



Coalition for Content Provenance and Authenticity

C2PA Technical Specification

PUBLIC DRAFT, 2021-12-09: Draft Specification (0.8)

Table of Contents

1. Introduction	2
1.1. Overview	2
1.2. Scope	2
1.3. Technical Overview	3
1.4. Establishing Trust	4
1.5. Example	4
1.6. Design Goals	5
2. Glossary	7
2.1. Introductory terms	7
2.2. Assets and Content	7
2.3. Core Aspects of C2PA	8
2.4. Additional Terms	10
2.5. Overview	10
3. Normative References	12
3.1. Core Formats	12
3.2. Schemas	12
3.3. Digital & Electronic Signatures	12
3.4. Other	13
4. Standard Terms	14
5. Assertions	15
5.1. General	15
5.2. Labels	15
5.3. Versioning	16
5.4. Multiple Instances	16
5.5. Assertion Store	16
5.6. Embedded vs Externally-Stored Data	16
5.7. Redaction of Assertions	17
6. Unique Identifiers	18
6.1. Using XMP	18
6.2. Other Identifiers	18
6.3. URI References	18
7. W3C Verifiable Credentials	21
7.1. General	21
7.2. VCStore	22

7.3. Using Credentials	23
7.4. Credential Security Considerations	23
8. Binding to Content	24
8.1. Overview	24
8.2. Hard Bindings	24
8.3. Soft Bindings	26
9. Claims	28
9.1. Overview	28
9.2. Syntax	28
9.3. Creating a Claim	30
9.4. Multiple Step Processing	34
10. Manifests	37
10.1. Use of JUMBF	37
10.2. Types of Manifests	41
10.3. Embedding manifests into assets	42
10.4. External Manifests	51
10.5. Embedding a Reference to the Active Manifest	51
11. Entity Diagram	53
12. Cryptography	54
12.1. Hashing	54
12.2. Digital Signatures	55
13. Trust Model	58
13.1. Overview	58
13.2. Identity of Signers	58
13.3. Signer Credential Trust	59
13.4. Credential Types	59
13.5. Identity In Assertions	65
13.6. Statements	65
14. Validation	66
14.1. Locating the Active Manifest	66
14.2. Locating the Claim	67
14.3. Validate the Signature	67
14.4. Validate the Time-Stamp	68
14.5. Validate the Credential Revocation Information	68
14.6. Validate the Assertions	69
14.7. Recursively Validating Integrity of Ingredients	72
14.8. Visual look of Validation	73

14.9. Validate the Asset's Content	74
15. User Experience	76
15.1. Approach	76
15.2. Principles	76
15.3. Disclosure Levels	76
15.4. Public Review, Feedback and Evolution	77
16. Information security	78
16.1. Threats and Security Considerations	78
16.2. Harms, Misuse, and Abuse	79
17. C2PA Standard Assertions	81
17.1. Introduction	81
17.2. Use of CBOR	81
17.3. Metadata About Assertions	82
17.4. Standard C2PA Assertion Summary	89
17.5. Data Hash	90
17.6. BMFF-Based Hash	92
17.7. Soft Binding	97
17.8. Cloud Data	99
17.9. Thumbnail	101
17.10. Actions	101
17.11. Ingredient	104
17.12. Depthmap	110
17.13. Exif Information	112
17.14. IPTC Photo Metadata	113
17.15. Use of Schema.org	115
17.16. Common Data Model: Actor	121
18. Open Topics	123
18.1. Assertions	123
18.2. Binding to Content	123
18.3. Trust Model	123
18.4. Validation	123
18.5. User Experience	123
19. Patent Policy	124



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Chapter 1. Introduction

1.1. 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. It is 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 requirements.

This specification has been, and continues to be, informed by scenarios, workflows and requirements gathered from industry experts and partner organizations, including the [Project Origin Alliance](#) and the [Content Authenticity Initiative \(CAI\)](#). It is also possible that regulatory bodies and governmental agencies could utilize this specification to establish standards for digital provenance.

1.2. 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](#). 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.

Prior to developing this specification, the C2PA created our [Guiding Principles](#) that enabled us to remain focused on ensuring that the specification can be used in ways that respect privacy and personal control of data with a critical

eye toward potential abuse and misuse. For example, the creators and publishers of the media assets always have control over whether provenance data is included as well as what specific pieces of data are included.

From the overarching goals section of the guiding principles:

IMPORTANT

C2PA specifications SHOULD NOT provide value judgments about whether a given set of provenance data is 'good' or 'bad,' merely whether the assertions included within can be validated as associated with the underlying asset, correctly formed, and free from tampering.

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

Other documents from the C2PA will address specific implementation considerations such as expected user experiences and details of our threat and harms modelling.

1.3. 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](#), 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](#) entity called a [Claim](#).

The [W3C Verifiable Credentials](#) of individual actors that are involved in the creation of the assertions can be added to the C2PA information to provide additional trust signals to the process of assessing trustworthiness of the asset.

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

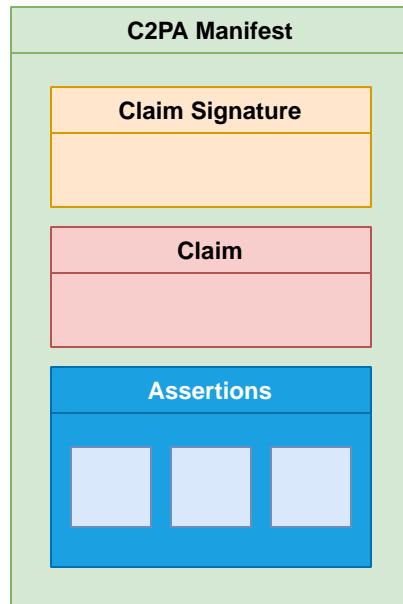


Figure 1. A C2PA Manifest and its constituent parts

1.4. Establishing Trust

The basis of making trust decisions in C2PA, our [Trust Model](#), 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.

1.5. Example

A very common scenario will be a user (called an actor in the C2PA ecosystem) taking a photograph with their C2PA-enabled camera (or phone). In that instance, the camera would create a C2PA manifest containing some such assertions including information about the camera itself, a thumbnail of the image and some cryptographic hashes that bind the photograph to the manifest. These assertions would then be listed in the Claim, which would be digitally signed and then the entire manifest would be embedded into the output JPEG. This manifest would remain valid indefinitely.

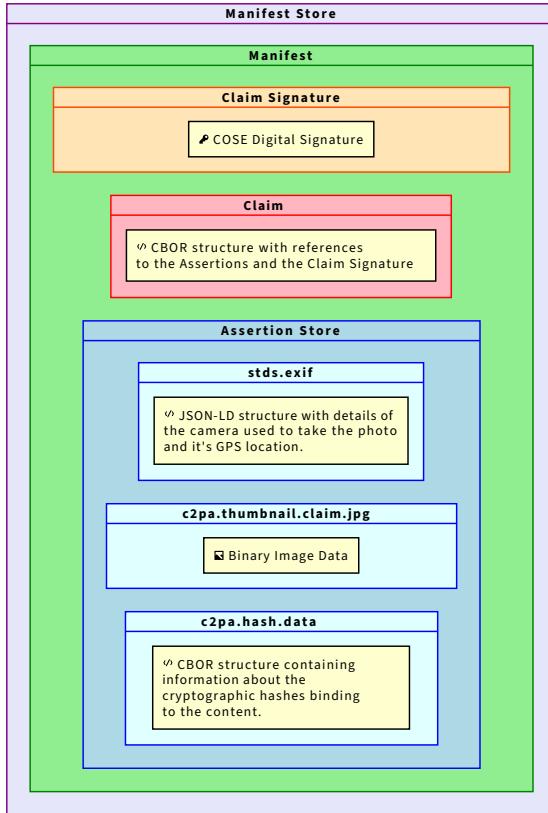


Figure 2. Example C2PA Manifest of a Photograph

A manifest consumer, such as a C2PA validator, could help users to establish the trustworthiness of the asset by first validating the digital signature and its associated credential. It can also check each of the assertions for validity and present the information contained in them, and the signature, to the user in a way that they can then make an informed decision about the trustworthiness of the digital content.

1.6. Design Goals

In the creation of the C2PA architecture, it was important to establish some clear goals for the work to ensure that the technology was usable across a wide spectrum of hardware and software implementations worldwide and accessible to all.

Some of those goals were:

- Maintain the provenance of the asset across multiple tools, from creation through all subsequent modification and publication/distribution.
- Support all standard asset formats supported by common authoring tools, across media types such as images, videos, audio, and documents.
- Create only the minimum required novel technology by relying on well-established techniques.
- Do not require cloud storage but allow for it.
- Allow flexibility in the nature of information stored.

- Allow for information to be subsequently redacted, provided that the author permits it.

Chapter 2. Glossary

2.1. Introductory terms

2.1.1. 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.

NOTE An organization or group of *actors* may also be considered an *actor* in the C2PA ecosystem.

2.1.2. 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.

2.1.3. 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*.

2.1.4. Manifest consumer

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

2.1.5. Validator

A *manifest consumer* whose role is to perform the actions described in [Chapter 14, Validation](#).

2.1.6. Action

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

2.2. Assets and Content

2.2.1. 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).

2.2.2. 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.

2.2.3. Asset

A file or stream of data containing *digital content*, *asset metadata* and optionally, a C2PA *manifest*.

NOTE For the purposes of this definition, we will extend the typical definition of "file" to include cloud-native and dynamically generated data.

2.2.4. 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*.

EXAMPLE: An audio stream that has been shortened or a document where pages have been added.

2.2.5. 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.

EXAMPLE: A video file that is re-encoded for reduced screen resolution or network bandwidth.

2.2.6. Composed asset

A composed asset is an *asset* that is created by building up a collection of multiple parts or fragments of *digital content* (referred to as ingredients) from one or more other *assets*. When starting from an existing *asset*, it is a special case of a *derived asset* - however a *composed asset* can also be one that starts from a "blank slate".

EXAMPLE: A video created by importing existing video clips and audio segments into a "blank slate". An image where another image is imported and super-imposed on top of the starting image.

2.3. Core Aspects of C2PA

2.3.1. Assertion

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

2.3.2. 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 *manifest*.

2.3.3. Claim signature

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

2.3.4. 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 manifest can either be embedded into an *asset* or be external to its *asset*.

NOTE A manifest can reference other manifests.

2.3.5. Origin

The *manifest* in the *provenance data* which represents the method or device that initially created the *asset*.

NOTE Details on how one determines which *manifest* is the *origin* are left for specification.

2.3.6. Active Manifest

The last manifest in the list of manifests which is the one with the set of *content bindings* that are able to be validated.

2.3.7. 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*.

2.3.8. Provenance data

The set of *manifests* for an *asset* and, in the case of a *composed asset*, its ingredients.

2.3.9. 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.

2.3.10. Content binding

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

2.3.11. Hard binding

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

2.3.12. 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 content.

2.3.13. 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.

2.4. Additional Terms

2.4.1. Fingerprint

A set of inherent properties computable from *digital content* that identifies the content or near duplicates of it.

EXAMPLE: An *asset* can become separated from its *manifest* due to removal or corruption of *asset* metadata. A *fingerprint* of the *digital content* of the *asset* could be used to search a database to recover the *asset* with an intact *manifest*.

2.4.2. 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 *manifest*.

2.4.3. Manifest Repository

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

2.5. Overview

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

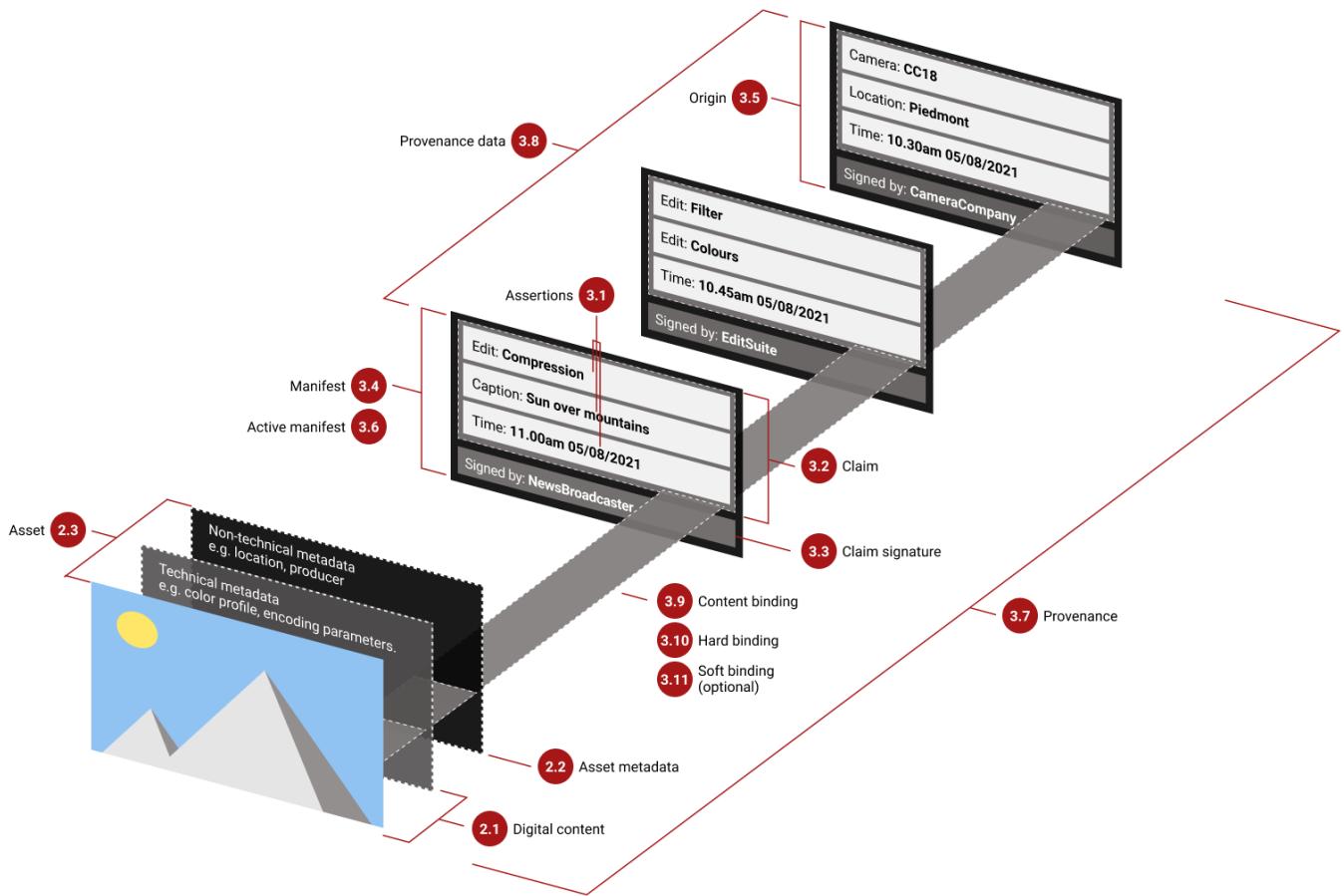


Figure 3. Elements of C2PA

Chapter 3. Normative References

3.1. Core Formats

- [CBOR](#)
- [JSON](#)
- [JSON-LD](#)
- [JPEG universal metadata box format \(JUMBF\)](#)
- [ISO Base Media File Format \(BMFF\)](#)

3.2. Schemas

- [CDDL](#)
- [JSON Schema](#)
- [Dublin Core Metadata Initiative](#)

3.3. Digital & Electronic Signatures

- [X.509 Certificates](#)
- [JSON Web Algorithms \(JWA\)](#)
- [CBOR Object Signing and Encryption \(COSE\)](#)
- [Using RSA Algorithms with COSE Messages](#)
- [Online Certificate Status Protocol \(OCSP\)](#)
- [Internet X.509 PKI Time-Stamp Protocol](#)
- [CBOR Object Signing and Encryption \(COSE\): Header parameters for carrying and referencing X.509 certificates](#)
- [Algorithms and Identifiers for the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List \(CRL\) Profile](#)
- [Internet X.509 Public Key Infrastructure: Additional Algorithms and Identifiers for DSA and ECDSA](#)
- [Algorithm Identifiers for Ed25519, Ed448, X25519, and X448 for Use in the Internet X.509 Public Key Infrastructure](#)
- [PKCS #1: RSA Cryptography Specifications Version 2.2](#)
- [Edwards-Curve Digital Signature Algorithm \(EdDSA\)](#)
- [JSON Advanced Electronic Signatures \(JAdES\)](#)
- [US Secure Hash Algorithms](#)

3.4. Other

- [eXtensible Metadata Platform \(XMP\)](#)
- [JSON-LD serialization of XMP](#)
- [IPTC Photo Metadata Standard](#)
- [EXIF](#)
- [UUID](#)

Chapter 4. Standard Terms

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [BCP 14](#), [RFC 2119](#), and [RFC 8174](#) when they appear in any casing (upper, lower or mixed).

Chapter 5. Assertions

5.1. 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).

Some examples of assertions are:

- Exif information (e.g. camera information such as maker, lens)
- Actions performed on the asset (e.g., clipping, color correction)
- Thumbnail of the asset or its ingredients
- Content bindings (e.g., cryptographic hashes)

Certain assertions may be redacted by subsequent claims (see [Section 5.7, “Redaction of Assertions”](#)), but they cannot be modified once made as part of a claim.

5.2. 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 (see ABNF below). 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`. The list of publicly known labels can be found in [Chapter 17, C2PA Standard Assertions](#).

```
namespaced-label = qualified-namespace label
qualified-namespace = entity / ( "stds." std-name )
entity = 1*( DIGIT / ALPHA / "-" )
std-name = 1*( DIGIT / ALPHA / "-" )
label = 1*("." 1(ALPHA / "_") *( DIGIT / ALPHA / "_" ) )
```

The period-separated components of a label follow the variable naming convention (`[a-zA-Z_][a-zA-Z0-9_]*`) specified in the POSIX or C locale, with the restriction that the use of a repeated underscore character (`__`) is reserved for labeling multiple assertions of the same type.

5.3. Versioning

When an assertion's schema is changed, it should be done in a backwards-compatible manner. This means that new fields may be added and existing ones may be marked as deprecated (i.e., can be read, but never written). Existing fields shall not be removed. The label would then consist of an incremented version number, for example moving from `c2pa.ingredient` to `c2pa.ingredient.v2`.

Deprecated fields for C2PA standard assertions shall be indicated in [Chapter 17, C2PA Standard Assertions](#). Tools which enable actors to create assertions shall prevent the actor from inserting data into deprecated assertion fields.

In addition, there are situations where a non-backwards compatible change is required. In that case, instead of increasing the label's version number, the assertion shall be given a new label. For example, `c2pa.ingredient` could be changed to the fictional `c2pa.component`.

5.4. Multiple Instances

Multiple assertions of the same type can occur in the same manifest, but since assertions are referenced by claims via their label, the assertion labels must be unique. This is accomplished by adding a double-underscore and a monotonically increasing index to the label. For example, if a manifest contains a single assertion of type `stds.schema-org.CreativeWork`, then the assertion label will be `stds.schema-org.CreativeWork`. If a manifest contains three assertions of this type, the labels will be `stds.schema-org.CreativeWork`, `stds.schema-org.CreativeWork_1` and `stds.schema-org.CreativeWork_2`.

When a label includes a version number, that version number is part of the label itself. As such, when there are multiple instances, the instance number continues to follow the label - e.g., `c2pa.ingredient.v2_2`.

5.5. Assertion Store

The set of assertions referenced by a [claim](#) in a manifest are collected together into a logical construct that is referred to as the **assertion store**. The assertions and assertion store shall be stored as described in [Section 10.1, “Use of JUMBF”](#); in particular, the assertion store shall be located in the same C2PA Manifest box as the claim that refers to its assertions.

For each manifest, there is a single assertion store associated with it. However, as an asset may have multiple manifests associated with it, each one representing a specific series of assertions, there may be multiple assertion stores associated with an asset.

5.6. 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. Unlike embedded assertions, external assertions are not retrieved nor validated as part of manifest validation, and are only retrieved and validated when specifically needed by an application. This is accomplished through a cloud data assertion (see [Section 17.8, “Cloud Data”](#)), and

the different validation rules are described in [Section 14.6, “Validate the Assertions”](#).

5.7. 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](#). 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](#) in the form of a [URI reference](#) to the redaction assertion in the [redacted_assertions](#) field of the claim.

NOTE Because each assertion’s [URI reference](#) includes the assertion label, it is also known what type of information (e.g., thumbnail, IPTC metadata, etc.) was removed. This enables both humans and machines to apply rules to determine if the removal was acceptable.

Unless the redaction of the assertion also requires modification to the digital content, an [update manifest](#) 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.

Chapter 6. Unique Identifiers

Every asset that is referenced from the [claim](#) shall be referenced via a unique identifier. In addition, these identifiers are used in various parts of a C2PA-enabled workflow, such as when identifying it as an [ingredient](#) in a derived or composed asset.

6.1. Using XMP

When an asset contains embedded XMP, that XMP shall include (at least) values for [xmpMM:DocumentID](#) and [xmpMM:InstanceID](#) as defined in [XMP Specification Part 2, 2.2](#). If an asset does not contain XMP at the time a claim is made, and the type of the asset supports it, an embedded XMP packet may be created as part of the process, and the identifiers shall be added to it.

NOTE 1

- NOTE** Some asset types are not suited for embedded XMP (e.g., text). It is possible to create XMP as a sidecar.

6.2. Other Identifiers

Instead of using XMP, a unique identifier for an asset could be a URI defined by standards such as [Decentralized Identifiers \(DID\)](#), [Handle](#), [EIDR](#) and [DOI](#).

Another standard unique identifier for an asset could be the cryptographic hash of the asset. When this method is used, the hash shall be represented using a standard [RFC 4122 UUID](#) following the recommendations at <https://datatracker.ietf.org/doc/html/draft-thiemann-hash-urn-01>.

EDITORS NOTE

- NOTE** Other methods may be defined here as they are developed.

