertions* were redacted, then a declaration to that effect is included. This data is a part of the *C2PA Manifest*.

## Claim signature

The digital signature on the *claim* using the private key of an *actor*. The *claim signature* is a part of the *C2PA Manifest*.

## C2PA Manifest

The set of information about the *provenance* of an *asset* based on the combination of one or more *assertions*

(including *content bindings*), a single *claim*, and a *claim signature*. A *C2PA Manifest* is part of a *C2PA Manifest Store*.

## C2PA Manifest Store

A collection of *C2PA Manifests* that can either be embedded into an *asset* or be external to its *asset*.

## Origin

The *C2PA Manifest* in the *provenance data* which represents the software or device that initially created the *asset*. Details on how one determines which *C2PA Manifest* is the *origin* are left for specification.

## Active Manifest

The last manifest in the list of *C2PA Manifests* inside of a *C2PA Manifest Store* which is the one with the set of *content bindings* that are able to be validated.

## Provenance

The logical concept of understanding the history of an *asset* and its interaction with *actors* and other *assets*, as represented by the *provenance data*.

## Provenance data

The set of *C2PA Manifest*s for an *asset* and, in the case of a *composed asset*, its *ingredients*. A *C2PA Manifest* can reference other *C2PA Manifest*s.

## Authenticity

A property of *digital content* comprising a set of facts (*provenance data* and *hard bindings*) that can be cryptographically verified as not having been tampered with.

## Content binding

Information that associates *digital content* to a specific *C2PA Manifest* associated with a specific *asset*, either as a *hard binding* or a *soft binding*.

## Hard binding

One or more cryptographic hashes that uniquely identifies either the entire *asset* or a portion thereof.

## Soft binding

A content identifier that is either (a) not statistically unique, such as a *fingerprint*, or (b) embedded as a *watermark* in the identified *digital content*.

## Trust signals

The collection of information that can inform an *actor’s* judgment of the trustworthiness of an *asset*. These are in addition to the *signer* of a *claim*, upon which the fundamental trust model relies.

# Additional Terms

## Watermark

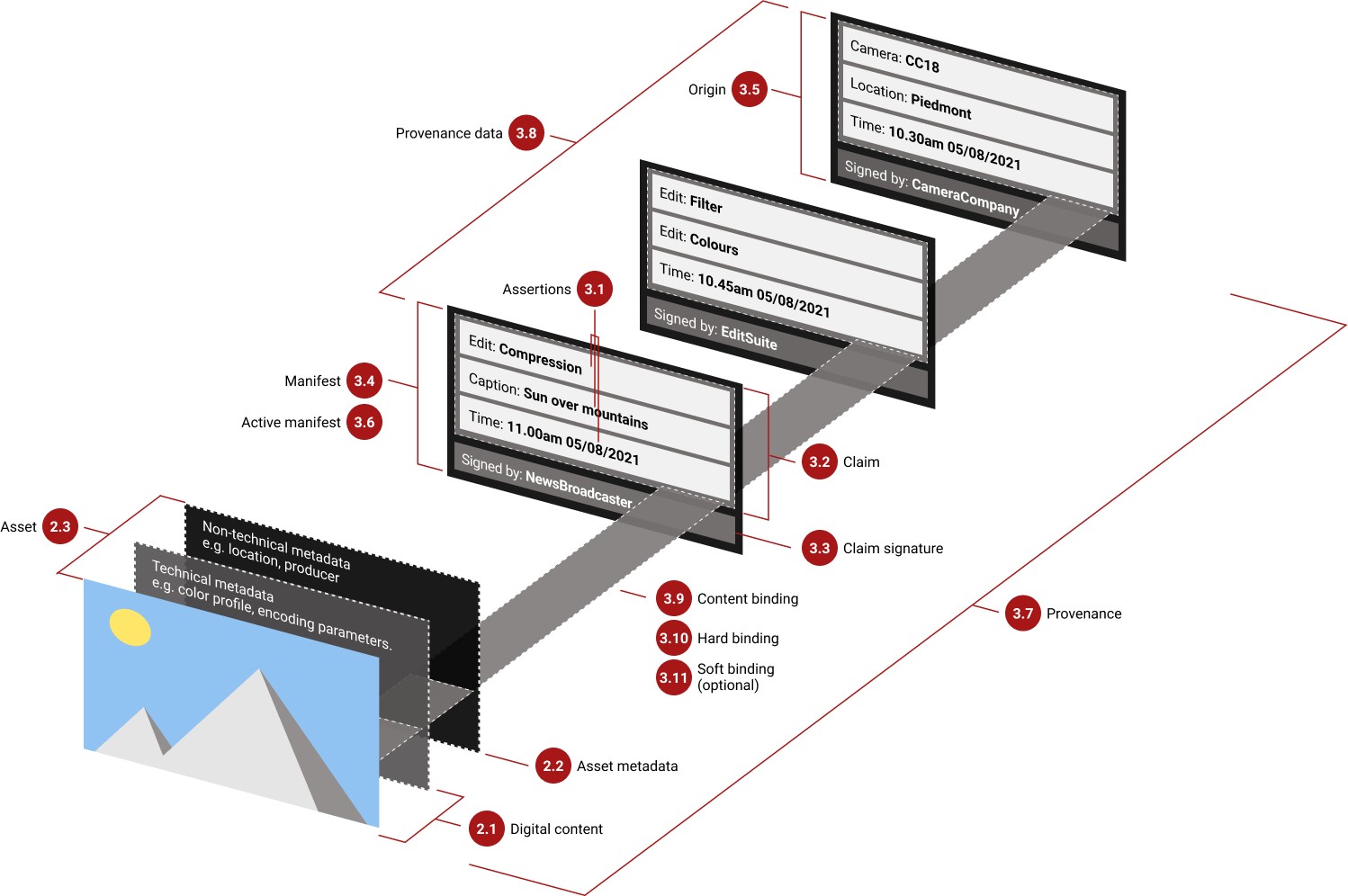
Information incorporated into the *digital content* (perceptibly or imperceptibly) of an *asset* which can be used, for example, to uniquely identify the *asset* or to store a reference to a *C2PA Manifest*.