6.3. URI References

Assertions and claims, whether they are stored internally to the asset (ie. embedded) or stored externally to the asset (e.g., in the cloud), shall be referenced via JUMBF URI references as defined in [ISO 19566-5, C.2](#). This URI shall be used as part of a [hashed_uri](#) data structure:

```
{
  "$schema": "http://json-schema.org/draft-07/schema",
  "$id": "http://ns.c2pa.org/hashed-uri/v1",
  "type": "object",
  "description": "The data structure used to store a reference to a local URL and its hash",
  "definitions": {
    "JUMBF_URI": {
      "$id": "#JUMBF_URI",
      "description": "JUMBF URI reference",
      "type": "string",
      "pattern": "^self#jumbf=[\\w\\d][\\w\\d\\.:/-]+[\\w\\d]$"
    }
  },
  "examples": [
    {
      "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-329BF39FA1E4/c2pa.assertions/c2pa.actions",
      "alg": "sha256",
      "hash": "hoOspQQ1lFTy/4Tp8Epx670E5QW5NwkNR+2b30KFXug="
    }
  ],
  "required": ["url", "hash"],
  "properties": {
    "url": {
      "$ref": "#/definitions/JUMBF_URI",
      "description": "JUMBF URI reference"
    },
    "alg": {
      "type": "string",
      "minLength": 1,
      "description": "A string identifying the cryptographic hash algorithm used to compute all hashes in this claim, taken from the C2PA hash algorithm identifier list. If this field is absent, the hash algorithm is taken from an enclosing structure as defined by that structure. If both are present, the field in this structure is used. If no value is present in any of these places, this structure is invalid; there is no default."
    },
    "hash": {
      "type": "string",
      "minLength": 1,
      "description": "CBOR byte string containing the hash value"
    }
  },
  "additionalProperties": false
}
```

This specification provides an equivalent [hashed-uri-map](#) data structure for schemas defined using [CDDL](#):

```

; The data structure used to store a reference to a URL within the same JUMBF and its hash.
We use a socket/plug here to allow hashed-uri-map to be used in individual files without
having the map defined in the same file
$hashed-uri-map /= {
  "url": url-regexp-type, ; JUMBF URI reference
  ? "alg": tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash
algorithm used to compute all hashes in this claim, taken from the C2PA hash algorithm
identifier list. If this field is absent, the hash algorithm is taken from an enclosing
structure as defined by that structure. If both are present, the field in this structure is
used. If no value is present in any of these places, this structure is invalid; there is no
default.
  "hash": bstr, ; byte string containing the hash value
}

; with CBOR Head (#) and tail ($) are introduced in regexp, so not needed explicitly
url-regexp-type /= tstr .regexp "self#jumbf=[\\w\\d][\\w\\d\\.\\/:-]+[\\w\\d]"

```

Because assertion stores must be located in the same C2PA Manifest box as the claim that refers to them, only **self#jumbf** URLs are permitted. These **self#jumbf** URLs may be relative to the entire C2PA Manifest Store, in which case they shall start with a / (U+002F, Slash), or relative to the current manifest. URLs shall not contain the sequence .. (a pair of U+002E, Full Stop).

NOTE

Cloud data assertions use a different schema for URLs which is described in [Section 17.8, “Cloud Data”](#).

Examples

- **self#jumbf=/c2pa/urn:uuid:f095f30e-6cd5-4bf7-8c44-ce8420ca9fb7/c2pa.assertions/c2pa.thumbnail.claim.jpeg** is relative to the entire store (since it starts with /),
- **self#jumbf=c2pa.assertions/c2pa.thumbnail.claim.jpeg** would be relative to the manifest of the box containing the URI.

The hash is performed over the canonical serialization of the assertion data, omitting the surrounding container. For example, if the assertion type is **c2pa.actions**, then the hash is calculated over the canonical serialization of **actions-map**.

Chapter 7. W3C Verifiable Credentials

7.1. General

In some use cases, the actors in the system may wish to provide their own [W3C Verifiable Credential](#), as they exist at that moment in time, to the claim generator to have them associated with one or more assertions. These actors may be individuals, groups or organizations.

W3C Verifiable Credentials are used in this specification to *decorate* the actors identified in assertions with more information, potentially providing additional trust signals. Although these W3C Verifiable Credentials can include proofs of their own authenticity, they are **not** a mechanism for verifying that a particular actor authorised a claim, assertion or piece of metadata. Any validation or usage of the W3C Verifiable Credential is out of scope of this specification and has no bearing on the [C2PA Trust Model](#).

For example, conveying a W3C Verifiable Credential for the actor identified as the **author** in an assertion might link that author's ID with an email address, social media ID, or real name, or it might identify that actor as a member of a particular professional body, or provide other qualifications relevant to the actor's involvement in the asset.

Such credentials shall be compliant with the [W3C Verifiable Credentials Data Model](#) using the [JSON-LD serialisation](#) described there.

NOTE JSON-LD serialization is mandated as it is the most commonly used of the three syntaxes presented in section 6 of the W3C Verifiable Credentials specification. It is also the one that aligns best with its extensibility model, which could be useful to some implementers.

An example of a compliant credential for an individual might be one issued by the National Press Photographers Association (NPPA), which links an identifier for a person to their name ("John Doe") and a statement about their membership of the NPPA. It might look like:

```
{
  "@context": [
    "https://www.w3.org/2018/credentials/v1",
    "http://schema.org"
  ],
  "type": [
    "VerifiableCredential",
    "NPPACredential"
  ],
  "issuer": "https://nppa.org/",
  "credentialSubject": {
    "id": "did:nppa:eb1bb9934d9896a374c384521410c7f14",
    "name": "John Doe",
    "memberOf": "https://nppa.org/"
  },
  "proof": {
    "type": "RsaSignature2018",
    "created": "2021-06-18T21:19:10Z",
    "proofPurpose": "assertionMethod",
    "verificationMethod": "did:nppa:eb1bb9934d9896a374c384521410c7f14#_Qq0UL2Fq651Q0Fjd6TvnYE-faHiOpRlPVQcY_-tA4A",
    "jws": "eyJhbGciOiJQUzI1NiIsImI2NCI6ZmFsc2UsImNyaxQi0lsiyjY0Il19DJB...ogZdgt1DkQxDfxn41QWDw_mmMCjs9qxg0zcZzqEJw"
  }
}
```

A W3C Verifiable Credential used with C2PA shall contain only a single `credentialSubject` and that `credentialSubject` shall have an `id` value.

NOTE

Although the example above and many examples in the W3C Verifiable Credentials data model specification use Decentralized Identifiers (DIDs) as the value of the `id` field, DIDs are not necessary for W3C Verifiable Credentials to be useful. Specifically, W3C Verifiable Credentials do not depend on DIDs and DIDs do not depend on W3C Verifiable Credentials. DID-based URLs are just one way to express identifiers associated with subjects, issuers, holders, credential status lists, cryptographic keys, and other machine-readable information associated with a W3C Verifiable Credential.

7.2. VCStore

The set of credentials in a manifest are collected together into a logical construct that is referred to as the `credential store` or `VCStore` (for short) and it shall be stored as described in [Section 10.1, “Use of JUMBF”](#). Unlike the assertion store, the VCStore shall always be included in the JUMBF - it shall not be stored separately.

For each manifest, there is a VCStore associated with it. However, as an asset may have multiple manifests associated with it, there may be multiple VCStores associated with an asset.

7.3. Using Credentials

Some assertions, such as [Creative Work](#) and [Actions](#), may contain references to Persons or Organisations which are responsible for various roles and responsibilities. These references to Actors are defined in [Section 17.16, “Common Data Model: Actor”](#).

```
{  
  "@context": "http://schema.org/",  
  "@type": "CreativeWork",  
  "copyrightHolder": {  
    "name": "BBC",  
    "legalName": "British Broadcasting Corporation",  
    "identifier": "https://www.bbc.co.uk/",  
    "credential": [  
      {  
        "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-  
329BF39FA1E4/c2pa.credentials/https://www.bbc.co.uk/",  
        "alg": "sha256",  
        "hash": "Auxjtmax46cC2N3Y9aFmB09Jfay8LEwJWzBUTZ0sUM8gA"  
      }  
    ]  
  },  
  "copyrightYear": 2021,  
  "copyrightNotice": "Copyright © 2021 BBC."  
}
```

7.4. Credential Security Considerations

In most W3C Verifiable Credential workflows, the information about the subject (e.g., the cryptographic keys) is fetched on demand at the time of validation. While that is an acceptable model, it does open up a possible attack vector by providing an attacker with an externally-visible signal about what the validator is validating. Therefore, C2PA also supports having the information captured and embedded at the time of signature. This not only prevents leakage, but also makes it very clear what data the signer is asserting about the credential’s subject.

Chapter 8. Binding to Content

8.1. Overview

A key aspect to the [standard C2PA manifest](#) 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.

8.2. Hard Bindings

8.2.1. Hashing using byte ranges

The simplest type of hard binding that can be used to detect tampering is a cryptographic hashing algorithm, as described in [Section 12.1, “Hashing”](#), 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](#) 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.

8.2.2. Hashing a BMFF-formatted asset

If the asset is based on [ISO BMFF](#) then a hard binding optimized for the box-based format (called [BMFF-based hash assertions](#)) may be used instead.

For a monolithic mp4 file asset where the `mdat` box is validated as a unit, the assertion is validated nearly identically to a data hash assertion. It simply uses a box exclusion list instead of byte ranges to define the range of bytes that are hashed (and those that are not).

For a monolithic mp4 file asset where the `mdat` box is validated piecemeal or an asset composed of fragmented mp4 files, the assertion itself must be combined with chunk-specific hashing information which is located as specified in [Section 10.3.2, “Embedding manifests into BMFF-based assets”](#). Validating a given chunk requires first validating the `merkle` field’s `initHash` over the corresponding initialization segment and then locating the correct entry in the `merkle` field’s `hashes` array and validating it against the hash of the chunk’s data plus (if needed) deriving the hash using the other `hashes` specified in the chunk’s C2PA-specific box.

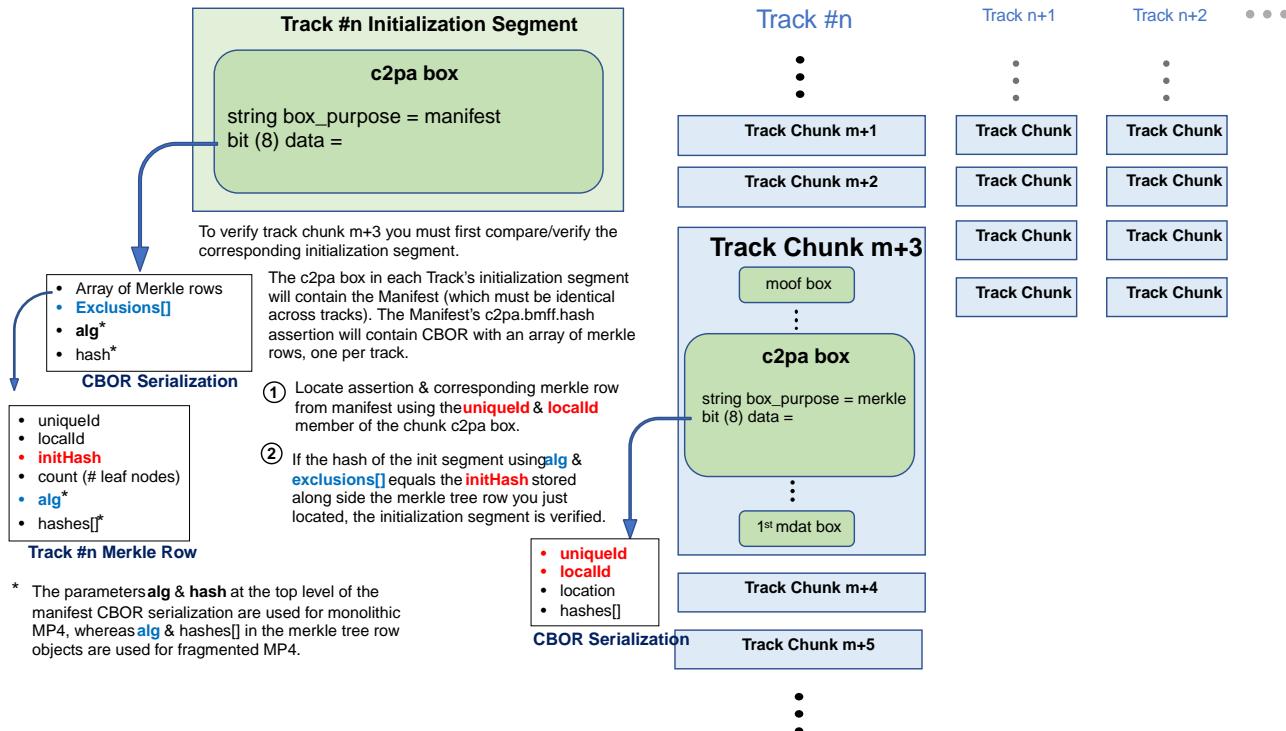


Figure 4. Validating the initialization segment

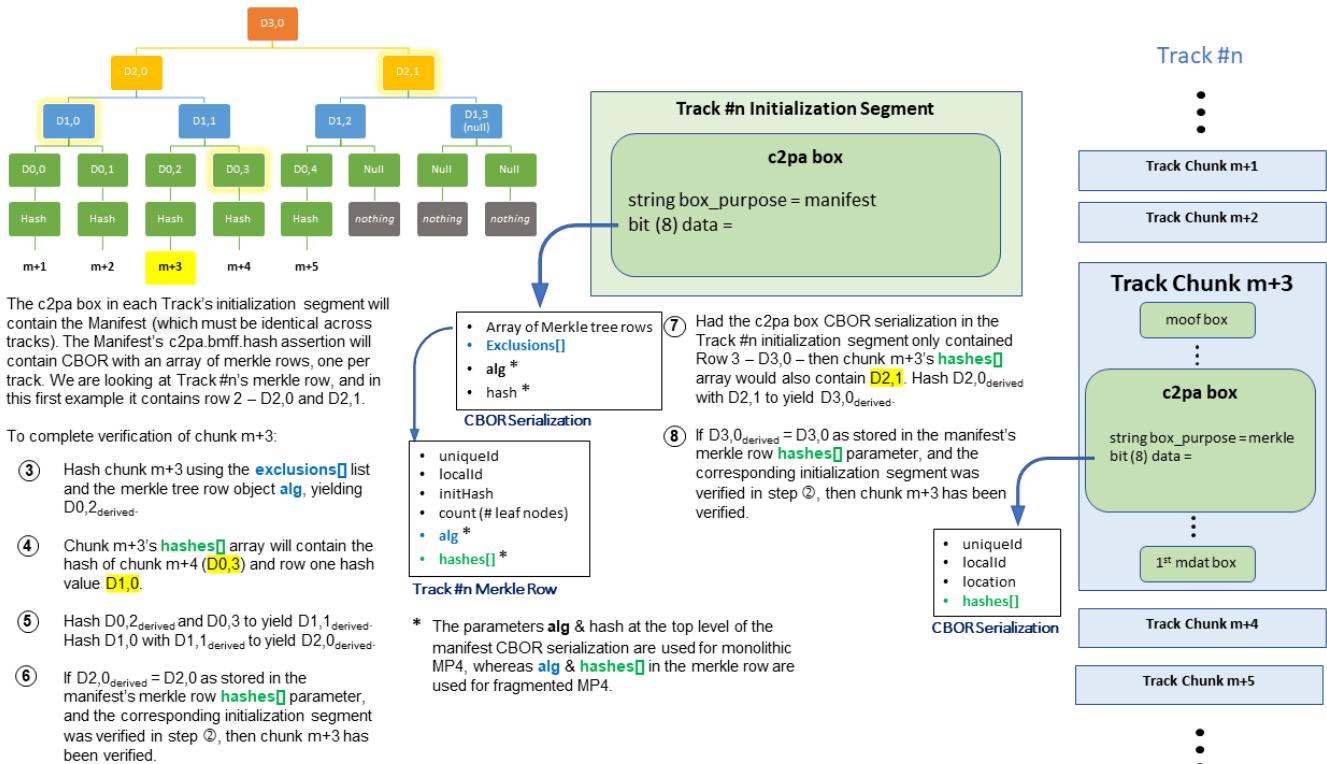


Figure 5. Validating the chunk's data

8.2.3. Asset Metadata Bindings

In those workflows which embed XMP or other forms of asset metadata into the asset, the asset's asset metadata

should not be excluded by [data hash assertions](#).

This means that by default all asset metadata (including Exif metadata and IPTC metadata in either IPTC-IIM or XMP format) will be included in the [data hash assertions](#), but with no provenance information such as **who** made the claims.

To explicitly assert the same claims in a C2PA assertion with verifiable provenance, the Exif or IPTC fields should be copied to a `stds.exif` or `stds.iptc.photo-metadata` assertion, as appropriate (see [Section 17.13, “Exif Information”](#) and [Section 17.14, “IPTC Photo Metadata”](#)).

NOTE

We recommend that existing Exif, IPTC-IIM and/or XMP asset metadata be left untouched in the asset. This will allow for compatibility with tools which do not yet support C2PA metadata.

8.3. Soft Bindings

Soft bindings are described using [soft binding assertions](#) 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. This section defines both:

- A list of algorithms that are allowed for generating soft bindings of new content as well as required for validating or locating existing content (the allowed list), and
- A list of algorithms that are required to be supported for validating or locating existing content but are not allowed for generating soft bindings of new content (the deprecated list).

This list of allowed algorithms will define the string algorithm identifier to be used as the algorithm identifier in the corresponding field and the content types over which it is applicable. In cases where there are different versions of an algorithm, each will be defined using different string algorithm identifiers. Any technical documentation sufficient for the soft binding algorithm to be uniquely identified and utilized, should be referenced.

Each algorithm should be defined along with the names and values of all parameters affecting the operation of that algorithm. When doing so, it shall describe the manner in which those parameters must be encoded within the `alg-params` field of the [soft binding assertion](#). An algorithm that is instantiated over a different parameter set will be considered a different algorithm.

Each algorithm may also define an encoding scheme for specifying the portion of digital content over which a soft binding is computed (namely, the `extent` field of the `scope` object within the [soft binding assertion](#)). An algorithm that encodes the `extent` differently will be considered a different algorithm.

It is recommended that the string identifiers for soft binding algorithms conform to how they are referred to in common practice.

There are no soft binding algorithms defined in the approved list nor in the deprecated list in this version of the specification.

NOTE

The C2PA is currently evaluating various soft binding algorithms. One of the many possible options includes the [ISCC - International Standard Content Code](#). The ISCC is an identifier and fingerprint for digital assets that supports all major content types (e.g., text, image, audio, video). The ISCC uses similarity-preserving hashes generated both from metadata and content.

Chapter 9. Claims

9.1. 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](#). The claim is then cryptographically hashed and signed as described in [Section 9.3.2.4, “Signing a Claim”](#). A claim has all the same properties as an assertion including being assigned a label (`c2pa.claim`).

9.2. Syntax

The [CDDL Definition](#) for this type is:

```
; CDDL schema for a claim map in C2PA
claim-map = {
  "claim_generator": tstr, ; A User-Agent string formatted as per
  http://tools.ietf.org/html/rfc7231#section-5.5.3, for including the name and version of the
  claims generator that created the claim
  "signature": jumbf-uri-type, ; JUMBF URI reference to the signature of this claim
  "assertions": [1* $hashed-uri-map],
  "dc:format": tstr, ; media type of the asset
  "instanceID": tstr .size (1..max-tstr-length), ; uniquely identifies a specific version of
  an asset
  ? "dc:title": tstr .size (1..max-tstr-length), ; name of the asset,
  ? "redacted_assertions": [1* jumbf-uri-type], ; List of hashed URI references to the
  assertions of ingredient manifests being redacted
  ? "alg": tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash
  algorithm used to compute all data hash assertions listed in this claim unless otherwise
  overridden, taken from the C2PA data hash algorithm identifier registry. This provides the
  value for the 'alg' field in data-hash and hashed-uri structures contained in this claim
  ? "alg_soft": tstr .size (1..max-tstr-length), ; A string identifying the algorithm used
  to compute all soft binding assertions listed in this claim unless otherwise overridden,
  taken from the C2PA soft binding algorithm identifier registry."
  ? "claim_generator_hints": generator-hints-map,
}

jumbf-uri-type = tstr .regexp "self#jumbf=[\\w\\d][\\w\\d\\.\\:/-]+[\\w\\d]"

generator-hints-map = {
  ? "Sec-CH-UA": [1* tstr], ; A human readable string naming the claim_generator
  * tstr => any
}
```

An example in [CBOR-Diag](#) is shown below:

```
{
  "claim_generator": "Joe's Photo Editor/2.0 (Windows 10)",
  "signature" : "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-
329BF39FA1E4/c2pa.signature",
  "alg" : "sha256",
  "dc:format": "image/jpeg",
  "assertions" : [
    {
      "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-
329BF39FA1E4/c2pa.assertions/c2pa.hash.data",
      "hash": b64'U9Gyz05tmpftkoEYP6XYNsMnUbns/KcktAg2vv7n1n8='
    },
    {
      "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-
329BF39FA1E4/c2pa.assertions/c2pa.thumbnail.claim.jpeg",
      "hash": b64'G5hfJwYeWTlflx0hmfcO9xDAK52aKQ+YbKNhRZe92c='
    },
    {
      "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-
329BF39FA1E4/c2pa.assertions/c2pa.claim.ingredient",
      "hash": b64'Yzag4o5j04xPyfANVtw7ETlbFSWZNfeM78qbSi8Abkk='
    }
  ],
  "redacted_assertions" : [
    "self#jumbf=c2pa/urn:uuid:5E7B01FC-4932-4BAB-AB32-
D4F12A8AA322/c2pa.assertions/stds.exif"
  ]
}
```

The [Media Type](#) of the ingredient shall be declared in `dc:format`. If present, the value of `dc:title` shall be a human-readable name for the asset.

If the asset contains XMP, then the asset's `xmpMM:InstanceID` should be used as the `instanceID`. When no XMP is available, then some other [unique identifier](#) for the asset shall be used as the value for `instanceID`.

NOTE

Some field names, such as `dc:format`, have namespace prefixes as their names and definitions are taken directly from the XMP standard. However, their usage in C2PA does not require the use of XMP.

The value of `claim_generator` is a human-readable string that will let a user know what software/hardware/system produced this Claim. This field shall be present and its value shall be a string that conforms to the User-Agent string format specified in section 5.5.3 of [HTTP/1.1 Semantics and Content](#). In addition, the claim generator may also provide information about itself using the `claim_generator_hints` object which allows using the specific fields from [W3C's proposed User Agent Client Hints specification](#).

Example using client hints

```
{  
  "claim_generator": "Joe's Photo Editor/2.0 (Windows 10)",  
  "claim_generator_hints" : {  
    "Sec-CH-UA": "Joe's Photo Editor",  
    "Sec-CH-UA-Full-Version": "2.0",  
    "Sec-CH-UA-Platform": "Windows"  
  }  
}
```

The **signature** field shall be present containing a [URI reference](#) to a [claim signature](#).

The **assertions** field shall be present containing one or more [URI references](#) to the [assertions](#) being made by this claim.

When present, the **redacted_assertions** field shall contain one or more [URI references](#) to redacted assertions.

9.3. Creating a Claim

9.3.1. Creating Assertions

Before the claim can be finalized, all [assertions](#) must be created and stored in a newly created [C2PA Assertion Store](#) as described [later in this document](#).

When creating a standard manifest, it may not be possible to know all of the required binding information at the time of claim creation, in which case use the [multiple step processing method](#) to setup and then later fill-in the information.

9.3.2. Preparing the Claim

9.3.2.1. Adding Assertions and Redactions

The claim shall contain the **assertions** field and its value is a list of all of the URI references for all assertions that were added to the assertion store that are being "claimed" by this claim. At least one of the assertions shall be either a [data hash assertion](#) or a [BMFF-based hash assertion](#).

If any assertions in ingredient claims are being redacted, their URI references shall be added to list which is the value of the **redacted_assertions** field.

9.3.2.2. Adding Ingredients

In many authoring scenarios, an actor does not create an entirely new asset but instead brings in other existing assets on which to create their work - either as a derived asset, a composed asset or an asset rendition. These existing assets are called ingredients and their use is documented in the provenance data through the use of an [ingredient assertion](#). When an ingredient contains one or more C2PA manifests, those manifests must be inserted into this asset to ensure that the provenance data is kept intact. Such ingredient manifests are added to the JUMBF as described in [Section](#)

10.1.1, “C2PA Box details”.

9.3.2.3. Connecting the Signature

The signature cannot be part of the signed payload, but since its label is pre-defined, then the full URI reference is also known. As such, we can include that in the claim by setting the value of the **signature** field of the claim to that URI reference.

NOTE This provides the explicit binding of the claim to its signature.

9.3.2.4. Signing a Claim

Producing the signature is specified in [Section 12.2, “Digital Signatures”](#). The **payload** field of **Sig_structure** shall be the serialized CBOR of the claim document. The serialized **COSE_Sign1_Tagged** structure resulting from the digital signature procedure is written into the C2PA Claim Signature box.

9.3.2.5. Time-stamps

If possible, the signer should use a RFC3161-compliant Time Stamp Authority (TSA) ([RFC 3161 section 1](#)) to obtain a trusted time-stamp proving that the signature itself actually existed at a certain date and time and incorporate that into the **COSE_Sign1_Tagged** structure as a countersignature. A manifest may contain multiple time-stamps.

NOTE Signers are encouraged to obtain and include time-stamps to ensure their manifests will remain valid. As described in [Chapter 14, Validation](#), manifests without time-stamps cease to be valid when the signing credential expires or becomes revoked.

All time-stamps shall be obtained as described in [RFC3161](#) with the following additional requirements:

- The **MessageImprint** of the **TimeStampReq** structure ([RFC 3161 section 2.4.1](#)) shall be computed by creating the **ToBeSigned** value in [RFC 8152 section 4.4](#) with the following values for elements of **Sig_structure**:
 - The **context** element shall be **CounterSignature**.
 - The **payload** element shall be as described in [Section 9.3.2.4, “Signing a Claim”](#)
 - The remaining elements of **Sig_structure** are as described in [Section 12.2, “Digital Signatures”](#).
- The **ToBeSigned** value is then hashed using a hash algorithm from the allowed list in [Section 12.1, “Hashing”](#) that the TSA supports, and that hash algorithm and value are placed in the **MessageImprint**. If the TSA does not support any hash algorithms from the allowed list, it cannot be used for time-stamping.
 - Where possible, the hash algorithm should use the same hash algorithm used in the digital signature of the claim.
- The **certReq** boolean of the **TimeStampReq** structure shall be asserted in the request to the TSA, to ensure its certificate chain is provided in the response.

Time-stamps shall be stored in a COSE unprotected header whose label is the string **sigTst**. If no time-stamps are

included, the header shall be absent. When present, the value of this header shall be a **tstContainer** defined by the following CDDL:

```
; CBOR version of tstContainer and related structures based on JSON schema at
https://forge.etsi.org/rep/esi/x19_182_JAdES/raw/v1.1.1/19182-jsonSchema.json
tstContainer = {
  "tstTokens": [1* tstToken]
}

tstToken = {
  "val": bstr
}
```

The content of the **TimeStampResp** structure received in reply from the TSA shall be stored as the value of the **val** property of an element of **tstTokens**.

NOTE The above definition is a CBOR adaptation of a subset of the schema from [JAdES section 5.3.4](#) and [its JSON schema](#), except with the modification that the content of **val** is a byte string containing the content of the **TimeStampResp**, and not a Base64-encoded version of the same.

9.3.2.6. Credential Revocation Information

If the signer's credential type supports querying online credential status, and the credential contains a pointer to a service to provide time-stamped credential status information, the signer should query the service, capture the response, and store it in the manner described for the signer's credential type in the [Trust Model](#). If credential revocation information is attached in this manner, a trusted time-stamp must also be obtained after signing, as described in [Section 9.3.2.5, “Time-stamps”](#).

9.3.3. Examples

9.3.3.1. Single Claim

Here is a visual representation of an image containing a single claim with multiple assertions that have been embedded inside it.

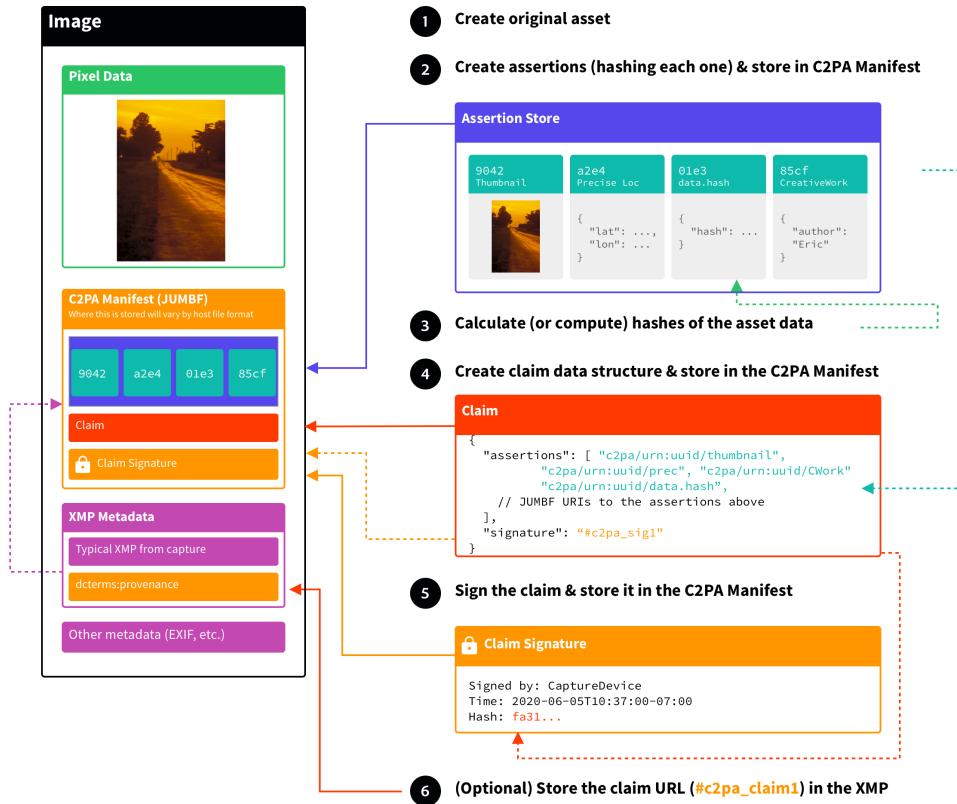


Figure 6. A single claim with assertions

9.3.3.2. Multiple Claims

In this example of creating a second claim for the [previous example](#), one of the original assertions has been redacted from the previous claim. The visual representation for this scenario would look like:

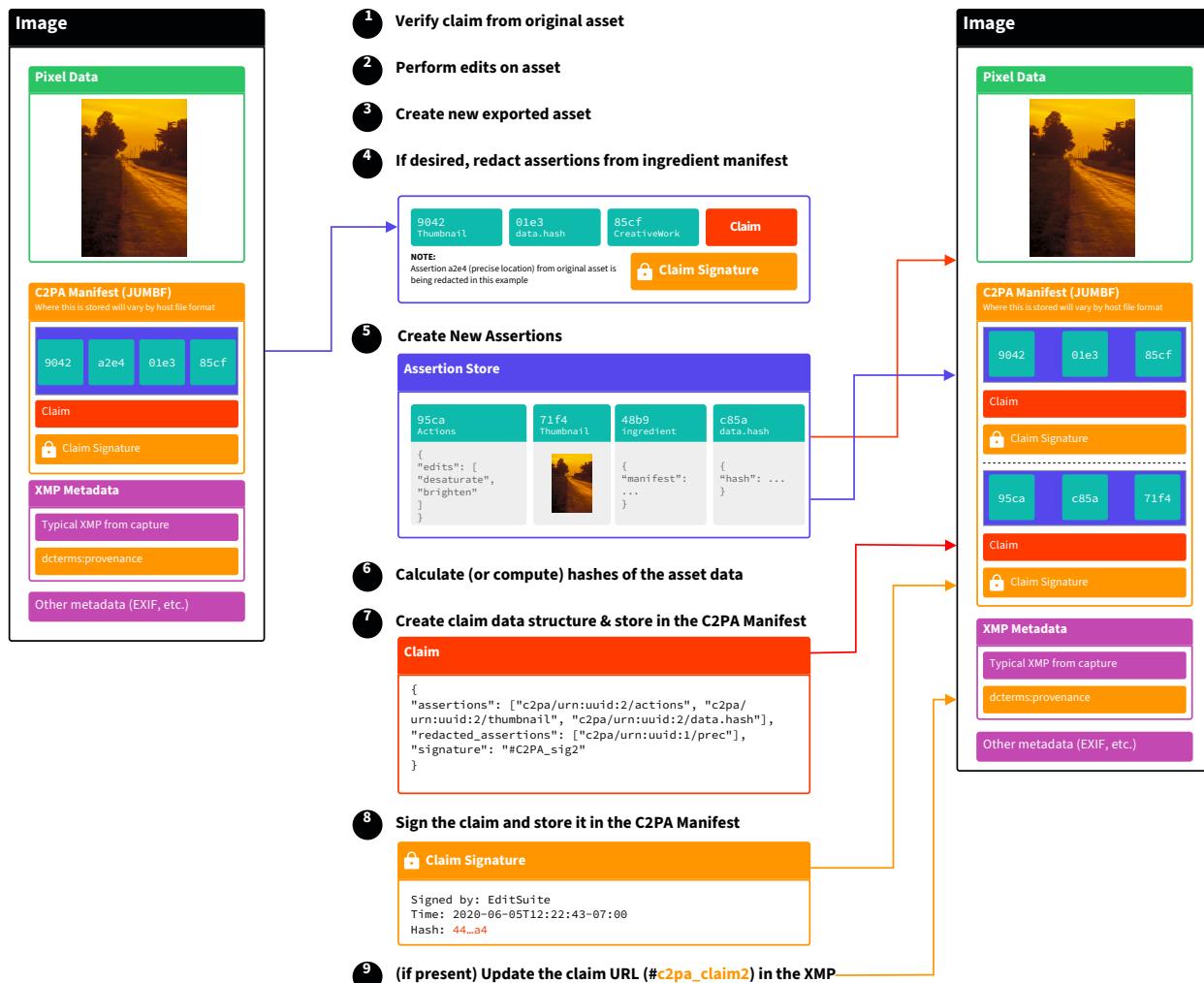


Figure 7. Redacting assertions in a secondary claim

9.4. Multiple Step Processing

Some asset file formats require file offsets of the C2PA Manifest Store and asset content to be fixed before the manifest is signed, so that content bindings will correctly align with the content they authenticate. Unfortunately, the size of a manifest and its signature cannot be precisely known until after signing, which could cause file offsets to change. For example, in [JPEG-1](#) files, the entire C2PA Manifest Store must appear in the file before the image data, and so its size will affect the file offsets of content being authenticated.

To accomplish this, a multiple step approach is taken, similar to how signatures in PDF are done.

9.4.1. Prepare the XMP

For those C2PA-enabled assets that contain embedded XMP, start by creating the XMP data stream and then serializing it into the asset in the standard location reserved for it in the format of the asset. The XMP stream may include the [active manifest reference](#).

NOTE

While it is possible to add the XMP data to the list of exclusions in a data hash assertion, doing so is not recommended as it would remove tamper detection of that asset metadata.

9.4.2. Create content bindings

When creating a [standard manifest](#), its claim shall include one or more content binding assertions in its list of assertions to ensure that the asset is tamper-evident.

In many cases, such as with JPEG-1, it is not possible to hash the asset in its entirety because the manifest will be embedded in the middle of the file, so the size or location manifest data will not be known at the time the asset hash is calculated. This circular dependency is avoided by allowing exclusion ranges to be specified during hashing. When exclusion ranges are specified, a single hash is performed, but only over the asset ranges that are not in any of the exclusions.

For example, if a manifest is embedded in the center of a JPEG-1 file in an APP11 segment, then the claim creator can exclude the APP11 segment(s) from the hash calculation.

In order to prevent insertion attacks, it is desirable to have only a single exclusion range when possible. When the size or location (or both) of the manifest in the asset is not known, then the [start](#) and [length](#) values in the data hash assertion shall both be zero and the size of the [pad](#) value should be large enough to accommodate writing in the values in the second pass. At least 16 bytes is recommended.

Create the data hash assertion and add it to the assertion store.

9.4.3. Create a temporary Claim and Signature

Add the newly created data hash assertion reference to the claim's assertion list providing a temporary hash value, such as empty spaces.

At this point, the temporary claim is complete and can be added to the C2PA Manifest being created.

Since the claim is only temporary at this time, it is not possible to sign it. To ensure the claim signature box contains a valid CBOR structure, create a temporary [COSE_Sign1_Tagged](#) structure as described in [RFC 8152 section 4.2](#). The [COSE_Sign1_Tagged](#) is a tag byte followed by a [COSE_Sign1](#) structure, which is a four-element CBOR array. Construct the array as follows:

- The first element is the [protected](#) header bucket ([RFC 8152 section 3](#)). Create an empty bucket by placing a [bstr](#) of size 0 in this position.
- The second element is the [unprotected](#) header bucket, which is a CBOR map. Create a map of 1 pair. Use the string [pad](#) as the label, and place a [bstr](#) of the desired padding size filled with zero bytes as the value. A 25 kilobyte size is recommended for the initial size of this padding.
- The third element is the [payload](#). Place the value [nil](#) (CBOR major type 7, value 22) here.
- The fourth element is [signature](#). Place a [bstr](#) of size 0 here.

9.4.4. Complete the C2PA Manifest

At this point all of the boxes that comprise the entire C2PA Manifest for the asset are completed and can be (if not already) constructed into its final form. The asset's C2PA Manifest, along with the manifests of any ingredients, are combined together to form the complete C2PA Manifest Store. The active manifest must be the last C2PA Manifest superbox in the C2PA Manifest Store superbox. The C2PA Manifest Store can then be embedded into the asset as discussed in [Section 10.3, “Embedding manifests into assets”](#).

9.4.5. Going back and filling in

Now that the C2PA Manifest Store has been embedded into the asset, the starting offset and the length of the active manifest can be updated in its data hash assertion. It is necessary that when doing so, you don't change the size of the assertion's box, only its data. This is done by adjusting the value of the `pad` field to be the necessary length to "fill up" the remaining bytes.

Preferred/deterministic CBOR serialization of `pad` uses a variable length integer to specify the length of the encoded binary data. When the length goes from zero to 1 byte, or 1 to 2 bytes (etc.), the length of the resulting pad jumps by two bytes. This means that not all paddings can be expressed using a single padding field. For example, 24-byte and 26-byte pads can be created, but a 25-byte pad cannot. If this situation arises, the desired padding can be split between `pad` and `pad2`. For example, to make a 25-byte pad, a claim generator can encode 19 bytes into `pad` (resulting in an encoded length of 20 bytes), and 4 bytes into `pad2` (resulting in 5 bytes.)

Once the data hash assertion has been updated, it can be hashed and the hash written over the empty spaces that were used previously to hold the location.

The claim is now complete, and it can be hashed and signed as described in [Section 9.3.2.4, “Signing a Claim”](#), with the resultant signature filling the pre-allocated space. The `pad` header can then be shrunk as required so that the claim signature box remains the same size; because this header is unprotected, changing it does not invalidate the claim signature.

If the serialized `COSE_Sign1_Tagged` structure exceeds the reserved size of the C2PA Claim Signature box, multiple step processing must be repeated with a larger padding size chosen in [Section 9.4.3, “Create a temporary Claim and Signature”](#). Revocation information retrieved during the previous attempt should be reusable if it is still within its validity interval ([RFC 6960 section 4.2.2.1](#)), but a new time-stamp will be required on the new claim with the file offsets changed as the result of added padding.

Chapter 10. Manifests

10.1. Use of JUMBF

In order to support many of the requirements of C2PA, C2PA Manifests needed to be stored (serialized) into a structured binary data store that enabled some specific functionality including:

- Ability to store multiple manifests (e.g., parents and ingredients) in a single container
- Ability to refer to individual elements (both within and across manifests) via URIs
- Ability to clearly identify the parts of an element to be hashed
- Ability to store pre-defined data types used by C2PA (e.g., JSON and CBOR)
- Ability to store arbitrary data formats (e.g., XML, JPEG, etc.)

In addition to supporting all of the requirements above, our chosen container format - [JUMBF, ISO 19566-5](#) - is also natively supported by the JPEG family of formats and is compatible with the box-based model (i.e. [ISOBMFF, ISO 14496-12](#)) used by many common image and video file formats. Using JUMBF enables all the same benefits (and a few extras, such as [URI References](#)) while being able to work with classic image formats, such as JPEG/JFIF and PNG as well as 3D and document (e.g., PDF) formats. This serialized format shall be used also in formats that do not natively support JUMBF, or when manifests are stored separately from the asset, such as in a separate file or URI location.

Since most of the standard assertions, as well the claim signature, are serialized as CBOR, using CBOR for the entire manifest was considered but not chosen because CBOR is not a container format. It could be used as one through having to re-define how CBOR would be used to provide the features natively supported by JUMBF.

NOTE For example, to store a "blob of JSON" inside of CBOR, and know that it is JSON (and not some other format) would require designing a data structure for storing such things. Then the parent structure would need to be defined as to how to carry that structure. This same concept would also have to be done for each of the native features of JUMBF.

While it would certainly be possible to re-implement all of the required functionality entirely in CBOR, it would be a lot of work and would not fully remove the need for a JUMBF/BMFF parser in all implementations.

A C2PA Manifest Consumer shall never process an assertion, assertion store, claim, claim signature or manifest that is not contained inside of a C2PA Manifest Store. Additionally, when a C2PA Manifest Consumer encounters a JUMBF box or superbox whose UUID it doesn't recognize, it shall skip over (and ignore) its contents.

NOTE This means that the C2PA Manifest Consumer can process private boxes that it knows about, but ignore ones of which it is unaware.

10.1.1. C2PA Box details

C2PA data is serialized into a JUMBF-compatible box structure, and the outermost box is referred to as the C2PA Manifest Store. An example C2PA Manifest Store, with a single manifest, might look like this:

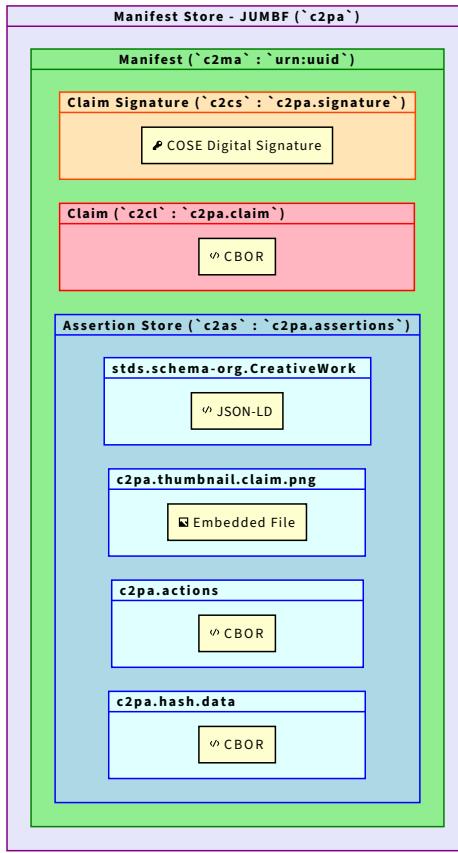


Figure 8. C2PA Manifest Store

The C2PA Manifest Store is a JUMBF superbox composed of a series of other JUMBF boxes and superboxes, each identified by their own UUID and label in their JUMBF Description box. The C2PA Manifest Store shall have a label of **c2pa**, a UUID of **0x63327061-0011-0010-8000-00AA00389B71 (c2pa)** and shall contain one or more C2PA Manifest superboxes. The C2PA Manifest Store may also contain JUMBF superboxes whose UUIDs are not defined in this specification.

NOTE

Allowing other superboxes enables custom extensions to C2PA as well as enabling the addition of new boxes in future versions of this specification without breaking compatibility.

Each C2PA Manifest superbox shall contain the data created at the time a claim is issued including the C2PA Assertion Store, a C2PA Claim, and a C2PA Claim Signature. It may also contain a [C2PA Credential Store](#). A C2PA Manifest may also contain JUMBF superboxes whose UUIDs are not defined in this specification.

The UUID for each C2PA Manifest shall be either **0x63326D61-0011-0010-8000-00AA00389B71 (c2ma)** or **0x6332756D-0011-0010-8000-00AA00389B71 (c2um)** depending on the [type of manifest](#). In order to enable uniquely identifying each manifest, they shall be labelled with a [RFC 4122, UUID](#) optionally proceeded by an identifier of the claim generator and a **:**. An example label for the fictitious ACME claim generator might look like

acme:urn:uuid:F9168C5E-CEB2-4FAA-B6BF-329BF39FA1E4.

The C2PA Assertion Store is a superbox that shall have a label of `c2pa.assertions` and a UUID of `0x63326173-0011-0010-8000-00AA00389B71 (c2as)`. It shall contain one or more JUMBF superboxes (called C2PA Assertion boxes) whose JUMBF type defines the BMFF type of the sub-boxes that contain the assertion data (see ISO 19566-5, Annex B, Table B.1 and ISO 19566-5/AMD-1, Annex B). These superboxes shall each have a label as defined in [Standard Assertions](#). The JUMBF Content Type (ISO 19566-5, Annex B) box(es) contained in each assertion superbox should be CBOR Content Type (`cbor`), JSON Content Type (`json`), Embedded File Content Type (`bfdb` & `bidb`) or UUID Content Type (`uuid`) though any Content Type defined in JUMBF and its amendments is permitted. The C2PA Assertion Store shall not contain any JUMBF boxes or superboxes that are not JUMBF Content Boxes.

Custom assertions containing other formats/serializations of data, such as encrypted data, are supported through the use of a UUID Content Box containing the custom UUID followed by the data (see ISO 19566-5, B.5).

The C2PA Claim box shall have a label of `c2pa.claim`, a UUID of `0x6332636C-0011-0010-8000-00AA00389B71 (c2cl)` and shall consist of a single CBOR Content Type box (`cbor`).

The C2PA Claim Signature box shall have a label of `c2pa.signature`, a UUID of `0x63326373-0011-0010-8000-00AA00389B71 (c2cs)` and shall consist of a single CBOR Content Type box (`cbor`).

10.1.1.1. Credential Storage

A C2PA Credential Store (VCStore) is a JUMBF superbox that shall contain only one or more JSON Content Type boxes (ISO 19566-5, Annex B.4). It shall not contain any other type of JUMBF box or superbox. It shall have a label of `c2pa.credentials` and a UUID of `0x63327663-0011-0010-8000-00AA00389B71 (c2vc)`.

When storing W3C Verifiable Credentials in a VCStore, each one shall be labelled with the value of the `id` field of the `credentialSubject` of the VC itself. Since the `id` is guaranteed to be unique, this ensures that the URI to that credential will be unique.