## Manifest Repository

A repository into which *C2PA Manifests* and *C2PA Manifest Stores* can be placed, and which can be searched using a *content binding*.

# Overview

This image shows how all these various elements come together to represent the C2PA architecture.



*Figure 3. Elements of C2PA*

# 

# 3. Assertions

# General

It is expected that each of the actors in the system that creates or processes an asset will produce one or more assertions about when, where, and how the asset was originated or transformed. An assertion is labelled data, typically (though not required to be) in a CBOR-based structure which represents a declaration made by an actor about an asset. Some of these actors will be human and add human-generated information (e.g., copyright) while other actors are machines (software/hardware) providing the information they generated (e.g., camera type).

Certain assertions may be redacted by subsequent claims, but they cannot be modified once made as part of a claim.

# Labels

Each assertion has a label defined either by the C2PA specifications or an external entity.

Labels are string values organized into namespaces using a period (.) as a separator. The namespace component of the label can be an entity, or a reference to a well-established standard. The most common labels will be defined by the C2PA and will begin with c2pa. Entity-specific labels shall begin with the Internet domain name for the entity similar to how Java packages are defined (e.g., com.litware, net.fineartschool). Well- established standards can use the "stds." prefix when describing their namespace. They are also versioned with a simple incrementing integer scheme (e.g., c2pa.actions.v2). If no version is provided, it is considered as v1.

# Embedded vs Externally-Stored Data

Some assertion data, due to its size or an infrequent need for it, may be externally hosted. Such data are not embedded in the assertion store, but instead are referenced by URI. This is accomplished through a cloud data assertion. Unlike embedded assertion data, cloud data is not retrieved nor validated as part of manifest validation, and are only retrieved and validated when specifically needed by an application

# Redaction of Assertions

Assertions that are present in an asset-embedded manifest may be removed from that asset’s manifest when the asset is [used as an ingredient](#_bookmark60). This process is called redaction.

Redaction involves removing either the entire assertion from the manifest’s assertion store or retaining the labelled assertion container but replacing its data with zeros (binary \0 values). In addition, a record that something was removed must be added to the [claim](#_bookmark47) in the form of a [URI reference](#_bookmark35) to the redaction assertion in the redacted\_assertions field of the claim.

Unless the redaction of the assertion also requires modification to the digital content, an [update manifest](#_bookmark63) shall be used to document the redaction as it makes a statement about the non-changes to the content. Claims generators shall not redact assertions with a label of c2pa.actions as this assertion type represents essential information in understanding the history of an asset.

# 4. Binding to Content

# Overview

A key aspect to the [standard C2PA manifest](#_bookmark62) is the presence of one or more data structures, called content bindings, that can uniquely identify portions of the asset. There are two types of bindings that are supported by C2PA - hard bindings and soft bindings. A hard binding (also known as a cryptographic binding) enables the validator to ensure that (a) this manifest belongs with this asset and (b) that the asset has not been modified, by determining values that can match only this asset and no other, not even other assets derived from it or renditions produced from it. A soft binding is computed from the digital content of an asset, rather than its raw bits. A soft binding is useful for identifying derived assets and asset renditions.

# Hard Bindings

## Hashing using byte ranges

The simplest type of hard binding that can be used to detect tampering is a cryptographic hashing algorithm,, over some or all of the bytes of an asset. This approach can be used on any type of asset.

When using this form of hard binding, one or more [data hash assertions](#_bookmark117) is used to define the range of bytes that are hashed (and those that are not). Because each data hash assertion defines a byte range and optional URL, it is flexible enough to be usable whether the asset is a single binary or represented in multiple chunks or portions, local or remote.

# Soft Bindings

Soft bindings are described using [soft binding assertions](#_bookmark15) such as via a perceptual hash computed from the digital content or a watermark embedded within the digital content. These soft bindings enable digital content to be matched even if the underlying bits differ, for example due to an asset rendition in a different resolution or encoding format. Additionally, should a C2PA manifest be removed from an asset, but a copy of that manifest remains in a provenance store elsewhere, the manifest and asset may be matched using available soft bindings.

Because they serve a different purpose, a soft binding shall not be used as a hard binding.

All soft bindings shall be generated using one of the algorithms listed as supported by this specification.

# Chapter 5. Claims

# Overview

A **claim** gathers together all the assertions about an asset from an actor at a given time including the set of assertions for [binding to the content](#_bookmark42). The claim is then cryptographically hashed and signed. A claim has all the same properties as an assertion including being assigned the label (c2pa.claim) and supporting the use of [assertion metadata](#_bookmark114).

## Claim Generator

The claim\_generator field is required and its value is a string that conforms to the User-Agent string format specified in section 5.5.3 of [HTTP/1.1 Semantics and Content](http://tools.ietf.org/html/rfc7231#section-5.5.3) to inform a manifest consumer about what software/hardware/system produced this Claim. Since the User-Agent string uses spaces to separate tokens, it is recommended to use an \_ (underscore, U+005F) to combine words inside a single tokenA Manifest Consumer shall use the value of claim\_generator\_info in determining information about the claim generator for itself or for presentation in a UX. If only a claim\_generator field is present, or both claim\_generator and claim\_generator\_hints are present, the Manifest Consumer shall consider this a 1.0- compliant claim. In that case, it shall use the value of claim\_generator, as is, for presentation in a UX.

# 

# 6. Manifests

# Types of Manifests

## Commonalities

All C2PA Manifests shall contain an [assertion store](#_bookmark29) with at least one [assertion](#_bookmark24), a [claim](#_bookmark47) and a [claim signature](#_bookmark52). It may also contain a [credential store](#_bookmark38).

## Standard Manifests

A standard C2PA Manifest (UUID: 0x63326D61-0011-0010-8000-00AA00389B71 (c2ma)) shall contain at least one [hard binding to content](#_bookmark42) assertion - either a c2pa.hash.data or a c2pa.hash.bmff based on the type of asset for which the manifest is destined. Because of this requirement, they are the predominant type of manifest that will be present in C2PA provenance data.