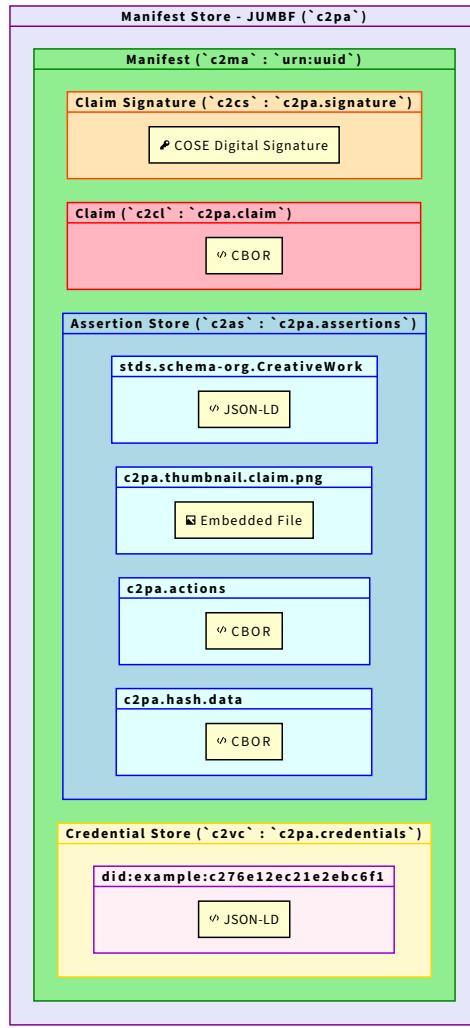


Figure 9. C2PA Manifest Store with Credentials

10.1.1.2. Ingredient Storage

When a C2PA Manifest includes [ingredient assertions](#), and an ingredient contains a C2PA manifest, that manifest must be brought into this asset to ensure that the provenance data is kept intact. Such ingredient manifests are added to the C2PA Manifest Store as a peer of the C2PA Manifest for the asset itself.

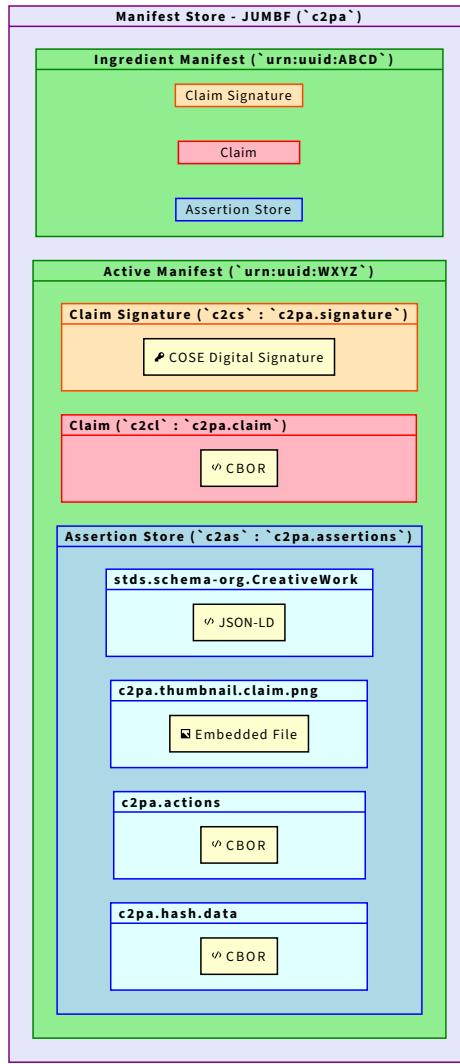


Figure 10. C2PA Manifest Store With an Ingredient

10.2. Types of Manifests

10.2.1. Commonalities

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

10.2.2. Standard Manifests

A standard manifest (UUID: [0x63326D61-0011-0010-8000-00AA00389B71 \(c2ma\)](#)) shall contain at least one [hard binding to content](#) 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.

10.2.3. Update Manifests

There are, however, provenance workflows where additional assertions need to be added but the digital content is not changed. In these workflows, an update manifest (UUID: [0x6332756D-0011-0010-8000-00AA00389B71](#) (`c2um`)) can be used. An update manifest shall not contain assertions of types `c2pa.hash.data` or `c2pa.hash.bmff` because the content has not changed and therefore the bindings need not be updated. It shall also not contain an assertion of type `c2pa.actions` because that assertion is defined to describe "changes to the digital content". The update manifest shall contain exactly one `c2pa.ingredient` assertion that (a) includes a `c2pa_manifest` field with a value that is the [URI reference](#) to that [C2PA Manifest](#) that is being updated and (b) has the value of `parentOf` for the `relationship` field.

NOTE

The ingredient's manifest (referenced via the `c2pa_manifest` field) can be either a standard manifest or an update manifest.

10.3. Embedding manifests into assets

10.3.1. Embedding manifests into non-BMFF-based assets

When embedding the C2PA Manifest Store into an asset, the location will vary based on the type of the asset. Here are some well-known types and the location to use:

JPEG

Refer to [Section 10.3.1.1, “Embedding manifests into JPEG”](#) for more information.

PNG

Refer to [Section 10.3.1.2, “Embedding manifests into PNG”](#) for more information.

SVG

Refer to [Section 10.3.1.3, “Embedding manifests into SVG”](#) for more information.

PDF

Refer to [Section 10.3.1.4, “Embedding manifests into PDFs”](#) for more information.

FLAC

Refer to [Section 10.3.1.5, “Embedding manifests into ID3”](#) for more information.

EDITORS NOTE

NOTE

C2PA is asking for feedback from the audio community if embedding the manifest in an ID3v2 container will work with FLAC or if we will need to use the native FLAC container.

MP3

Refer to [Section 10.3.1.5, “Embedding manifests into ID3”](#) for more information.

BMFF-based formats

The box specified in [Section 10.3.2, “Embedding manifests into BMFF-based assets”](#).

NOTE

A C2PA Manifest Store can be embedded in BMFF-based downloadable audio files using codecs such as the Advanced Audio Codec (AAC) or the Apple Lossless Audio Codec (ALAC).

Additional locations for other file formats will be added in the future.

NOTE

Non-BMFF-based audio formats which are being considered for addition to this specification include MP3, Ogg Vorbis and the native container version of the Free Lossless Audio Codec (Native FLAC).

NOTE

Many classic image formats such as GIF and BMP do not support the embedding of arbitrary data, so that the use of an [external manifest](#) is required.

10.3.1.1. Embedding manifests into JPEG

The C2PA Manifest shall be embedded as the data contained in an **APP11** Marker as defined in [JPEG XT, ISO/IEC 18477-3](#).

Since a single marker segment in JPEG-1 cannot be larger than 64K, it is likely that multiple **APP11** segments will be required, and they shall be constructed as per the JPEG-1 standard and [ISO 19566-5, D.2](#). When writing multiple segments, they shall be written in sequential order and they shall be contiguous (i.e. one segment immediately following the next).

10.3.1.2. Embedding manifests into PNG

The C2PA Manifest shall be embedded using an ancillary, private, not safe to copy, chunk type of '**caBX**' (as per [PNG, 4.7.2](#)). It is recommended that the '**caBX**' chuck precede the '**IDAT**' chunks.

NOTE

Although PNG supports it, it's considered bad-form to have a data block after the '**IDAT**' and before the '**IEND**'. (The exception being animated PNG blocks)

10.3.1.3. Embedding manifests into SVG

[SVG](#) is an XML-based format that can exist either stand-alone or embedded into other text-based formats such as HTML. As such, it is necessary to Base64 encode the binary C2PA manifest to perform the embedding. While this section describes how to do that, the use of an [external manifest](#) is preferred.

The C2PA Manifest shall be embedded as the Base64-encoded value of a **c2pa:manifest** element in the [metadata element](#) of the SVG.

An example might look like this (with the actual C2PA manifest data left out)

```

<?xml version="1.0" standalone="yes"?>
<svg width="4in" height="3in" version="1.1"
      xmlns = 'http://www.w3.org/2000/svg'>
  <metadata>
    <c2pa:manifest>...Base64 data goes here...</c2pa:manifest>
  </metadata>
</svg>

```

10.3.1.4. Embedding manifests into PDFs

The C2PA Manifest shall be embedded using an embedded file specification (ISO 32000, 7.11.3). The file specification dictionary shall have an **AFRelationship** key whose value is *C2PA_Manifest*. If the C2PA manifest is embedded into an encrypted PDF, the embedded file stream shall use an **Identity** crypt filter.

The embedded file specification shall be the value (via indirect object) of the **AF** key in the document catalog dictionary. It shall also be referenced (via indirect object) either from the *EmbeddedFiles* NameTree (*/Catalog/Names/EmbeddedFiles*) or from a *FileAttachment* annotation. The annotation approach shall be used when adding a C2PA Manifest to a PDF that already has an existing PDF signature in order to avoid invalidating its **DocMDP** restrictions.

It is possible to add a PDF signature (certifying or approval) to a PDF following the inclusion of a C2PA Manifest Store, but only if it is known at the time of the creation of the active manifest. Since the PDF signature will change the contents of the PDF after the C2PA manifest is signed, the size and location of value of its **Contents** key must be determined before C2PA signing and added to the list of exclusions in the **data.hash** assertion, so that the C2PA signature is not invalidated by the addition of the PDF signature.

NOTE

Adding the PDF signature in addition to the C2PA's claim signature improves compatibility with the existing PDF ecosystem.

10.3.1.5. Embedding manifests into ID3

The C2PA Manifest Store shall be embedded into a ID3v2-compatible, compressed audio file (e.g., MP3 or FLAC) file as the Encapsulated object data of a General Encapsulated Object (GEOB) as defined in https://id3.org/id3v2.3.0#General_encapsulated_object. The GEOB's **MIME type** field shall be present and shall use the value for the media type for JUMBF as described in [Section 10.4, “External Manifests”](#).

10.3.2. Embedding manifests into BMFF-based assets

10.3.2.1. The '**uuid**' Box for C2PA

All BMFF-based C2PA assets, whether they are timed (e.g., videos with or without audio tracks), untimed (e.g., still photos) or mixed (e.g., live or animated photos) audiovisual media, shall use a '**uuid**' box that adheres to the following syntax and semantics defined below.

EDITORS NOTE

- NOTE** The reason that a '**uuid**' box instead of a '**c2pa**' box is being used is that browsers based on Chromium will immediately fail playback when they encounter any unknown top-level boxes.

Some file formats that are BMFF-based and would be supported via this method include:

- MPEG-4 code-points, either complete (**.mp4**) or fragmented (**.m4s**); downloadable audio files (**.m4a**)
- HEIF (**.heif**, **.heic**)
- AVIF (**.avif**)

10.3.2.1.1. Definition

Box Type: '**uuid**' Extended Box Type: 0xD8, 0xFE, 0xC3, 0xD6, 0x1B, 0x0E, 0x48, 0x3C, 0x92, 0x97, 0x58, 0x28, 0x87, 0x7E, 0xC4, 0x81 Container: File Mandatory: No Quantity: Zero or more

The Coalition for Content Provenance and Authenticity ('**uuid**' with aforementioned uuid) box embeds provenance into BMFF. One such box contains a manifest store, and there may be one or more auxiliary boxes containing additional information required for validation.

10.3.2.1.2. Syntax

```
aligned(8) class ContentProvenanceBox extends FullBox(`'uuid'`, extended_type = 0xD8 0xFE
0xC3 0xD6 0x1B 0x0E 0x48 0x3C 0x92 0x97 0xE8 0x28 0x87 0x7E 0xC4 0x81, version = 0, 0) {
    string box_purpose;
    bit(8) data[];
}
```

10.3.2.1.3. Semantics

box_purpose	[indicates purpose of box]
data	[depends on box_purpose]

The box_purpose and fields that depend on it are described below for each box purpose.

Regarding unique ids:

There are cases, such as fragmented MP4 (fMP4), where the id for a subset of the asset, such as the track_id field of the '**tkhd**' box, is only locally unique to a subset of the overall asset rather than globally unique to the asset.

NOTE

Because a globally unique id is required to determine what to hash, a unique id is included. This unique id does not equal any value from the original asset; each value is instead defined when the manifest is created. The unique id is then combined with an associated local id to form an id that's globally unique to the entire asset.

10.3.2.2. Box Containing the Manifest

The box containing the manifest shall appear before the first '**mdat**' box in the file.

The fields in the corresponding box described above shall be set as follows.

box_purpose	For a manifest, this value shall be manifest .
data	When box_purpose is manifest , the first 8 bytes shall be the absolute file offset to the first auxiliary ' uuid ' C2PA box with box_purpose equal to merkle . If this file contains no such boxes, the first 8 bytes shall be zero. Those 8 bytes shall be followed by the raw manifest bytes followed by zero or more unused padding bytes.

For fragmented mp4 files, an identical '**uuid**' C2PA box of type **manifest** shall be present in each initialization segment; the manifest must be identical.

10.3.2.3. Auxiliary '**c2pa**' Boxes for Large and Fragmented Files

Some files have one or more very large '**mdat**' boxes (e.g., large video or image files which may be downloaded and rendered progressively) or large numbers of independent '**mdat**' boxes (e.g., fMP4 where each fragment can be downloaded independently).

In these cases, it is unreasonable to require a client to completely download all '**mdat**' box(es) before validating any portion of the asset. Avoiding that necessity is resolved by using multiple hashes.

For each large '**mdat**' box, subsets of the box have individual hashes that can be validated independently; how to determine these subsets is specified below. For fMP4 content where each '**mdat**' box can be downloaded independently, each fragment has its own individual hash.

In the simplest case, all of these hashes are stored in the manifest. Each subset has an auxiliary '**uuid**' C2PA box that declares how to locate its hash in the manifest; refer to the note regarding unique ids above for why this is the case.

However, for sufficiently large assets, including every subset's hash in the manifest itself would increase the size of the manifest to one or more megabytes.

Avoiding such a large manifest for a large asset is achieved by using one or more Merkle Trees. * For a large non-fragmented asset that contains one or more '**mdat**' boxes in a single large file, one Merkle Tree is used for each '**mdat**' box. * For a large fragmented asset that contains a set of '**mdat**' boxes for a single track which may be spread across multiple files, one Merkle Tree is used for each track.

In either case: * Each leaf node of any given Merkle Tree is the subset's hash. * The manifest stores one row of each Merkle Tree. * The auxiliary '**uuid**' C2PA box that exists for each subset indicates which Merkle Tree row in the manifest it requires and which leaf node it represents. It also includes any additional hash(es) from the Merkle Tree

necessary to derive a hash in the manifest's Merkle Tree row.

The selection of which Merkle Tree row to store in the manifest creates a size tradeoff within the asset. Specifically, storing a single hash per Merkle Tree in the manifest minimizes the size of the manifest but requires $\log_2(\text{subsets})$ to be stored in each subset-specific box. Each time the number of hashes stored in the manifest for a Merkle Tree is doubled (by moving "down" one Merkle Tree row), the number of hashes stored in each subset-specific box decreases by one. Thus, increasing the size of the manifest decreases the size of the entire asset and vice-versa, and since hashes for individual subsets are replicated across subsets as required to derive a manifest-specified hash, the tradeoff is not 1 to 1.

Making this size tradeoff is left up to the implementation creating the manifest; this spec neither mandates nor recommends that any specific Merkle Tree row be stored in the manifest. That said, because the simplest case of storing all subset hashes in the manifest is equivalent to using a Merkle Tree where the leaf nodes are stored in the manifest, the same Merkle Tree construction is used for multiple hashes in all cases. That construction is defined as follows.

The portion of the manifest containing the BMFF Hash shall include the **merkle** field. Refer to [Section 8.2.2, “Hashing a BMFF-formatted asset”](#) for more information.

For large '**mdat**' boxes that can be validated piecemeal, two or more auxiliary '**uuid**' C2PA boxes with `box_purpose` set to '**merkle**' as described below shall be included in the single asset file. They shall follow the last '**mdat**' box in the file.

For timed-media where an '**stco**' or '**co64**' box is present, the hash used for a given leaf node in the merkle tree shall be computed over an individual subset of samples as defined by that box.

For untimed-media where an '**iloc**' box is present (such as HEIF or AVIF), the hash used for a given leaf node in the merkle tree shall be computed over an individual item as defined by that box.

Regardless of the subset-defining mechanism, all such auxiliary '**uuid**' C2PA boxes shall occur in the same sequence as the subsets they hash as specified by the '**stco**', '**co64**', or '**iloc**' box regardless of their location in the '**mdat**'. They shall also be grouped such that a single Merkle Tree's auxiliary '**uuid**' C2PA boxes are sequential with no intervening boxes.

EDITORS NOTE

NOTE

Are there other scenarios where an '**mdat**' can be large enough to be worth dividing where neither '**stco**', '**co64**', nor '**iloc**' is present? If so, what box(es) should be used to decide on subset division points?

For fragmented mp4 files, one auxiliary '**uuid**' C2PA box with `box_purpose` set to '**merkle**' as described below shall be included in each fragment file per '**moof**' / '**mdat**' pair in that file. Each one shall be between the '**moof**' and '**mdat**' box it covers. The hash used for a given leaf node in the merkle tree shall be over all data following the last '**mdat**' box preceding that '**uuid**' C2PA box (or start of file) and preceding the first '**moof**' box following that '**uuid**' C2PA box (or end of file) except data excluded by the exclusion list.

NOTE

This specification does not enable support for fMP4 assets where individual fragments contain more than one '**moof**' box and/or '**mdat**' box.

Regardless of how the asset is structured, the fields in the corresponding box described above shall be set as follows.

box_purpose	For an auxiliary ' uuid ' C2PA box, this value shall be merkle .
data	When box_purpose is merkle , this value shall contain raw CBOR bytes indicating how to validate a portion of the asset as defined as follows. If there are multiple auxiliary ' uuid ' C2PA boxes with box_purpose merkle for a given merkle tree in a single file, each shall be followed by sufficient padding bytes (zero or more) to make all auxiliary ' uuid ' C2PA boxes for that merkle tree a fixed size.

NOTE

When there are more than one of these boxes in a single file, i.e., the case where there are large '**mdat**'(s) being validated piecemeal, a fixed size is required in order to enable a progressively downloading client to only download the boxes it needs to begin validation rather than the entire merkle tree. Such a client can download enough of the first of these boxes based on the file offset in the **manifest** to determine if its **uniqueId** and **localId** match the '**mdat**' it is trying to validate. If they do, it can determine the file offset to the box it needs to validate by multiplying the subset number by that size then download just that box. Otherwise, it can determine the file offset to the beginning of the next merkle tree by multiplying that fixed size by the current merkle tree's total number of leaf nodes, and it can repeat this process until it locates the box it needs. The total download size for this subset of boxes is very small relative to the size of a single subset.

NOTE

A fragmented mp4 asset may also be stored as a single flat mp4 file with a single **moov** for all tracks and then **moof/mdat** pairs for each fragment. Such a file shall be handled as if the individual fragments were moved into their own files, manifest and subset processing was performed, and then the files were appended together.

10.3.2.3.1. Schema and Example

The [CDDL Definition](#) for this type is:

```

; The data structure used to store sufficient information to validate a single 'mdat' box or
; a portion of an 'mdat' box when a Merkle tree is used",
bmff-merkle-map = {
  "uniqueId": int, ; A unique integer used to differentiate local ids
  "localId": int, ; A local id indicating which 'mdat' box this entry pertains to. This may
not be globally unique
  "location": int, ; Zero-based index into the leaf-most merkle tree row corresponding to
this 'mdat' box or portion of this 'mdat' box
  ? "hashes": [1* bstr], ; An ordered array representing the set of additional hashes
required to reach a hash in the merkle tree specified in the manifest from leaf-most (peer
of this node) to root-most (child of node in manifest). Note that this array may not be
present, e.g. if the manifest itself contains the leaf-most row of the merkle tree. Null
hashes are not included in this array. The algorithm used shall be determined using the
`alg` field from the corresponding entry in the `merkle` field array in the BMFF hash
structure.
}

```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```
{
  "uniqueId": 1339,
  "localId": 4402,
  "location": 2203,
  "hashes": [
    b64 'TWVub3JhaA=='
  ]
}
```

For files that use a '[tkhd](#)' box to indicate individual tracks, the localId in the preceding CBOR shall be set to the track_id field of the '[tkhd](#)' box pertaining to the '[mdat](#)' being hashed.

EDITORS NOTE

NOTE How do we handle the case where there are multiple '[mdat](#)' boxes for untimed media, e.g. a file that includes multiple image '[mdat](#)' boxes? How do we handle the case where there are multiple '[mdat](#)' boxes of different types, e.g. a file that has both audio/video '[mdat](#)' boxes as well as an image '[mdat](#)' box (e.g. for a thumbnail)? For scenarios such as these, a [tkhd](#) box alone cannot be used to reference the complete set of different [mdat](#) boxes when there are more than one; it may not be present at all. What box value(s) should be used for localId instead?

10.3.2.4. Dynamic stream generation

Many adaptive bitrate streaming (ABR) implementations store a single version of an asset, e.g., as a flat MP4 or in another intermediate format, and generate individual asset streams using various codecs, bitrates, etc. at consumption time. As a result, such a server must either hash said streams and create a manifest each time the content is consumed or, if generation is deterministic, create and cache the hashes and manifests once and then embed them at consumption time.

This also means that such a server must have a signing identity of its own that will be trusted by validators or be able

to sign the generated files on behalf of the content creator in a secure and trustworthy manner.

10.3.2.5. Exclusion List Requirements

For all `c2pa.hash.bmff` assertions, the following entries shall always appear on the exclusion list.

The entire '`uuid`' C2PA box shall be excluded. (The '`data`' field is ensuring that other '`uuid`' boxes are not excluded.)

```
xpath = "/uuid"
data = [ { offset = 8, data = "2P7D1hsOSDySl1goh37EgQ==" } ]
```

The entire '`ftyp`', '`iloc`', and '`tfra`' boxes shall be excluded.

```
xpath = "/ftyp"
```

```
xpath = "/meta/iloc"
```

```
xpath = "/mfra/tfra"
```

The entire '`stco`' and '`co64`' box except for the first 16 bytes shall be excluded.

```
xpath = "/moov/trak/mdia/minf/stbl/stco"
subset = [ { offset = 16, length = 0 } ]
```

```
xpath = "/moov/trak/mdia/minf/stbl/co64"
subset = [ { offset = 16, length = 0 } ]
```

The '`base_data_offset`' field of the '`tfhd`' box shall be excluded. Note that this field only appears when the lowest order bit of flags is set to 1. This field must be excluded because it is a file offset and including it would invalidate the hash once the manifest was embedded into the file. Failing to include the remaining fields of this box in the hash would enable the audio/video speed to be increased or decreased without invalidating the hash, thus defeating a stated goal of C2PA.

```
xpath = "/moof/traf/tfhd"
flags = "AAAB"
exact = false
subset = [ { 16, 8 } ]
```

The '`data_offset`' field of the '`trun`' box shall be excluded. Note that this field only appears when the lowest order bit of flags is set to 1. This field must be excluded because it is a file offset and including it would invalidate the hash once the manifest was embedded into the file. Failing to include the remaining fields of this box in the hash would enable the audio/video speed to be increased or decreased without invalidating the hash, thus defeating a

stated goal of C2PA.