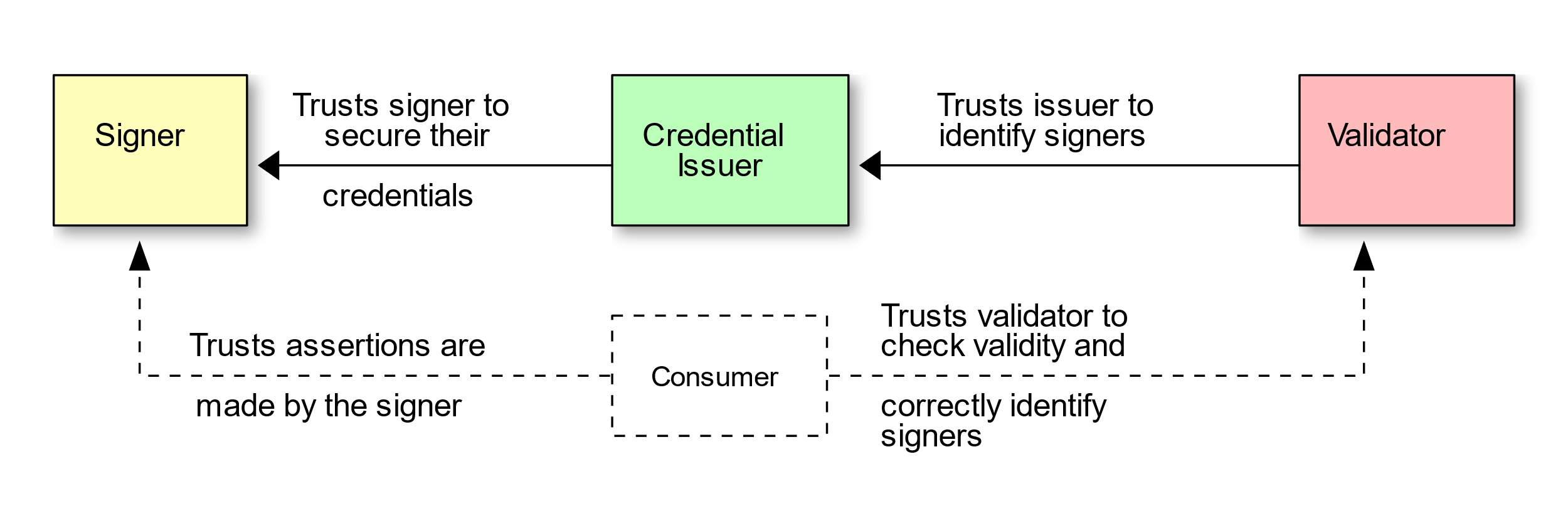
# External Manifests

In some cases, it may not be possible (or practical) to embed a C2PA Manifest Store in an asset. In those cases, keeping the C2PA Manifests externally to the asset is an acceptable model for providing providence to assets. The manifest should be stored in a location, referred to as a manifest repository, that is easily locatable by a manifest consumer working with the asset, such as [by reference or URI](#_bookmark92). As the C2PA Manifest Store is a JUMBF box, it shall be served with the JUMBF Media Type, application/x-c2pa-manifest-store. Some common reasons to use an external manifest are:

* It may not be practical, such as when the size of the C2PA Manifest Store is larger than the asset’s digital content.
* It may not be appropriate, such as when it would modify an asset that should not be modified. NOTE: a good example of this is creating a manifest for a pre-existing asset.

# 8. Trust Model

# Overview



The above model shows, in yellow, green and red, the three entities specified in the trust model, which is concerned with trust in a signer’s identity. In dashed lines, below, is the consumer (who is not specified in the trust model), who uses the identity of the signer, along with other trust signals, to decide whether the assertions made about an asset are true.

# Identity of Signers

Identity in the trust model is the means by which a cryptographic signing key is associated with an actor for the basis of making trust decisions based on any structure (including, but not limited to, claims and manifests) signed with that key. The identity of a signatory is not necessarily a human actor, and the identity presented may be a pseudonym, completely anonymous, or pertain to a service or trusted hardware device with its own identity, including an application running inside such a service or trusted hardware.

# 10. References

* [www.c2pa.org](http://www.c2pa.org)
* <https://contentauthenticity.org/>
* [www.digimarc.com](http://www.digimarc.com)
* Content Authenticity Initiative and C2PA - IPTC Photo Metadata Conference 2021

# 11. Conclusion

C2PA emerged from a collective effort to tackle the growing problem of disinformation online. By establishing an open, royalty-free technical standard for content provenance, the coalition aims to equip creators and consumers with the tools to discern truth from falsehood. This standard allows creators to embed metadata within their content, detailing its origin, modifications, and journey through the digital realm. The ramifications of C2PA extend far beyond a simple technical solution. This technology empowers individuals to become discerning consumers of information.

By having access to a content's history, users can make informed decisions about its credibility. This fosters a more responsible online environment where accountability and transparency take center stage. However, C2PA's journey is far from complete. As with any new technology, challenges and uncertainties lie ahead. Ensuring widespread adoption of the standard across platforms and industries remains a crucial task. Additionally, robust verification mechanisms need to be established to ensure the authenticity of the embedded metadata.

Despite these challenges, C2PA presents a promising path towards a more trustworthy digital landscape. Its potential to combat the spread of misinformation and promote responsible content creation is undeniable. As stakeholders collaborate and refine the C2PA standard, we can anticipate a future where information's journey is transparently documented, empowering users to navigate the digital world with greater confidence. In conclusion, C2PA represents a significant step forward in the fight for online authenticity. While its path will undoubtedly have its hurdles, the potential benefits of a verifiable and traceable digital space are immense.