```
xpath = "/moof/traf/trun"  
flags = "AAAB"  
exact = false  
subset = [ { 16, 4 } ]
```

10.3.2.6. Timed-media streams that are neither audio nor video

Timed-media streams that are neither audio nor video, such as text streams for captions, that the claim generator wishes to make tamper evident shall be handled the same way as audio and video streams.

10.3.2.7. External references

Externally referenced content declared inside BMFF boxes, such as in a '['dref'](#)', '['url'](#)', or '['urn'](#)' box, that the claim generator wishes to make tamper evident shall **not** exclude the referencing box AND shall include a separate [cloud data assertion](#) for each external reference to be hashed.

10.3.2.8. Size requirements

If a BMFF-based asset uses 32-bit sizes or offsets in any box(es), e.g. the '['stco'](#)' box, and adding boxes to conform to this specification will push the file size over 4 gigabytes, it is the responsibility of the manifest creator to edit the file to use appropriate sizes and offsets, e.g. by replacing the '['stco'](#)' box with a '['co64'](#)' box, before creating the manifest.

10.4. 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 manifest 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](#). As the C2PA Manifest Store is a JUMBF box, it shall be served with a JUMBF Media Type.

Editors Note

NOTE The C2PA will work with the ISO JPEG Systems committee to register a media type for JUMBF with IANA.

Some common reasons to use an external manifest are:

- It may not be technically possible, such as with a [.txt](#) file.
- It may not be practical, such as when the size of the manifest is larger than the asset's digital content.
- It may not be appropriate, such as when it would modify an asset that shouldn't be modified.

NOTE: a good example of this is creating a manifest for a pre-existing asset.

10.5. Embedding a Reference to the Active Manifest

If the asset has embedded XMP, it is recommended that the claim generator add a [dcterms:provenance](#) key to the

XMP, the value (a URI reference) being where to locate the active C2PA Manifest. The URI shall either be a JUMBF URI for an embedded manifest or a standard URI for any non-embedded scenarios, whether stored remotely or locally on the same storage system as the asset itself.

An example JUMBF URI would be stored as:

```
<dcterms:provenance>self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-  
329BF39FA1E4</dcterms:provenance>
```

Chapter 11. Entity Diagram

The following diagram provides a look at how all of the pieces of the C2PA system integrate and relate to each other.

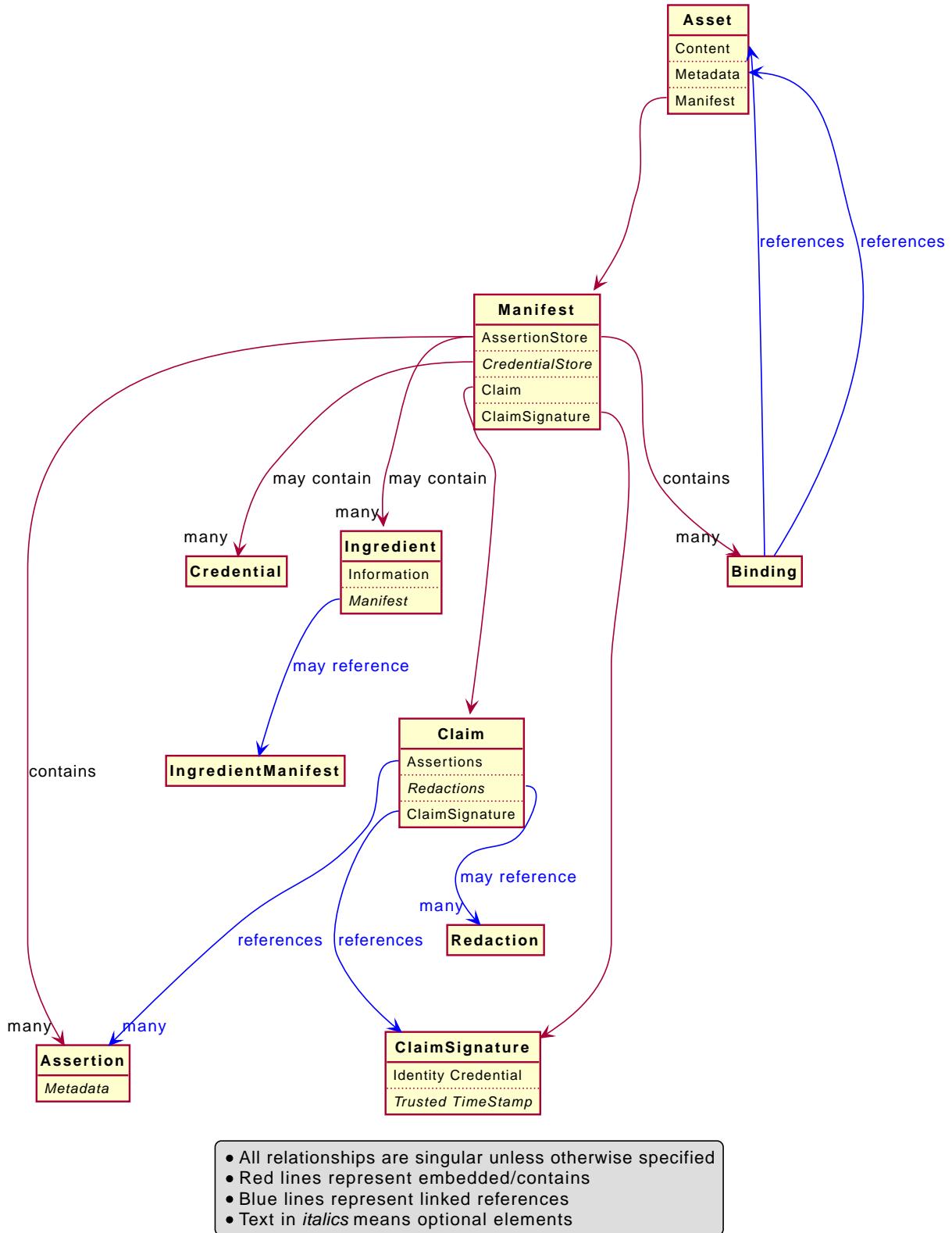


Figure 11. C2PA Entity Diagram

Chapter 12. Cryptography

12.1. Hashing

All cryptographic hashes that are stored in a C2PA manifest shall be generated using one of the hash algorithms as described in this section. This section defines both:

- A list of hash algorithms that are allowed for generating hashes of new content as well as required for validating hashes of existing content (the allowed list), and
- A list of hash algorithms that are required to be supported for validating hashes of existing content but are not allowed for generating hashes of new content (the deprecated list).

NOTE

This section does not govern algorithms used for soft bindings as described in [Section 8.3, “Soft Bindings”](#).

An algorithm must appear in no more than one list. An algorithm that is instantiated over multiple output lengths (such as the various lengths of SHA2) will each be considered different algorithms, and each instantiation must be listed separately. If an algorithm does not appear in either list, it is forbidden and must not be used or supported. Algorithms can be removed from the lists in order to implement forbidding an algorithm. For this reason, implementations must not support additional algorithms on an optional basis.

Implementers should consult this section in the current version of the specification when releasing software updates and ensure their supported algorithms conform to it.

These lists establish the allowed algorithms for creating hashes and a string algorithm identifier to be used as the algorithm identifier (usually called `alg`) in the corresponding field of C2PA data structures. The outputs of hash functions shall be stored as their binary values encoded into CBOR as byte strings (major type 2) with a declared length. Wherever a field contains the output of a hash function, an algorithm identifier string field shall be present within the same structure, or within an enclosing structure, to declare which algorithm was used. A hash algorithm identifier field should be present in exactly one of these places, but if more than one is present within the structure and its enclosing structures, the nearest identifier must be used. Nearest is defined first as an identifier that is a sibling field of the hash value, and then the immediately enclosing structure, up to the root structure.

The allowed list is:

- SHA2-256 ("sha256")
- SHA2-384 ("sha384")
- SHA2-512 ("sha512")

NOTE

The SHA-3 family of hash algorithms are not on the allowed list for consistency with the digital signature algorithm allowed list, because COSE has not yet established digital signature algorithms that use a SHA-3 algorithm as the hash algorithm. A future version of this specification will incorporate SHA-3 for hashing as well as COSE digital signature algorithms that use SHA-3 when they are available.

The deprecated list is empty.

12.2. Digital Signatures

All digital signatures that are stored in a C2PA manifest shall be generated using one of the digital signature algorithms and key types listed as described in this section. This section defines both:

- A list of digital signature algorithms and key types that are allowed for generating signatures of new content as well as required for validating signatures of existing content (the allowed list), and
- A list of digital signature algorithms and key types that are required to be supported for validating signatures of existing content but are not allowed for generating signatures of new context (the deprecated list).

These lists establish the allowed algorithms and key types by referencing an algorithm identifier from the relevant standards that define algorithms for COSE and their mappings to CBOR identifiers, including but not limited to [RFC 8152](#) and [RFC 8230](#). These standards also specify the hash algorithm used in the signature scheme. Nothing in [Section 12.1, “Hashing”](#) shall apply to this use of hash algorithms; if a digital signature algorithm is present in the digital signature algorithm and key type registry, the use of its specified hash algorithm in the signature scheme shall be allowed and followed.

NOTE

Parenthetical notes in the lists below are explainers provided only as an aid to the reader.

The allowed list is:

- [ES256](#) (ECDSA with SHA-256)
- [ES384](#) (ECDSA with SHA-384)
- [ES512](#) (ECDSA with SHA-512)
- [PS256](#) (RSASSA-PSS using SHA-256 and MGF1 with SHA-256)
- [PS384](#) (RSASSA-PSS using SHA-384 and MGF1 with SHA-384)
- [PS512](#) (RSASSA-PSS using SHA-512 and MGF1 with SHA-512)
- [EdDSA](#) (Edwards-Curve DSA)
 - Ed25519 instance only. No other EdDSA instances are allowed.

The deprecated list is empty.

Implementations must check that keys provided for signing or verification operations are correct for the chosen

algorithm, as required by [RFC 8152 section 8.1](#) for ECDSA, [RFC 8152 section 8.2](#) for EdDSA, and [RFC 8230 section 2](#) and [section 4](#) for RSASSA-PSS. These requirements are summarized here for convenience:

- ECDSA requires elliptic curve keys on the P-256, P-384, or P-521 elliptic curves.
 - Although it is recommended to use P-256 keys with [ES256](#), P-384 keys with [ES384](#), and P-521 keys with [ES512](#), it is not required. Implementations must accept keys on any of these curves for all ECDSA algorithm choices.
- Ed25519 requires elliptic curve keys on the X25519 elliptic curve.
- RSASSA-PSS requires RSA keys with a modulus length of at least 2048 bits.

Implementations must refuse to generate or verify signatures with keys that are not correct for the algorithm choice. Implementations may refuse RSA keys with modulus length greater than 16384 bits.

The signature for the CBOR-encoded claim is produced by CBOR Object Signing and Encryption (COSE) as described in [RFC 8152 sections 4.2](#) and [4.4](#). COSE is always used in "detached content" mode as described in [section 4.1](#).

NOTE

In "detached content" mode, the signed data is stored externally to the [COSE_Sign1_Tagged](#) structure, and the [payload](#) field of the [COSE_Sign1_Tagged](#) structure is always [nil](#). The contents of the [payload](#) field of [Sig_structure](#) in memory, when constructed to compute or verify a digital signature, must be populated with that external data as described by the particular use of digital signature in this specification. The [payload](#) field of [Sig_Structure](#) can never be [nil](#).

For example, when computing or verifying a claim signature, the [payload](#) field of the [Sig_structure](#) will contain the contents of the Claim JUMBF box, as described in [Section 9.3.2.4](#), "Signing a Claim" and [Section 10.1](#), "Use of JUMBF".

- The signature is computed or verified as described in [RFC 8152 section 4.4](#). The following additional requirements apply to the construction of [Sig_structure](#):
 - The value for the [context](#) element shall be [Signature1](#) except where a particular use of digital signatures in this specification specifies using [CounterSignature](#) instead. [Signature](#) shall not be used.
 - The value for the [payload](#) element will be specified by each use of digital signatures in this specification.
 - The [external_aad](#) element shall be a bstr of length zero. External authenticated data must not be used.
 - The [alg](#) header specifying the signature algorithm shall be present in the [body_protected](#) element. [RFC 8152 section 3.1](#)

NOTE

The [alg](#) header is a standard COSE header, and therefore is always included in the protected header map with the integer 1 as its label, as established in the [IANA Cose Header Parameters Registry](#). The literal string [alg](#) is never used as the label. The [sign_protected](#) element is always omitted when using [COSE_Sign1](#).

- All digital signatures in C2PA structures shall be a **COSE_Sign1_Tagged** structure as defined in [RFC 8152 section 4.2](#). **COSE_Sign1_Tagged** contains a **COSE_Sign1** structure. The following additional requirements apply to the construction of **COSE_Sign1_Tagged**:
 - The same **alg** header in the **Sig_structure** above shall be present in the **protected** header bucket.
 - The **payload** field shall be **nil**.

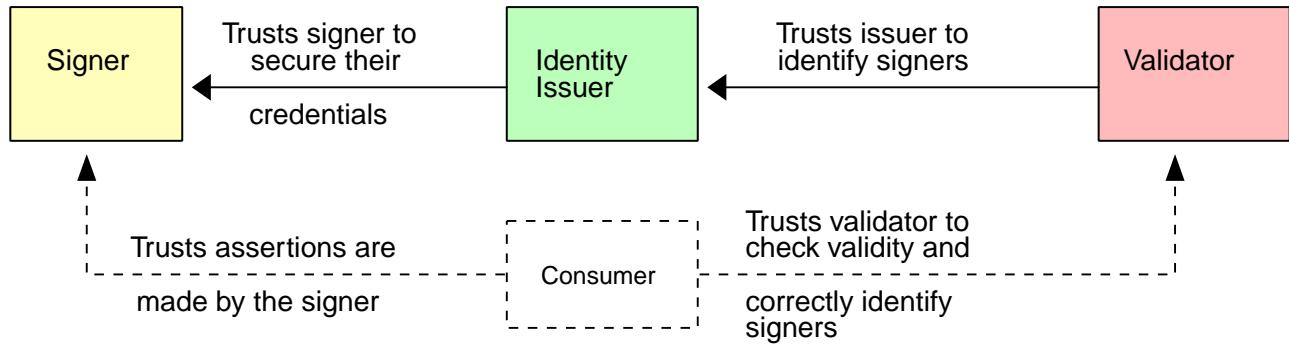
NOTE COSE defines **nil** to be major type 7, value 22 in [section 1.3](#) and uses this value exclusively for detached content. A byte array (major type 2) of length zero cannot be used to indicate detached content.

When producing a signature, if the claim generator can also act as a validator, the claim generator should validate that the signing credential is acceptable according to [Chapter 13, Trust Model](#) for the credential's type, and produce a warning if it is not. The claim generator should still allow signing with that credential if so desired. This may be desirable if it is known the local claim generator's validator has a different configuration than validators used by the expected audience of the asset.

When verifying a signature, the **body_protected** field of **Sig_structure** is populated from the contents of the **protected** header bucket of the **COSE_Sign1_Tagged** structure, but because of the use detached content mode, the **payload** field of the **COSE_Sign1_Tagged** structure will be **nil**. The contents of the **payload** field of **Sig_structure** must be populated from the same external source that was used in the generation of the signature. These are defined in the places where the digital signature is used in this specification.

Chapter 13. Trust Model

13.1. 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.

13.2. 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.

The credential should be listed in the COSE unprotected headers of the [COSE_Sign1_Tagged](#) structure used for digital signatures in all C2PA manifests. The credential may appear in the protected headers, though as all credential types are themselves signed objects and so carry their own integrity protection, this is not necessary. Regardless, exactly one instance of an identity credential must appear in the union of the protected and unprotected headers. [COSE_Sign1_Tagged](#) structures with no credentials, or two or more credentials, must be rejected. Repeating the same credential more than once, including separately in the protected and unprotected headers, is also an instance of two or more credentials and must be rejected.

Each credential type will define the following data to be provided to the validator:

- How the credential is stored in the header value,
- How a trust chain is computed from the signer to an entry in the validator's trust anchor store for C2PA signers,
- The public key in the credential used to validate the signature, and

- The time validity period of the credential.

The name of the header to indicate credential type, how the credential is stored in the header value, and how trust chains are constructed are specified for each credential type in [Section 13.4, “Credential Types”](#).

13.3. Signer Credential Trust

As part of the validation of an asset, the signer’s credential is validated as being trusted to sign manifests. Credential types are defined in [Section 13.4, “Credential Types”](#).

A validator shall maintain the following lists:

- A list of X.509 certificate trust anchors for C2PA signers,
- A list of accepted Extended Key Usage (EKU) values for C2PA signers, and
- A list of X.509 certificate trust anchors for Time Stamp Authorities.

In this section, "user" refers to human actors that are using C2PA-compliant validators in consumption and authoring scenarios.

A validator should allow these trust anchor stores to be configured by the user, and should provide default options or offer lists maintained by external parties that the user may opt into to populate the validator’s trust anchor store for C2PA signers or Time Stamp Authorities.

A validator may also allow the user to create and maintain a private credential store of signing credentials for each credential type. This store is intended as an "address book" of credentials they have chosen to trust based on an out-of-band relationship. For example, a journalist may choose to add sources to their private credential store to facilitate accepting and validating media with C2PA provenance data attached, even though the sources themselves would have no reason to be on an externally-maintained trust list used broadly by the general public. Credentials in the private credential store may be self-issued, and may be anonymous or pseudonymous (that is, containing no identifying information about the signer, or information that only identifies by pseudonym). The private credential store shall only apply to validating signed C2PA manifests, and shall not apply to validating time-stamps. The private credential store shall only allow trust in signer certificates directly; entries in the private credential store cannot issue credentials and shall not be included as trust anchors during validation.

A validator must not be pre-configured with any entries in a private credential store.

A validator must only add entries to a private credential store in response to a user request to trust the credential. Similarly, a validator must only remove entries from a private credential store in response to a user request to stop trusting the credential.

13.4. Credential Types

Each credential type defined in this section may be used to sign C2PA manifests.

13.4.1. X.509 Certificates

X.509 Certificates are stored as defined by the draft specification [CBOR Object Signing and Encryption \(COSE\): Header parameters for carrying and referencing X.509 certificates \(draft-ietf-cose-x509\)](#) section 2. This is a draft at the time of publication of this specification, and therefore an integer header map label has not been permanently assigned for this header. Therefore, this specification shall use the string **x5chain** as the label for this header. The definition of **x5chain** at time of this specification's publication is copied below. The definition here shall be used in this specification; further updates to the draft and eventual publication as a standard will be incorporated by a future version of this specification.

x5chain: This header parameter contains an ordered array of X.509 certificates. The certificates are to be ordered starting with the certificate containing the end-entity key followed by the certificate which signed it and so on. There is no requirement for the entire chain to be present in the element if there is reason to believe that the relying party already has, or can locate the missing certificates. This means that the relying party is still required to do path building, but that a candidate path is proposed in this header parameter.

The trust mechanism MUST process any certificates in this parameter as untrusted input. The presence of a self-signed certificate in the parameter MUST NOT cause the update of the set of trust anchors without some out-of-band confirmation. As the contents of this header parameter are untrusted input, the header parameter can be in either the protected or unprotected header bucket.

This header parameter allows for a single X.509 certificate or a chain of X.509 certificates to be carried in the message.

- * If a single certificate is conveyed, it is placed in a CBOR byte string.
- * If multiple certificates are conveyed, a CBOR array of byte strings is used, with each certificate being in its own byte string.

When creating the **x5chain** header as part of signing, the signer's certificate and all intermediate certificate authorities shall be included in the header's value. The trust anchor's certificate (also called the root certificate) should not be included.

The **subjectPublicKeyInfo** element of the first or only certificate will be the public key used to validate the signature. The **validity** element of the **tbsCertificate** sequence provides the time validity period of the certificate.

13.4.1.1. Certificate Profile

This section defines the requirements to validate that an X.509 certificate is acceptable as a signing credential as described in [Section 14.3, “Validate the Signature”](#).

All certificates must fulfil the following requirements.

- The **algorithm** field of the **signatureAlgorithm** field shall be one of the following values:

ecdsa-with-SHA256

RFC 5758 section 3.2

ecdsa-with-SHA384

RFC 5758 section 3.2

ecdsa-with-SHA512

RFC 5758 section 3.2

sha256WithRSAEncryption

RFC 8017 appendix A.2.4

sha384WithRSAEncryption

RFC 8017 appendix A.2.4

sha512WithRSAEncryption

RFC 8017 appendix A.2.4

id-RSASSA-PSS

RFC 8017 appendix A.2.3

id-Ed25519

RFC 8410 section 3

- If the `algorithm` field of the `signatureAlgorithm` field is **id-RSASSA-PSS**, the `parameters` field is of type **RSASSA-PSS-params**. Its fields shall have the following requirements: [RFC 8017 section A.2.3](#)
 - The `hashAlgorithm` field shall be present.
 - The `algorithm` field of the `hashAlgorithm` field shall be one of the following values: [RFC 8017 appendix B.1](#)
 - **id-sha256**
 - **id-sha384**
 - **id-sha512**
 - The `maskGenAlgorithm` field shall be present.
 - The `algorithm` field of the `parameters` field of the `maskGenAlgorithm` field shall be equal to the `algorithm` field of the `hashAlgorithm` field.
- If the `algorithm` field of the `algorithm` field of the certificate's `subjectPublicKeyInfo` is **id-ecPublicKey**, the `parameters` field shall be one of the following named curves: [RFC 5480 section 2.1.1.1](#)
 - **prime256v1**
 - **secp384r1**

- `secp521r1`
- If the `algorithm` field of the `algorithm` field of the certificate's `subjectPublicKeyInfo` is `rsaEncryption` or `rsaPSS`, the `modulus` field of the `parameters` field shall have a length of at least 2048 bits.

All certificates except those in the private credential store for X.509 certificates must fulfill the following additional requirements to be acceptable.

- Version must be v3. [RFC 5280 section 4.1.2.1](#)
- The `issuerUniqueID` and `subjectUniqueID` optional fields of the `TBSCertificate` sequence must not be present. [RFC 5280 section 4.1.2.8](#)
- The Basic Constraints extension must follow [RFC 5280 section 4.2.1.9](#). In particular, it must be present with the `cA` boolean asserted if the certificate issues certificates, and not asserted if it does not.
- The Authority Key Identifier extension must be present in any certificate that is not self-signed. [RFC 5280 section 4.2.1.1](#)
- The Subject Key Identifier extension must be present in any certificate that acts as a CA. It should be present in end entity certificates. [RFC 5280 section 4.2.1.2](#)
- The Key Usage extension must be present and should be marked as critical. Certificates used to sign C2PA manifests must assert the `digitalSignature` bit. The `keyCertSign` bit must only be asserted if the `cA` boolean is asserted in the Basic Constraints extension. [RFC 5280 section 4.2.1.3](#)
- The Extended Key Usage (EKU) extension must be present and non-empty in any certificate where the Basic Constraints extension is absent or the `cA` boolean is not asserted. These are commonly called "end entity" or "leaf" certificates. [RFC 5280 section 4.2.1.12](#)
 - The `anyExtendedKeyUsage` EKU (2.5.29.37.0) must not be present.
 - If the configuration store contains a list of EKUs, a certificate that signs C2PA manifests must be valid for at least one of the listed purposes.
 - If the configuration store does not contain a list of EKUs, a certificate that signs C2PA manifests must be valid for the `id-kp-emailProtection` (1.3.6.1.5.5.7.3.4) purpose.
 - The `id-kp-emailProtection` purpose is not implicitly included by default if a list of EKUs has been configured. If desired, it must explicitly be added to the list in the configuration store.
 - A certificate that signs time-stamping countersignatures must be valid for the `id-kp-timeStamping` (1.3.6.1.5.5.7.3.8) purpose.
 - A certificate that signs OCSP responses for certificates must be valid for the `id-kp-OCSPSigning` (1.3.6.1.5.5.7.3.9) purpose.
 - If a certificate is valid for either `id-kp-timeStamping` or `id-kp-OCSPSigning`, it must be valid for exactly one of those two purposes, and not valid for any other purpose.
 - A certificate should not be valid for any other purposes outside of the purposes listed above, but the presence of any EKUs not mentioned in this profile and not in the list of EKUs in the configuration store shall not cause

the certificate to be rejected.

13.4.1.1.1. Certificate Trust Chain

When validating a certificate as the signing credential, if the certificate is present in the private credential store for X.509 certificates, the certificate is accepted. The private credential store is not consulted when validating time-stamps.

If the certificate is not present in the private credential store, or the validator does not implement one, the trust chain shall be built and validated according to the procedure in [RFC 5280 section 6](#) for the particular purpose required (signing, time-stamping, or OCSP signing) and for the appropriate trust anchor store for that purpose. Any failure of that validation algorithm shall mean the chain must be rejected. The private credential store is never included when building certificate chains; certificates in the private credential store cannot act as CAs.

Only end entity certificates shall be used to sign C2PA manifests or time-stamps. A CA certificate must not be used for these purposes. Any CA certificate (where the `CA` boolean in the Basic Constraints extension is asserted) being used to validate a signature on a C2PA manifest, time-stamp, or OCSP response must be rejected.

13.4.1.2. Certificate Revocation

X.509 certificates support revocation status queries. C2PA uses the Online Certificate Status Protocol (OCSP) and OCSP stapling to implement revocation. C2PA does not use Certificate Revocation Lists (CRLs).

NOTE

Using CRLs requires downloading the entire list of revoked certificates for each Certificate Authority encountered, which can be time-consuming. Although a CRL could be included in the same way an OCSP response is stapled, the potential size of a CRL relative to an OCSP response also makes this undesirable.

A conforming CA should include an AuthorityInfoAccess (AIA) extension ([RFC 5280 section 4.2.2.1](#)) to provide access information for an Online Certificate Status Protocol (OCSP) service operated by the CA.

If the certificate has an AIA extension, revocation information shall be stored in an unprotected header of the `COSE_Sign1` structure with the string label `rVals` and the value shall follow the following CDDL:

```
; CBOR version of rVals and related structures based on JSON schema in
https://www.etsi.org/deliver/etsi_ts/119100_119199/11918201/01.01.01_60/ts_11918201v010101p.
pdf section 5.3.5.2
rVals = {
  "ocspVals": [1* bstr]
}
```

NOTE

The above definition is a CBOR adaptation of a subset of the schema from [JAdES section 5.3.5.2](#), which only stores OCSP responses, and stores them as binary strings.

Before signing a claim, if a signer's certificate has the AIA extension, a signer should query the OCSP service indicated therein, capture the response, and store it in an element of the `ocspVals` array of the `rVals` header.

Validating the Certificate Revocation Information

A validator must follow the requirements of [RFC 6960](#), in particular [section 3](#), when constructing an OCSP query and accepting an OCSP response. If the response is not accepted, or contains a `certStatus` of `unknown`, nothing can be concluded about the certificate's revocation status, and therefore:

- If the unusable response is in an `rVals` header, the validator must proceed as if the header was absent.
- If the unusable response is received in reply to an OCSP query at validation time, the validator must proceed as if it chose not to make the query.

An accepted OCSP response in the `rVals` header establishes that the signer's certificate was not revoked at the time of signing if all of the following requirements are met:

- The manifest has a valid time-stamp, and the attested time falls within the `(thisUpdate, nextUpdate)` interval of the response,
- The `certStatus` field of the response is `good`, or `revoked` but with a `revocationReason` of `removeFromCRL`, and
- The signer of the response is an "authorized responder" as defined by [RFC 6960 section 4.2.2.2](#).

NOTE

The `removeFromCRL` is unique amongst the values of `revocationReason` because it is equivalent to a `good` response. Despite being a type of `revoked` response, this response indicates the certificate had temporarily been put "on hold" (the `certificateHold` reason) previously due to some concern about its integrity, but that the concern has been resolved and the issuer is stating the certificate remains trustworthy.

Validators must check the `revocationReason` of any `revoked` response to disambiguate the `removedFromCRL` case from an actual revocation.

If the `rVals` header is not present or does not contain an accepted OCSP response, or if the manifest does not have a time-stamp, but the certificate has an AIA extension, the validator may choose to query the OCSP responder, as described in [Section 14.5, "Validate the Credential Revocation Information"](#). If it does, and the response is accepted per the requirements of RFC 6960, it shall establish the signer's certificate was not revoked at the time of signing if either of the following requirements is fulfilled:

- The manifest has a valid time-stamp, and the attested time falls within the `(thisUpdate, nextUpdate)` interval of the response, or
- The manifest does not have a valid time-stamp but the current time falls within the `(thisUpdate, nextUpdate)` interval of the response,

And both of the following requirements are fulfilled:

- The `certStatus` field of the response is `good`, or `revoked` but with a `revocationReason` of `removeFromCRL`, and

- The signer of the response is an "authorized responder" as defined by [RFC 6960 section 4.2.2.2](#).

If the `certStatus` field of the response is `revoked` but with a `revocationReason` that is not `removeFromCRL`, it shall establish the signer's certificate was not revoked at the time of signing if both of the following requirements are met:

- The manifest has a valid time-stamp, and the attested time falls within the `(thisUpdate, nextUpdate)` interval of the response, and
- The `revocationTime` in the response is after the attested time-stamp.

Otherwise, the certificate shall be considered revoked at the time of signing and the claim shall be rejected.

13.5. Identity In Assertions

Some assertions (such as `stds.schema-org.CreativeWork`) allow a person's identity to be associated in a defined way with the asset. This identity is purely scoped via the definition in each assertion and does not imply any larger involvement or responsibility for any assertion made in the claim, or the asset itself. All assertions, as stated below, are made by a signer.

13.6. Statements

A validator is a manifest consumer that will produce some `validation` statements about that asset. The actor consuming the asset, usually through their user agent and its user interface, then has to interpret those statements to arrive at a set of conclusions of their own about the provenance of the asset they are consuming. These conclusions will be drawn from these statements, the set of trust relationships that consumer currently has with the actors in the asset, and the contents of the asset itself.

A validator can make the following true or false statements about the asset they are validating, and no more.

1. The active manifest has not been modified since the active manifest was signed
2. The portions of the asset that are covered by content bindings have not been modified since the active manifest was produced
3. The claim was produced by a claim generator (typically software), and signed by an actor identified in the `subject` field of the signing credential
4. The assertions of the active manifest are statements by the signer and their contents are not verified
5. The assertions of the active manifest have not been modified since the active manifest was produced
6. The assets referred to by ingredient assertions are not (necessarily) available at validation of the active manifest, and therefore their hashes cannot be validated
7. The ingredient assertion may contain a `ClaimReview`, that indicates the active manifest signer's assessment of the validation state of the ingredient asset's hashes at the time of adding the ingredient
8. The content of ingredient assertions, like all other assertions, is not independently validated

Chapter 14. Validation

The active manifest of an asset is valid only if all the steps in this section are successful. This validation must be completed before a validator presents a successful result to a human user or begins to render any content. Validating content as it is rendered to the user is described in [Section 14.9, “Validate the Asset’s Content”](#).

14.1. Locating the Active Manifest

The last C2PA Manifest superbox in the C2PA Manifest Store superbox is the active manifest, but locating the C2PA Manifest Store may involve looking in a number possible locations.

14.1.1. Embedded

The manifest consumer shall look inside the asset for an embedded C2PA Manifest Store in the [standard locations for embedding manifests](#) to see if one is present.

If there are multiple C2PA Manifest Stores present in an asset, they shall all be considered as invalid and the validation should treat this as if no manifests were located. In the case where this asset is being added as an ingredient, none of these embedded C2PA manifests shall be included in the ingredient assertion.

14.1.2. By Reference or URI

If there is no embedded C2PA Manifest Store, the following attempts should be made to locate one at a remote location.

- If the asset was retrieved via an HTTP connection, the manifest consumer should look in the header of the HTTP response for a [Link](#) header, as defined in [RFC 8288](#), containing a parameter of [rel=c2pa-manifest](#). If present, a C2PA Manifest Store can be retrieved from that URI reference.

Note 1

NOTE HTTP refers to the *Hypertext Transfer Protocol* defined in [RFC 7230](#), not the specific URL scheme <http://>.

- If the asset has XMP and the XMP contains a [dcterms:provenance](#) key, the provided URI should be used to locate the active manifest.
- If no C2PA Manifest Store has been located, the manifest consumer should look for files at the same path or URI, but with a filename extension of [.c2pa](#). If the C2PA Manifest Store is not found, a manifest consumer may look in whatever additional places it deems most appropriate to locate it. For example, a child folder of a file system.

Note 2

NOTE A manifest consumer is not restricted to only the above locations, it can choose to look in additional locations as well.

NOTE The C2PA will work with the ISO JPEG Systems committee to register a media type for JUMBF with IANA.

14.1.2.1. Validating a Match

A manifest consumer may wish to validate that the located C2PA Manifest Store is indeed the one associated with asset.

If the C2PA Manifest Store was located then the hard binding assertions present in its active manifest shall be used to validate that it is the matching manifest and whether the asset has been modified without manifest updates. If the hard bindings do not match, it is unknown if that is because of (a) modification of the asset or (b) the wrong C2PA Manifest Store was located. Accordingly, the manifest consumer shall treat this as non-matching hard bindings and reject the manifest with a failure code of `assertion.dataHash.mismatch` if the data hash assertion is used, or `assertion.bmffHash.mismatch` if the BMFF hash assertion is used.

14.2. Locating the Claim

Once an active manifest has been located, the claim is found by locating, within the active manifest, the JUMBF Superbox with a label of `c2pa.claim` and a UUID of `0x6332636C-0011-0010-8000-00AA00389B71` (`c2cl`). There shall only be one such box in the active manifest, if more than one is located, the manifest shall be rejected with a failure code of `claim.multiple`.

14.3. Validate the Signature

Retrieve the URI reference for the signature from the value of the claim's `signature` field and resolve the URI reference to obtain the COSE signature. The signature must be embedded in the same manifest as described in [Section 10.1.1, “C2PA Box details”](#). If the signature URI does not refer to a location within the same manifest box (a `self#jumbf` location), the claim must be rejected. If no such field is present or the URI cannot be resolved, then the claim must be rejected with a failure code of `claimSignature.missing`.

Validate that the credential used in the signature is acceptable according to [Chapter 13, Trust Model](#) for the credential's type. If a chain of trust cannot be built from the credential to an entry in the trust anchor list, the claim must be rejected with a failure code of `signingCredential.untrusted`. If the credential is not acceptable per the requirements of the credential's type, then the claim must be rejected with a failure code of `signingCredential.invalid`. If the signature algorithm is not on the allowed or deprecated list in [Section 12.2, “Digital Signatures”](#), then the claim must be rejected with a failure code of `algorithm.unsupported`. After confirming the credential type and signing algorithm are acceptable, validation should proceed according to the specified procedure in [Section 12.2, “Digital Signatures”](#). If validation of the signature fails, then the claim must be rejected with a failure code of `claimSignature.mismatch`.

For the remainder of this chapter, headers refer to the union of the set of protected and unprotected header parameters in the COSE signature. Unless otherwise specified in [Section 12.2, “Digital Signatures”](#) or [Section 13.4](#),

“[Credential Types](#)”, a header may appear in either bucket. [RFC8152 section 3](#) describes COSE headers.

14.4. Validate the Time-Stamp

If the `sigTst` header is not present, the claim is valid if the current time is within the validity period of the signer’s credential. If it is not, the claim must be rejected with a failure code of `signingCredential.expired`.

If the `sigTst` header is present, the claim is valid if the `tstTokens` array contains at least one `tstToken` whose `val` property is an RFC3161-compliant `TimeStampResp` which satisfies the following requirements:

- The time-stamp contains a message imprint as described in [Section 9.3.2.5, “Time-stamps”](#) that matches the claim being validated. The claim must be rejected with a failure code of `timeStamp.mismatch` if it does not match.
- The time attested by the Time Stamp Authority (TSA) falls within the validity period of the signing credential. The claim must be rejected with a failure code of `signingCredential.expired` if it does not.
- The attested time falls within the validity period of the TSA’s signing certificate. The claim must be rejected with a failure code of `timeStamp.outsideValidity` if it does not.
- A trust chain can be built to an entry in the TSA trust store as described in [Section 13.3, “Signer Credential Trust”](#). Locating the TSA’s certificate in the `TimeStampResp` is described in [RFC 3161 section 2.4.1](#). The claim must be rejected with a failure code of `timeStamp.untrusted` if a trust chain cannot be built.

NOTE

At this time, the revocation status of a Time Stamp Authority’s certificate is neither captured at signing time nor validated at validation time.

14.5. Validate the Credential Revocation Information

If the signer’s credential type does not support revocation status, or the credential’s issuer did not provide a method to query its revocation status, the validator presumes the credential is not revoked.

If the signer’s credential type supports revocation, and the credential’s issuer provided a method to query its revocation status:

- If the `rVals` header is present, its contents shall be validated as described in [Section 13.4, “Credential Types”](#) for the signer credential type.
- If the `rVals` header is not present, and the signer’s credential is considered valid by the requirements of [Section 14.3, “Validate the Signature”](#) and [Chapter 13, Trust Model](#), a validator may choose to query the credential status method to determine if the credential is currently revoked.
 - If the validator does not query the credential status, the validator shall presume the credential is not revoked.
 - If the validator does query the credential status, it shall determine the status from the response as described in [Section 13.4, “Credential Types”](#) for the signer’s credential type.

NOTE

Querying the credential status method can reveal to an observer the identity of the asset being validated, and so this query is optional.

NOTE

When a signer's credential is revoked, this does not invalidate manifests that were signed before the time of revocation. The inclusion of the `rVals` header combined with a time-stamp provides proof that the signer's credential was valid at the time of signing. Signers are encouraged to include revocation information and time-stamps to avoid the necessity of a query at validation time.

In all cases, if the credential is deemed revoked at the time of signing, the claim shall be rejected with a failure code of `signingCredential.revoked`.

14.6. Validate the Assertions

14.6.1. Validate the correct assertions for the type of manifest

Depending on the `type of manifest`, there are assertions that are either required or forbidden. A validator shall check for required and not-permitted assertions.

1. If it is a `standard manifest`

- a. Validate that there is at least one `hard binding to content` assertion - either a `c2pa.hash.data` or a `c2pa.hash.bmff` based on the type of asset for which the manifest is destined. If not, the manifest must be rejected with a failure code of `claim.hardBindings.missing`.
- b. Validate that there are zero or one `c2pa.ingredient` assertions whose `relationship` is `parentOf`. If there is more than one, the manifest must be rejected with a failure code of `manifest.multipleParents`.

2. If it is an `update manifest`

- a. Validate that there are not any `c2pa.hash.data`, `c2pa.hash.bmff` or `c2pa.actions` assertions. If there are, the manifest must be rejected with a failure code of `manifest.update.invalid`.
- b. Validate that there is exactly one `c2pa.ingredient` assertion and whose `relationship` is `parentOf`. If there is not (i.e., either it is missing, there are more than one, or the value of `relationship` is not `parentOf`), the manifest must be rejected with a failure code of `manifest.update.wrongParents`.

14.6.2. Preparing the list of redacted assertions

For each manifest, there may be a set of its assertions that were redacted from it. The list of those assertions is not found in the manifest itself but instead in a manifest that references it as an ingredient. Therefore a validator, when processing a claim, shall gather the set of redacted assertion for each ingredient manifest based on each `hashed_uri` listed in the `redacted_assertions` field. A claim's `redacted_assertions` field shall never include a `hashed_uri` to any of its own assertions.

14.6.3. Assertion Validation

Each assertion in the `assertions` field of the claim is a `hashed_uri` structure. For each assertion, the validator must:

1. If the URI reference in the `url` field is in the [list of redacted assertions](#):
 - a. If the assertion's label is `c2pa.actions`, the claim must be rejected with a failure code of `assertion.action.redacted` as `c2pa.actions` assertions shall not be redacted.
 - b. Otherwise, the redacted assertion is considered valid, and validation continues to the next assertion.
2. For all other assertions:
 - a. Resolve the URI reference in the `url` field to obtain its data. If the URI does not refer to a location within the same manifest box (a `self#jumbf` location), the claim must be rejected. If the URI cannot be resolved and the data retrieved, the claim must be rejected with a failure code of `assertion.missing`.
 - i. If the assertion's label is `c2pa.cloud-data`:
 - A. If the `label` field of the external assertion is `c2pa.hash.data` or `c2pa.hash.bmff`, the claim must be rejected with a failure code of `assertion.cloud-data.hardBinding`.
 - B. If the manifest is an update manifest and the `label` field of the external assertion is `c2pa.actions`, the claim must be rejected `assertion.cloud-data.actions`.
 - ii. Compute a hash of the data using the hash algorithm, determined by following the procedure described in [Section 12.1, “Hashing”](#):
 - A. If an `alg` field is present in the `hashed_uri` structure, that determines the hash algorithm.
 - B. If an `alg` field is not present in the `hashed_uri` structure, an `alg` field must be present in an enclosing structure, and the nearest instance present determines the hash algorithm.
 - C. If no `alg` field is found in any of these locations:
 - I. If an `alg` field is present in the claim, that determines the hash algorithm.
 - II. If no `alg` field is present in the claim, the claim must be rejected with a failure code of `assertion.hashedURI.mismatch`.
 - iii. Locate the value of the `alg` field in the allowed list or the deprecated list in [Section 12.1, “Hashing”](#) to determine the hash algorithm. If it is not present in either list, the claim must be rejected with a failure code of `algorithm.unsupported`.
 - iv. Compare the computed hash value with the value in the `hash` field. If they do not match, the claim must be rejected with a failure code of `assertion.hashedURI.mismatch`.
 - v. Otherwise, the assertion is valid and validation continues to the next assertion.

Then, for each element of the claim's own `redacted_assertions` array, if any element of the claim's `assertions` array has a `url` field equal to that value, the claim must be rejected with a failure code of `assertion.selfRedacted`. A claim cannot redact its own assertions, only those of its ingredients.

NOTE As described in [Chapter 7, W3C Verifiable Credentials](#), any proofs present inside a Verifiable Credential are not validated. Like all contents of an assertion, C2PA only guarantees the contents of the credential are integrity-protected.

14.6.4. External Data Validation

The contents of a [cloud data assertion](#), which contains the URI references to and hashes of external data, are validated like any other assertion, but those references are not retrieved and validated as part of standard validation. A validator must first successfully validate a claim before attempting to retrieve external data listed in a cloud data assertion. A validator must not attempt to retrieve external data from a rejected claim. As the retrieval of external data is optional, the inability to retrieve or validate external data shall not cause a claim to become rejected.

If a validator chooses to retrieve any of the external data in a cloud data assertion, the validator must:

1. First, determine the hash algorithm to be used.
 - a. If an `alg` field is present in the `hashed_ext_uri` structure, that determines the hash algorithm. If an `alg` field is not present, the validator must abort the attempt to retrieve the external data.
 - b. Locate the value of the `alg` field in the allowed list or the deprecated list in [Section 12.1, “Hashing”](#) to determine the hash algorithm. If it is not present in either list, the validator must abort the attempt to retrieve the external data.

NOTE The `alg` field is mandatory in `hashed_ext_uri`, so no recursive procedure to determine the hash algorithm is required.

2. Resolve the URI reference in the `url` field to obtain its data. If the URI cannot be resolved and the data retrieved, the validator must abort the attempt to retrieve the external data.
3. If the size of the retrieved data is not equal to the value of the `size` field, the validator must return a failure code of `assertion.hashedURI.mismatch` to the application and not provide the retrieved data.
4. If the content type of the retrieved data, as determined by the `Content-Type` header of the HTTP response, is not equal to the value of the `content_type` field, the validator must return a failure code of `assertion.hashedURI.mismatch` to the application and not provide the retrieved data.
5. Compute a hash of the data using the hash algorithm, determined by following the procedure described in [Section 12.1, “Hashing”](#):
 - a. Compare the computed hash value with the value in the `hash` field. If they do not match, the validator must return a failure code of `assertion.hashedURI.mismatch` to the application and not provide the retrieved data.
 - b. Otherwise, the retrieved data is successfully provided to the application.

14.7. Recursively Validating Integrity of Ingredients

A validator must perform the above validation steps for the asset being presented and its manifest. If any of the above steps conclude the manifest is invalid, that manifest must be rejected with the indicated failure code.

An asset's manifest may list one or more ingredients. For standard manifests, a validator may choose to optionally recurse through any ingredient manifests, however the `parentOf` ingredient of an update manifest shall be validated by the procedure below. There is no requirement that signers of ingredient manifests are trusted by the validator, and building of trust chains of signers of ingredients shall not be attempted. Instead, as the ingredient is included by the signer of the active manifest, and if the signer of the active manifest is accepted per the rules above, ingredient manifests will share in that trust for the purposes of this recursive validation. Applications should not display data from ingredient manifests with failed integrity checks. If the application chooses to display such data, it must flag the display with a warning about the failed integrity check, and that the data cannot be reliably attributed to the ingredient manifest's signer nor to the asset's manifest's signer.

When ingredients are being added to an asset as part of an authoring workflow, the ingredient may undergo full validation and the results expressed in a [validation status](#). In this scenario, the ingredient's manifest is considered the active manifest for validation purposes, before the ingredient is added to another asset's ingredients.

For consumption scenarios, it is expected that problems with ingredient manifests from a standard manifest would be ignored during normal consumption use but may be surfaced as a warning if a user opts to explore the provenance history. However, the `parentOf` ingredient of an update manifest shall be surfaced.

In authoring scenarios, it may be desirable to more prominently raise warnings so that a creator making use of such an asset with a flawed provenance history can make an informed decision of how to proceed.

If the manifest is an update manifest, or if a validator chooses to validate the ingredients of a standard manifest, then for each ingredient, it must recursively:

- If the ingredient does not have a `c2pa_manifest` field
 1. If the manifest is a standard manifest, then the ingredient is accepted.
 2. If the manifest is an update manifest, then the claim must be rejected with a failure code of `manifest.update.wrongParents`.
- If the ingredient does have a `c2pa_manifest` field:
 1. Create an [list of redacted assertions](#) which is the concatenation of the claim's `redacted_assertions` array with any pre-existing redacted assertions list from previous recursive calls. This list is therefore only the redacted assertions from the active manifest's claim and any ingredients along the path to the current ingredient.
 2. Resolve the URI reference in the `url` field to obtain the ingredient claim's manifest. If the URI reference cannot be resolved, the ingredient claim is rejected with a failure code of `claim.missing`.
 3. Determine the hash algorithm identifier as determined by following the procedure described in [Section 12.1, “Hashing”](#):

- a. If an `alg` field is present in the `hashed_uri` structure, that determines the hash algorithm.
 - b. If an `alg` field is not present in the `hashed_uri` structure, an `alg` field must be present in an enclosing structure, and the nearest instance present determines the hash algorithm.
 - c. If no `alg` field is found in any of these locations, the claim must be rejected with a failure code of `ingredient.hashedURI.mismatch`.
4. Compute the hash of the ingredient manifest's data using that algorithm.
 5. Compare the computed hash with the value in the `hash` field. If the hashes are not equal, the claim must be rejected with a failure code of `ingredient.hashedURI.mismatch`.
 6. If the ingredient contains a `validationStatus` field, and its `code` field equals a failure code, as defined at [Section 17.11.6.2, “Failure codes”](#), the ingredient's claim is admitted. An admitted manifest is treated like an accepted manifest, but with an explicit indication that it may contain validation errors that are known to the signer. If admitted in this way, the validator must present the contents of the `validationStatus` field as part of any exploration of the provenance history. Validators should perform full validation if exploration of the provenance history of the ingredient is requested to indicate where there are validation errors.

NOTE

The presence of a `validationStatus` with a failure code is an explicit statement by the signer that they acknowledge and override validation errors in the ingredient's claim itself.

7. Otherwise, validate the ingredient claim and assertions as described beginning in [Section 14.3, “Validate the Signature”](#), except skip establishing signer credential trust, as this is not applicable to ingredients.
 - a. When validating the assertions as described in [Section 14.6, “Validate the Assertions”](#), provide a [list of redacted assertions](#) as an input.
 - b. The validator may optionally recursively validate the ingredient's ingredients. If it does and those are accepted, the ingredient's claim is accepted. If any are rejected, the ingredient's claim is rejected. If the validator chooses not to recurse further, the ingredient's claim is accepted.

14.8. Visual look of Validation

Here is a visual representation of the process of validating a claim (and its assertions).

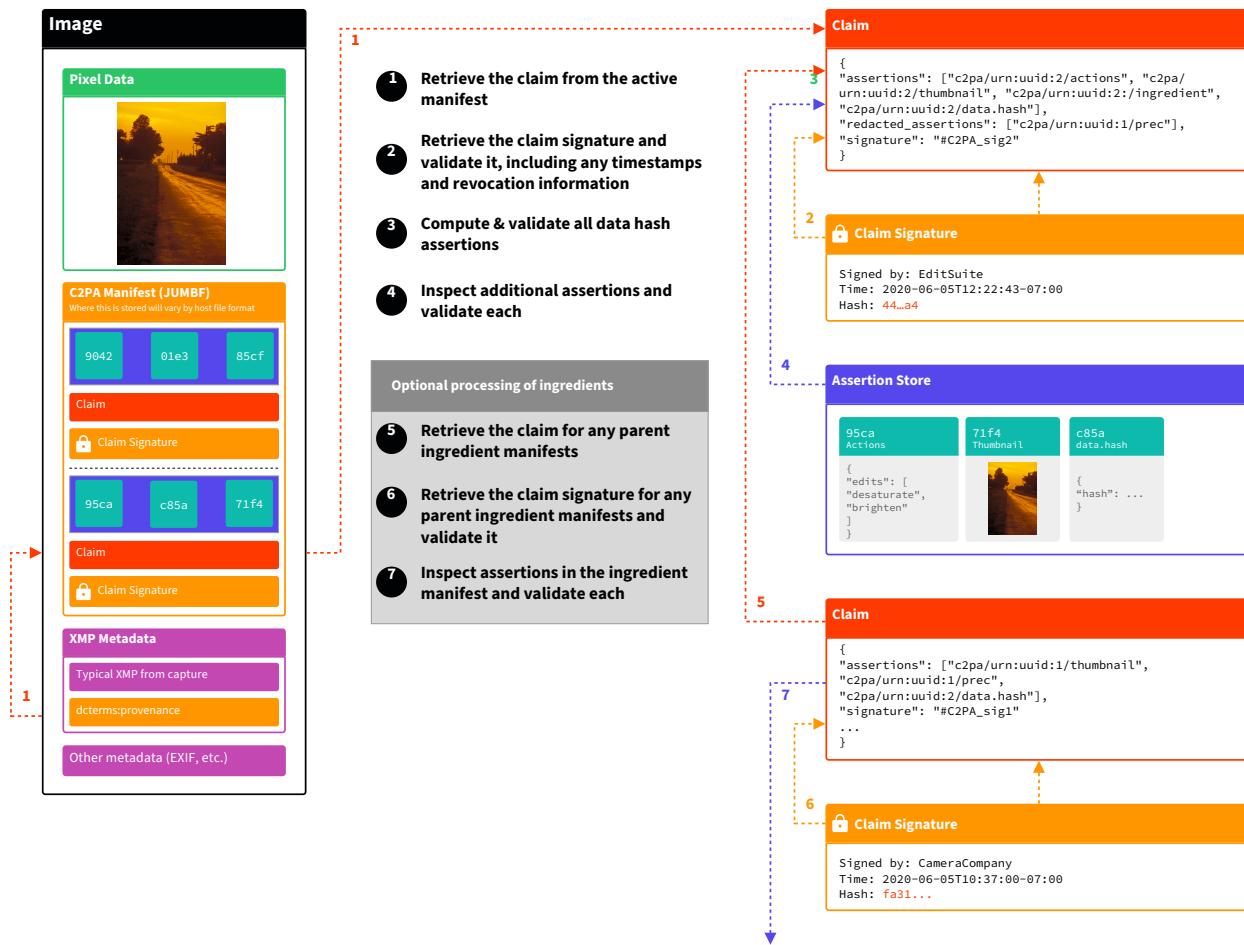


Figure 12. Validating a Claim

14.9. Validate the Asset's Content

If the active manifest is an update manifest, its [Section 8.2, “Hard Bindings”](#) are inherited from the `parentOf` ingredient’s manifest. If that manifest is also an update manifest, the search for a standard manifest shall recurse through the chain of ingredients. If no standard manifest is found, then the manifest shall be rejected with a failure code of `manifest.update.wrongParents`.

14.9.1. Validating a data hash

Once a standard manifest (and its bindings) has been located, the exclusion range(s) shall be extracted from the `c2pa.hash.data` assertion.

If any update manifests were encountered then the `length` value of the exclusion range whose `start` value is the offset of the start of the entire C2PA manifest store shall be treated as the current length of the entire C2PA manifest store plus any file format specific extras.

The hash algorithm (`alg`) specified in that `c2pa.hash.data` shall be computed over the bytes of the asset, minus those specified in the exclusion range(s).

If the resultant hash does not match the value of the `hash` field in the `c2pa.hash.data`, then the manifest shall be rejected with a failure code of `assertion.dataHash.mismatch`.

14.9.1.1. Hashing of JPEG-1 files

In JPEG-1 files, the file format extras described above would include any `APP11` markers and their respective segment length bytes for `APP11` segments. Because the segment lengths are inside the exclusion range, a validator shall match the total length of the exclusion range with that of the total length of all `APP11` segments representing the C2PA Manifest to ensure that the length was not tampered with.

NOTE

A JPEG-1 file can contain `APP11` segments for reasons other than C2PA (e.g., JPEG 360 or JPSec) and those are not included in these calculations.

14.9.2. Validating a BMFF-hash

For any portions of an asset rendered for presentation to a user, including but not limited to audio, video, or text, the corresponding hard bindings corresponding to the rendered content must be validated in accordance with [Section 8.2, “Hard Bindings”](#). If at any time content fails to be validated, the validator must clearly signal to the user that some of the content does not match the claim, and if possible, should indicate what part of the content did not validate. If any content is absent for which content bindings exist, discovery of this absence is also a validation failure. The validator must continue to report validation has failed, even if later portions of the content validate correctly.

For content that is not wholly available before rendering begins, such as during adaptive bitrate streaming (ABR) and progressive download, absence of not-yet-available portions of content is not considered a validation failure. As the content becomes available, the validator must validate each portion of the content before it is rendered as previously described. In addition, the validator must validate that the sequence of said content is the same as when the manifest was produced. Unless the player has explicitly signalled the validator that a discontinuity is expected (e.g., when the consumer performs a manual seek operation via the UI), the validator must clearly signal to the user that an unexpected discontinuity has occurred whenever the sequence does not match.

For content that is intentionally not being rendered as the claim generator originally intended, such as during fast-forward, rewind, or playback at a different speed, the validator may not be able to validate the content. In this case, the validator must clearly signal to the user that the content cannot be validated during the corresponding operation.

The `assertion.bmffHash.mismatch` failure code is used for all failures described in this section.

Chapter 15. User Experience

15.1. Approach

The C2PA intends to provide clear recommendations and guidance for implementers of provenance-enabled user experiences (UX). Developing these recommendations is an ongoing process that involves diverse stakeholders, with the results balancing uniformity and familiarity with utility and flexibility for users across contexts, platforms, and devices. These recommendations will be provided as a companion document.

15.2. Principles

The UX recommendations aim to define best practices for presenting C2PA provenance to consumers. The recommendations strive to describe standard, readily recognizable experiences that:

- provide asset creators a means to capture information and history about the content they are creating, and
- provide asset consumers information and history about the content they are experiencing, thereby empowering them to understand where it came from and decide how much to trust it.

User interfaces designed for the consumption of C2PA provenance must be informed by the context of the asset. We have studied 4 primary user groups and a collection of contexts in which C2PA assets are encountered. These user groups have been defined in the [C2PA Guiding Principles](#) as Consumers, Creators, Publishers and Verifiers (or Investigators). To serve the needs of each of these groups across common contexts, exemplary user interfaces are presented for many common cases. These are recommendations, not mandates, and we expect best practices to evolve.

15.3. Disclosure Levels

Because the complete set of C2PA data for a given asset can be overwhelming to a user, we describe 4 levels of progressive disclosure which guide the designs:

- Level 1: An indication that C2PA data is present and its cryptographic validation status.
- Level 2: A summary of C2PA data available for a given asset. Should provide enough information for the particular content, user, and context to allow the consumer to understand to a sufficient degree how the asset came to its current state.
- Level 3: A detailed display of all relevant provenance data. Note that the relevance of certain items over others is contextual and determined by the UX implementer.
- Level 4: For sophisticated, forensic investigatory usage, a standalone tool capable of revealing all the granular detail of signatures and trust signals is recommended.

15.4. Public Review, Feedback and Evolution

The team authoring the UX recommendations is cognizant of its limitations and potential biases, recognizing that feedback, review, user testing and ongoing evolution is a key requirement for success. The recommendations will therefore be an evolving document, informed by real world experiences deploying C2PA UX across a wide variety of applications and scenarios.

Chapter 16. Information security

16.1. Threats and Security Considerations

This section provides a summary of information security considerations and processes for technology described in the C2PA core specification. More detailed content will be provided in future releases of C2PA material including the Guidance document.

16.1.1. Context

Information security is a principal concern of C2PA. C2PA maintains a threat model and security considerations for the C2PA specification. This effort complements other security-related work within C2PA. Associated documentation is currently in development. A draft version of this content is expected to be included in a future public release of C2PA material, and will be referenced here when available.

The C2PA is developing security considerations documentation that includes:

- A summary of relevant security features of C2PA technology
- Security considerations for practical use of C2PA technology
- Threats to C2PA technology and respective treatment of those threats, including countermeasures

16.1.2. Threat modelling process overview

The C2PA builds security into our designs as they are being developed, but also expects that security design and threat modelling will continue as the system, ecosystem, and threat landscape evolve.

To this end, the C2PA uses a focused threat modelling process to support development of a strong security and privacy design. Outcomes of the effort directly support development of explicit threats and security considerations documentation, but also facilitate security thinking throughout the design process.

The threat modelling process combines synchronous (live) threat modelling sessions with focused groups of SMEs with asynchronous development of content. The number of attendees in each synchronous session is kept small to promote efficient discussions, but all members of the C2PA have the opportunity to participate via either modality.

Like other security activities, we expect our threat modelling process to evolve with the C2PA ecosystem. Process documentation is considered a guide rather than a strict directive on how threat modelling works within the C2PA.

16.1.2.1. References

A variety of references and experiences are used to inform threat modelling and related security activities for the C2PA. This section provides a subset of public documents for reference.

- [IETF on security considerations](#)

- IETF on privacy considerations (guidelines)
- W3C security and privacy self-review questionnaire
- OAuth2 threat model (example)
- Threat modelling: Designing for Security
- OWASP Threat modelling
- Microsoft Threat modelling

16.2. Harms, Misuse, and Abuse

16.2.1. Introduction

The C2PA is built on a set of [Guiding Principles](#) designed to prevent harms, threats to human rights, or disproportionate risks to vulnerable groups globally.

To ensure that the C2PA is meeting those aspects of their principles, an ongoing harms, misuse, and abuse assessment aims to identify any such abuses. In addition, the specification is being reviewed to:

- Anticipate and mitigate potential abuse and misuse;
- Address common privacy concerns of its users; and
- Consider the needs of users and stakeholders throughout the world.

The following subsections describe the framework and the process carried out to date, followed by the methodology, an overview of the assessment and an outline for public review and feedback. Due diligence actions will be developed parallel to the public review and feedback process.

16.2.2. Considerations

The harms, misuse, and abuse assessment is an ongoing process. The information presented below should not be considered the end result of a comprehensive evaluation, but a basis for broader, ongoing, and more profound discussions, that are centred on impacted communities, that could lead towards the mitigation of potential abuse and misuse and the protection of human rights.

There are two critical aspects of the approach:

Ongoing

The harms, misuse, and abuse assessment and identification of guidance, potential mitigations or compliance areas, system-change requirements and no-go areas accompanies the design, development, implementation and use stages of the C2PA standard.

Multi-disciplinary and diverse

The harms, misuse, and abuse assessment is a collaborative effort that includes multi-disciplinary experts and a

broad range of stakeholders with lived, practical and technical experience of the issues and from diverse geographical locations, cultural backgrounds and individual identities.

16.2.3. Assessment

Harms modelling focuses on analysing how a socio-technical system might negatively impact users and other stakeholders. The process of harms modelling systematically requires combining knowledge about a system architecture and its user affordances, with historical and contextual evidence about the impact of similar existing systems on different social groups. This combined information frames the ability to comprehensively anticipate harm.

The details of the process used by C2PA to model harms, misuse and abuse can be found in our [Harms Modelling documentation](#).

Chapter 17. C2PA Standard Assertions

17.1. Introduction

This section of the document lists the standard set of assertions for use by C2PA implementations, describing their syntax, usage, etc. To keep things simple, all example JUMBF URLs have been shortened for illustrative purposes - full URLs are necessary in the actual data.

All C2PA standardized assertions use the JSON JUMBF content type, the CBOR JUMBF content type, or the Embedded File content type from ISO 19566-5:AMD-1. Entity-specific assertions can be any of those, any of the other JUMBF content types from ISO 19566-5, B.1 (such as XML) or may create its own (as per the instructions in ISO 19566-5, Table B.1). The Codestream content type shall not be used for a C2PA assertion.

NOTE CBOR is not currently defined in 19566-5, but an upcoming update will define the type as `cbor`.

Unless otherwise mentioned, all assertions documented in this standard set of assertions shall be serialized as CBOR. For all assertions are of type CBOR, their schemas shall be defined using [CDDL](#). For those defined using JSON, their schemas shall be defined using [JSON Schema](#).

All assertions shall have a label as described in [Section 5.2, “Labels”](#) and shall be versioned as described in [Section 5.3, “Versioning”](#).

17.2. Use of CBOR

All CBOR encoded data in a C2PA manifest shall comply with the "Core Deterministic Encoding Requirements" of [CBOR](#).

When converting JSON data, based on one of the JSON Schemas documented here, to CBOR for the purposes of including as a valid C2PA assertion, the following shall be done:

- A JSON `null` shall be serialized as Null (major type 7, additional information 22).
- A JSON `false` shall be serialized as False (major type 7, additional information 20).
- A JSON `true` shall be serialized as True (major type 7, additional information 21).
- A JSON number shall be serialized either as an integer (major type 0 or 1) or a floating-point value (major type 7, additional information 25 through 27).
- A JSON string that does not have an alternate semantic type (e.g., date-time or uri) shall either be serialized as a fixed-length UTF-8 string (major type 3) without any null termination, if human consumable, or a byte string (major type 2) if not. For Base64-encoded strings, unless specified otherwise, they shall be encoded in CBOR as a byte string (major type 2) after decoding from Base64.
- A JSON string that is identified as a `date-time` shall be serialized with `tag 0` (major type 6)

- A JSON string that is identified as a **URI** shall be serialized with **tag 32** (major type 6)
- A JSON array shall be serialized as an array (major type 4).
- A JSON object shall be serialized as a map (major type 5), with each key being serialized as a UTF-8 string (major type 3).
 - If a JSON object contains a key that is defined (by a specific schema) as optional, then either the key with an associated value shall be included in the map or neither its key nor any value shall be included in the map.
 - Indefinite-length items (additional information 31) shall not be used. Accordingly, break codes (major type 7 and additional information value 31) shall not be used.
- A JSON **null**, a JSON array of length zero, nor a JSON string of length zero shall be used unless they are (a) the value of a required (by a specific schema) key in a JSON object or (b) there are explicitly defined semantics (by a specific schema) for their usage.

NOTE

The above list is based on an inversion of [section 6.1 of CBOR](#) with additions from C2PA best practices.

17.3. Metadata About Assertions

In many cases, it is useful or even necessary to provide additional information about an assertion, such as the date and time when it was generated or other data that may help manifest consumers to make informed decisions about the provenance or veracity of the assertion data. Below shows the core schemas used inside other assertions.

In CDDL:

```
;Describes additional information about an assertion, including a hashed-uri reference to
it. We use a socket/plug here to allow hashed-uri-map to be used in individual files without
having the map defined in the same file
$assertion-metadata-map /= {
  ? "reviewRating": [1* rating-map], ; Ratings given to the assertion (may be empty)
  ? "dateTime": tdate, ; The ISO 8601 date-time string when the assertion was
created/generated
  ? "reference": $hashed-uri-map, ;hashed_uri reference to another assertion that this
review is about
  * $$assertion-metadata-map-extension
}

$source-type /= "signer"
$source-type /= "claimGenerator.REE"
$source-type /= "claimGenerator.TEE"
$source-type /= "localProvider.REE"
$source-type /= "localProvider.TEE"
$source-type /= "remoteProvider.1stParty"
$source-type /= "remoteProvider.3rdParty"
$source-type /= "humanEntry.anonymous"
$source-type /= "humanEntry.identified`"

source-map = {
  "type": $source-type, ; A value from among the enumerated list indicating whether the
source of the assertion is a claim generator running in a rich execution environment (REE),
```

```

a claim generator running in a trusted execution environment (TEE), a local data provider in
REE (e.g. the location API from a mobile operating system), a local data running in a TEE
(e.g. a trusted location trusted app from a chipset vendor), a remote data provider such as
a server (e.g. Google's geolocation API service), entry by a human who wishes to remain
anonymous, or a human who is credentialled with a W3C Verifiable Credential that's included
in the asset.
    ? "details": tstr .size (1..max-tstr-length), ; A human readable string giving details
about the source of the assertion data, e.g. the URL of the remote server that provided the
data
    ? "actors" : [1* actor-map] ; array of hashed_uri references to W3C Verifiable Credentials
}

actor-map =
    ? "identifier": tstr .size (1..max-tstr-length), ; An identifier for a human actor, used
when the "type" is humanEntry.identified
    ? "credentials" : [1* $hashed-uri-map / $hashed-ext-uri-map] ; array of hashed_uri
references to W3C Verifiable Credentials
}

int-range = 1..5

$review-code /=
    "actions.unknownActionsPerformed"
$review-code /=
    "actions.missing"
$review-code /=
    "actions.possiblyMissing"
$review-code /=
    "depthMap.sceneMismatch"
$review-code /=
    "ingredient.modified"
$review-code /=
    "ingredient.possiblyModified"
$review-code /=
    "thumbnail.primaryMismatch"
$review-code /=
    "stds.iptc.location.inaccurate"
$review-code /=
    "stds.schema-org.CreativeWork.misattributed"
$review-code /=
    "stds.schema-org.CreativeWork.missingAttribution"

rating-map =
    "value": int-range, ; "A value from 1 (worst) to 5 (best) of the rating of the item"
    ? "code": $review-code, ; A label-formatted string that describes the reason for the
rating
    ? "explanation": tstr .size (1..max-tstr-length), ; A human readable string explaining why
the rating is what it is
}

```

In JSON Schema:

```
{
    "$schema": "http://json-schema.org/draft-07/schema",
    "$id": "http://ns.c2pa.org/assertion-metadata/v1",
    "type": "object",
    "description": "Assertion that describes additional information about an assertion,
including a hashed-uri reference to it",
    "examples": [
        {
            "dataSource": {
                "type": "localProvider.REE",
                "details": "Dilip's Photo Editor for Windows v5.6"
            },
            "reviewRatings": [
                {
                    "value": 2,
                    "code": "actions.unknownActionsPerformed",
                    "explanation": "A 3rd party filter was used."
                }
            ]
        }
    ]
}
```

```

        ],
        "dateTime": "2021-06-28T16:34:11.457Z"
    },
    {
        "reference": {
            "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-329BF39FA1E4/c2pa.assertions/stds.schema-org.CreativeWork",
            "alg": "sha256",
            "hash": "ho0spQQ1lFTy/4Tp8Epx670E5QW5NwkNR+2b30KFXug="
        },
        "reviewRatings": [
            {
                "value": 1,
                "code": "stds.schema-org.CreativeWork.missingAttribution",
                "explanation": "Producer Thomas Chu was not credited."
            },
            {
                "value": 1,
                "code": "stds.schema-org.CreativeWork.misattributed",
                "explanation": "Editor of asset is Aswhini Viswanathan, not Stacey Higgs."
            }
        ]
    },
    "definitions": {
        "ACTOR": {
            "type": "object",
            "properties": {
                "credentials": {
                    "type": "array",
                    "description": "An array of hashed uris to W3C Verifiable Credentials",
                    "minItems": 1,
                    "items": {
                        "oneOf": [
                            {
                                "$ref": "http://ns.c2pa.org/hashed-uri/v1"
                            },
                            {
                                "$ref": "http://ns.c2pa.org/hashed-ext-uri/v1"
                            }
                        ],
                        "description": "hashed-uri reference to a W3C Verifiable Credential (VC) associated with the person or organization who entered the assertion content."
                    }
                }
            }
        },
        "identifier": {
            "type": "string",
            "description": "An identifier for a human actor, used when the 'type' is humanEntry.identified",
            "minLength": 1
        },
        "required": ["credentials"],
        "additionalProperties": false
    },
    "DATASOURCE": {
        "type": "object",
        "properties": {
            "type": {
                "enum": [
                    "signer",
                    "claimGenerator.REE",
                    "claimGenerator.TEE",
                    "claimGenerator.HSM"
                ]
            }
        }
    }
}

```

```

        "localProvider.REE",
        "localProvider.TEE",
        "remoteProvider.1stParty",
        "remoteProvider.3rdParty",
        "humanEntry.anonymous",
        "humanEntry.identified"
    ],
    "description": "A value from among the enumerated list indicating whether the source of the assertion is a claim generator running in a rich execution environment (REE), a claim generator running in a trusted execution environment (TEE), a local data provider in REE (e.g. the location API from a mobile operating system), a local data running in a TEE (e.g. a trusted location trusted app from a chipset vendor), a remote data provider such as a server (e.g. Google's geolocation API service), entry by a human who wishes to remain anonymous, or a human who is credentialled with a Verifiable Credential that's included in the asset."
},
"details": {
    "type": "string",
    "description": "A human readable string giving details about the source of the assertion data, e.g. the URL of the remote server that provided the data"
},
"actors": {
    "type": "array",
    "description": "An array of the actors that undertook this action.",
    "minItems": 1,
    "items": {
        "type": "string",
        "minLength": 1,
        "$ref": "#/definitions/ACTOR",
        "description": "list of actors"
    }
},
"anyOf": [
{
    "not": {
        "properties": {
            "type": {
                "const": "humanEntry.credentialled"
            }
        },
        "required": ["type"]
    }
},
{
    "required": ["actors"]
}
],
"required": ["type"],
"additionalProperties": false
},
"RATING": {
    "type": "object",
    "properties": {
        "value": {
            "type": "integer",
            "minimum": 1,
            "maximum": 5,
            "description": "A value from 1 (worst) to 5 (best) of the rating of the item"
        },
        "code": {
            "enum": [
                "actions.unknownActionsPerformed",

```

```

        "actions.missing",
        "actions.possiblyMissing",
        "depthMap.sceneMismatch",
        "ingredient.modified",
        "ingredient.possiblyModified",
        "thumbnail.primaryMismatch",
        "stds.iptc.location.inaccurate",
        "stds.schema-org.CreativeWork.misattributed",
        "stds.schema-org.CreativeWork.missingAttribution"
    ],
    "description": "A label-formatted string that describes the reason for the rating"
},
"explanation": {
    "type": "string",
    "minLength": 1,
    "description": "A human readable string explaining why the rating is what it is"
}
},
"required": ["value"]
},
"properties": {
    "dataSource": {
        "$ref": "#/definitions/DATASOURCE",
        "description": "A description of the source of the assertion data, selected from a predefined list"
    },
    "reviewRatings": {
        "type": "array",
        "description": "An array of review ratings",
        "minItems": 1,
        "items": {
            "$ref": "#/definitions/RATING",
            "description": "The rating given to the assertion"
        }
    },
    "dateTime": {
        "type": "string",
        "minLength": 1,
        "format": "date-time",
        "description": "The ISO 8601 date-time string when the assertion was created/generated"
    },
    "reference": {
        "$ref": "http://ns.c2pa.org/hashed-uri/v1",
        "description": "hashed_uri reference to another assertion that this review is about"
    }
},
"required": [],
"additionalProperties": true
}

```

In most cases, this assertion specific metadata will appear directly inside of other assertions (e.g., ingredients) as the value of their `metadata` field. However, assertions not in JSON or CBOR, such as thumbnails, will need their information stored as a separate, independent `assertion metadata` assertion. The label for the `assertion metadata` assertion is `c2pa.assertion.metadata`.

17.3.1. Details

17.3.1.1. Data Source

This `dataSource` field is an optional field that allows the claim signer to inform downstream manifest consumers about the source from which the assertion contents originated. If no `dataSource` is provided for a given assertion, the `dataSource` is considered to be the `Signer`.

NOTE

By default, all assertions are sourced to the Signer, as the Trust Model is rooted in trust of the Signer. Where a different source is indicated, it will be a useful Trust Signal to a manifest consumer.

The value of the field is a `dataSource` object that is composed of three fields: `type`, `details`, and if applicable, `credential`.

The `dataSource type` field defines the type of the `dataSource`. It is assembled with labels in the format described in [Section 5.2, “Labels”](#). The value can be one of the following specification-defined values, or entity-specific labels can be used as an extension mechanism.

Value of <code>type</code>	Meaning
<code>signer</code>	The assertion contents came from the Signer
<code>claimGenerator.REE</code>	Assertion contents came from a claim generator running in a rich execution environment (REE), such as a desktop or mobile operating system
<code>claimGenerator.TEE</code>	Assertion contents came from a claim generator running in a trusted execution environment (TEE), such as a trusted OS
<code>localProvider.REE</code>	Assertion contents came from a data source running in an REE on the same physical computing device as the claim generator
<code>localProvider.TEE</code>	Assertion contents came from a data source running in a TEE on the same physical computing device as the claim generator
<code>remoteProvider</code>	Assertion contents came from a remote data source controlled by the signer or claim generator vendor
<code>remoteProvider.external</code>	Assertion contents came from an external, remote data source that is not the signer or claim generator vendor
<code>humanEntry.anonymous</code>	Assertion contents were entered by a human that wishes to remain anonymous
<code>humanEntry.identified</code>	Assertion contents were entered by an identified human actor(s), specified in the <code>actors</code> field of the <code>dataSource</code> .

The `details` field is a human-readable string that provides additional information about the `dataSource`, e.g. the name of the API used to provide the assertion contents, or the URL of the server from which the contents were provided. For example, a broad location assertion source may have a `type` value of `remoteProvider.3rdParty`, with the `details` value set to www.googleapis.com/geolocation/v1/geolocate.

If the value of the `type` field is `humanEntry.identified`, then an `actors` field shall contain an array of one or more actor objects as defined in [Section 17.16, “Common Data Model: Actor”](#).

17.3.1.2. Review Ratings

When present, the `reviewRatings` array provides a place for the claim generator to provide one or more `rating` objects on the quality (or lack thereof) of an assertion. A `reviewRatings` shall not be present if a `dataSource` object is present with a `type` field whose value is either `humanEntry.anonymous` or `humanEntry. credentialized`.

The `value` field of the `rating` object shall be present with any integer value from 1 (worst) through 5 (best). If present, the `explanation` field shall contain a human-consumable string description of the type of rating. In addition, an optional machine-readable `code` field which defines assertion-specific evaluation outcome codes may be provided. The value of the `code` field is assembled with labels in the format described in [Section 5.2, “Labels”](#). The value can be one of the following specification-defined values, or entity-specific labels can be used as an extension mechanism.

Value of code	Applicable Assertion	Meaning
<code>actions.unknownActionsPerformed</code>	<code>c2pa.actions</code>	The <code>actions</code> assertion does not contain a full list of all actions performed in the authoring tool (e.g. because of the use of a 3rd party filter whose effect is unknown to the authoring tool)
<code>actions.placedInredientNotFound</code>	<code>c2pa.actions</code>	The <code>actions</code> assertion being reviewed has a <code>placed</code> action without a resolvable <code>ingredient</code> URI. <code>value</code> should be 1.
<code>ingredient.actionMissing</code>	<code>c2pa.ingredient</code>	The <code>ingredient</code> assertion being reviewed does not have at least one action that references it in its claim. <code>value</code> should be 1.
<code>ingredient.notVisible</code>	<code>c2pa.ingredient</code>	The <code>ingredient</code> assertion being reviewed is not visible in the digital content bound to that manifest. <code>value</code> should be 1.
<code>depthMap.sceneMismatch</code>	<code>c2pa.depthmap.GDepth</code>	The contents of the depth map assertion do not correspond to the scene portrayed in the primary presentation in the asset (e.g. because of a picture-of-picture attack)
<code>thumbnail.primaryMismatch</code>	<code>c2pa.thumbnail.claim</code>	The thumbnail’s contents do not match the contents of the primary presentation in the asset.
<code>stds.schema-org.CreativeWork.misattributed</code>	<code>stds.schema-org.CreativeWork</code>	One or more of the roles listed in a <code>CreativeWork</code> assertion is misattributed to the wrong actor (e.g. the wrong credit is given for the <code>editor</code> role)
<code>stds.schema-org.CreativeWork.missingAttribution</code>	<code>stds.schema-org.CreativeWork</code>	An attribution for a role in a <code>CreativeWork</code> assertion is missing (e.g. a person who played the role of <code>producer</code> isn’t credited)

Value of code	Applicable Assertion	Meaning
stds.iptc.location.inaccurate	stds.iptc	The reported location is inaccurate (e.g. reported to be in New York, NY but appears to be in another city entirely)

17.3.1.3. References

Because the `reference` field of the `assertion metadata` assertion is a standard `hashed_uri`, it is also possible to have an `assertion metadata` assertion refer to assertions in other manifests than the active one. For example, the active manifest could include an `assertion metadata` assertion that validates the `precise location` assertion present in an ingredient's manifest.

17.3.1.4. DateTime

If a `dateTime` field is present, its value shall be a date time string that complies with ISO 8601.

17.4. Standard C2PA Assertion Summary

The standard C2PA assertions are:

Type	Assertion	Schema	Serialization
Assertion Metadata	c2pa.assertion.metadata	C2PA	CBOR
Data Hash	c2pa.hash.data	C2PA	CBOR
BMFF-based Hash	c2pa.hash.bmff	C2PA	CBOR
Soft Binding	c2pa.soft-binding	C2PA	CBOR
Cloud Data	c2pa.cloud-data	C2PA	CBOR
Thumbnail	c2pa.thumbnail.claim (claim creation time) c2pa.thumbnail.ingredient (importing an ingredient)	C2PA	Embedded File
Actions	c2pa.actions	C2PA	CBOR
Ingredient	c2pa.ingredient	C2PA	CBOR
GDepth Depthmap	c2pa.depthmap.GDepth	https://developers.google.com/depthmap-metadata/reference	CBOR
Exif information	stds.exif	C2PA	JSON-LD

Type	Assertion	Schema	Serialization
IPTC Photo Metadata	stds.iptc.photo-metadata	C2PA	JSON-LD
Claim Review	stds.schema-org.ClaimReview	Schema.org ClaimReview	JSON-LD
Creative Work	stds.schema-org.CreativeWork	Schema.org CreativeWork	JSON-LD

17.5. Data Hash

Portion(s) of a non-BMFF-based asset that a claim generator wishes to uniquely identify with a hard binding (aka cryptographic hash) shall be described using data hash assertions. The types of hashes which can be created and stored in such an assertion are described in [Section 12.1, “Hashing”](#) and the value shall be present in the **hash** field.

Each data hash assertion defines a specified range of bytes over which the hash has been computed. By default, this range is the entire asset containing this assertion or the entire asset referred to by the **url** value in the assertion. If only a portion of the asset shall be hashed, then the range(s) to be excluded shall be present in the array value of the **exclusions** field.

NOTE

The **url** value provides flexibility for hashing assets that may be represented in multiple chunks or portions, local or remote.

A Data Hash assertion shall have a label of **c2pa.hash.data**.

A Data Hash assertion must not appear in a [_Cloud Data assertion](#).

17.5.1. Schema and Example

The [CDDL Definition](#) for this type is:

```

; Also check optionality within the hash-map
; The data structure used to store the cryptographic hash of some or all of the asset's data
; and additional information required to compute the hash.
data-hash-map = {
    ? "exclusions": [1* EXCLUSION_RANGE-map],
    ? "alg":tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash
algorithm used to compute the hash in this assertion, taken from the C2PA hash algorithm
identifier list. If this field is absent, the hash algorithm is taken the `alg` value of the
enclosing structure. If both are present, the field in this structure is used. If no value
is present in any of these places, this structure is invalid; there is no default.
    "hash": bstr, ; byte string of the hash value
    "pad": bstr, ; byte string used for filling up space
    ? "pad2": bstr, ; byte string used for filling up space
    ? "name": tstr .size (1..max-tstr-length), ; (optional) a human-readable description of
what this hash cover
    ? "url": uri, ; (optional) a file or http(s) URL to where the bytes that are being hashed
can be found.
        ; This is useful for cases where the data lives in a different file chunk or
side-car than the claim.
}

EXCLUSION_RANGE-map = {
    "start": int, ; Starting byte of the range
    "length": int, ; Number of bytes of data to exclude
}

```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```
{
    "exclusions": [
        {
            "start": 9960,
            "length": 4213
        }
    ],
    "alg" : "sha256",
    "hash": 'Auxjtmax46cC2N3Y9aFmB09Jfay8LEwJWzBUTZ0sUM8gA=',
    "pad" : 'abc',
    "name": "JUMBF manifest"
}
```

Normally, the **start** and **length** values of an **exclusion** shall be written in their preferred serialization (i.e., "as short as possible"). However, when a data hash assertion needs to be created but the **start** and **length** values are not yet known, they shall be created "as large as possible", which would be as a 32-bit integer.

The **pad** value shall always be present but shall be a byte string of length 0 unless used to replace (aka "pad") bytes during multiple pass processing.

NOTE

[Section 9.4, “Multiple Step Processing”](#) describes how to fill in the correct values and adjust the padding.

17.5.2. Special consideration for JPEG-1

When hashing a JPEG-1 (.jpg) file into which the C2PA Manifest will be embedded, the APP11 marker (**FFEB**) and the segment's length (**Lp**) of all APP11 segments containing the JUMBF data shall be included in the exclusion range.

NOTE

All the APP11 segments containing the C2PA Manifest JUMBF are contiguous so that only a single range is required.

17.5.3. Special consideration for PNG

When hashing a PNG (.png) file into which the C2PA Manifest will be embedded, it is important that the **Length** and the '**caBX**' (representing the chunk type) of the chunk containing the JUMBF data be included in the exclusion range.

17.6. BMFF-Based Hash

Portion(s) of a BMFF-based asset that a claim generator wishes to uniquely identify with a hard binding (aka cryptographic hash) shall be described using BMFF-based Hash assertions.

A BMFF-based Hash assertion shall have a label of **c2pa.hash.bmff**.

A BMFF-based Hash assertion must not appear in a [Cloud Data assertion](#).

To create the hash specified in the value field of a bmff hash, all bytes of the file are added to the hash excluding those BMFF boxes which match any exclusion entry in the exclusions array. Any box included in the hash also includes its box headers; any box excluded from the hash also excludes its box headers.

A box matches an exclusion entry in the exclusions array if and only if all of the following conditions are met.

- The box's location in the file exactly matches the exclusion entry's xpath field.
- If length is specified in the exclusion entry, the box's length exactly matches the exclusion entry's length field.
Note: The length includes the box headers.
- If version is specified in the exclusion entry, the box is a FullBox and the box's version exactly matches the exclusion entry's version field.
- If flags (byte array of exactly 3 bytes) is specified in the exclusion entry and the box is a FullBox. If exact is set to true or not specified, the box's flags (bit(24), i.e., 3 bytes) also exactly matches the exclusion entry's flags field. If exact is set to false, the bitwise-and of the box's flags (bit(24), i.e., 3 bytes) with the exclusion entry exactly matches the exclusion entry (i.e. the box has at least those bits set but may also have additional bits set).
- If data (array of objects) is specified in the exclusion entry, then for each item in the array, the box's binary data at that item's offset field exactly matches that item's bytes field.

The xpath syntax shall be limited to the following strict subset.

- Only abbreviated syntax shall be used.
- Only full paths shall be used.
- Only node selection via `node` or `node[integer]` shall be used.
- Descendent syntax, i.e., `//`, shall NOT be used.
- All nodes shall be BMFF **4cc** codes.

Complete Syntax:

```

xpath = '/' nodes
nodes = node
    | node '/' nodes
node = box4cc
    | box4cc '[' integer ']'

```

Where:

`box4cc` is any 4cc allowed by ISO/IEC 14496-12 for a BMFF box.
`integer` is any non-zero positive integer with no leading zeros.

A non-leaf xpath node shall only point to a container box that has no fields of its own (i.e. contains no data, only child boxes) and that does not inherit from FullBox. This ensures that a c2pa validator does not need to be aware of the syntax and semantics of unusual boxes that contain other boxes. If a child box of such an unusual box needs to be excluded in full or in part, the exclusion's xpath shall point to the unusual box itself and the subset member shall exclude the byte ranges containing the excluded child box data. For example, the `sgpd` box contains other boxes but is unusual in that it inherits from FullBox; as such, if excluding child box(es) from `sgpd` is required, the assertion shall use an xpath pointing to the `sgpd` itself (e.g. `/moof/traf/sgpd`) and shall use the subset member to exclude the desired bytes.

If the manifest is embedded into the file, the box containing it shall be one of the entries in the exclusions array and all file offsets shall be excluded wherever they appear. Refer to [Section 10.3.2, “Embedding manifests into BMFF-based assets”](#) for more information.

Because the box that contains the manifest is at the root and always excludes file offsets, the hash can be computed before the manifest is created and the manifest can then be added without changing any existing portion of the file or invalidating the hash.

If a non-root excluded box is removed after the manifest is created, it shall be replaced with a `free` box of the same size to ensure that the hash of its parent box is not invalidated due to its size changing. If it is expected that a non-root excluded box may be added after the manifest is created, then, at manifest creation time, a `free` box shall be inserted with sufficient space for the excluded box and that `free` box shall also be excluded by an exclusion entry using its full xpath. When the excluded box is added, the `free` box shall be shrunk (or removed) to make space for the added box.

17.6.1. Schema and Example

The [CDDL Definition](#) for this type is:



```

bmff-hash-map = {
  "exclusions": [1* exclusions-map],
  ? "alg": tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash algorithm used to compute this hash, taken from the C2PA hash algorithm identifier list. If this field is absent, the hash algorithm is taken from an enclosing structure as defined by that structure. If both are present, the field in this structure is used. If no value is present in any of these places, this structure is invalid; there is no default.
  ? "hash": bstr, ; For non-fragmented mp4, this shall be the hash of the entire BMFF file excluding boxes listed in the exclusions array; if the merkle field is present, this hash also excludes all 'mdat' boxes. For fragmented mp4, this field is required to be absent.
  ? "merkle": [1* merkle-map], ; A set of merkle tree rows and the associated data required to enable verification of a single 'mdat' box, multiple 'mdat' boxes, and/or individual fragment files within the asset.
  ? "name": tstr .size (1..max-tstr-length), ; optional) a human-readable description of what this hash covers.
  ? "url": uri, ; (optional) a file or http(s) URL to where the bytes that are being hashed lived. This is useful for cases where the data lives in a different file chunk or side-car than the claim.
}

;(optional) CBOR byte string of exactly 3 bytes.
flag-type = bytes

flag-t = flag-type .eq 3

exclusions-map = {
  "xpath": tstr, ; Location of box(es) to exclude from the hash starting from the root node as an xpath formatted string of version https://www.w3.org/TR/xpath-10/ with highly constrained syntax.
  ? "length": int, ; (optional) Length that a leafmost box must have to exclude from the hash.
  ? "data": [1* data-map], ; (optional) The data in the leafmost box at the specified offset must be identical to the specified data for the box to be excluded from the hash.
  ? "subset": [1* subset-map], ; (optional) Only this portion of the excluded box shall be excluded from the hash. Each entry in the array must have a monotonically increasing offset. No subset within the array may overlap. The last entry may have a length of zero; this indicates that the remainder of the box from that offset onward is excluded.
  ? "version": int, ; (optional) Version that must be set in a leafmost box for the box to be excluded from the hash. Shall only be specified for a box that inherits from FullBox.
  ? "flags": flag-t, ; (optional) byte string of exactly 3 bytes. The 24-bit flags that must be set in a leafmost box for the box to be excluded from the hash. Shall only be specified for a box that inherits from FullBox.
  ? "exact": bool, ; (optional) indicates that flags must be an exact match. If not specified, defaults to true. Shall only be specified for a box that inherits from FullBox and when flags is also specified.
}

data-map = {
  "offset": int,
  "value" : bstr,
}
subset-map = {
  "offset": int,
  "length": int,
}

; Each entry in a map is a merkle tree rows and the associated data required to enable validation of a single
; 'mdat' box or multiple 'mdat' boxes within the asset.",
merkle-map = {
  "uniqueId": int, ; 1-based unique id used to differentiate across files to determine which merkle tree should be used to validate a given 'mdat' box.
  "localId": int, ; Local id used to differentiate across multiple 'mdat' boxes within a
}

```

```

single file to determine which merkle tree should be used to validate that 'mdat' box.
  "count": int, ; Number of leaf nodes in the Merkle tree. Null nodes are not included in
this count.
  ? "alg": tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash
algorithm used to compute the hashes in this merkle tree, taken from the C2PA hash algorithm
identifier list. If this field is absent, the hash algorithm is taken the `alg` value of the
enclosing structure as defined by that structure. If both are present, the field in this
structure is used. If no value is present in any of these places, this structure is invalid;
there is no default.
  ? "initHash": bstr, ; For fragmented mp4, this field is required to be present and shall
be the hash of the initialization segment for chunks hashed by this merkle tree. For non-
fragmented mp4, this field is required to be absent.
  "hashes": [1* bstr], ; An ordered array representing a single row of the Merkle tree which
may be the leaf-most row, root row, or any intermediate row. The depth of the row is
implied by (shall be computed from) the number of items in this array.
}

```

An example in CBOR Diagnostic Format ([.cbordiag](#)) for a monolithic mp4 file asset where the **mdat** box is validated as a unit is shown below:

```
{
  "exclusions": [
    {
      "xpath": "/moov[1]/pssh"
    },
    {
      "xpath": "/emsg",
      "data": [
        {
          "offset": 20,
          "value": b64'r3avWCpXHkmKHATFsV0Q5g'
        }
      ]
    }
  ],
  "name": "Example BMFF-hash object",
  "hash": b64'EiAuxjtmax46cC2N3Y9aFmB09Jfay8LEwJWzBUTZ0sUM8gA'
}
```

An example in CBOR Diagnostic Format ([.cbordiag](#)) for an asset composed of fragmented mp4 files is shown below:

```
{
  "exclusions": [
    {
      "xpath": "/c2pa"
    },
    {
      "xpath": "/moov[1]/pssh"
    },
    {
      "xpath": "/emsg",
      "length": 200,
      "data": [
        {
          "offset": 5,
          "value": b64'9Q=='
        }
      ]
    }
  ]
}
```

```

        },
        {
          "offset": 20,
          "value": b64'UAJXD79SlkG9rfnmcsqTUA=='
        },
        {
          "offset": 70,
          "value": b64'0xKM'
        }
      ],
      "subset": [
        {
          "offset": 5,
          "length": 7
        },
        {
          "offset": 20,
          "length": 28
        },
        {
          "offset": 45,
          "length": 63
        },
        {
          "offset": 80,
          "length": 112
        }
      ],
      "version": 1,
      "flags": b64'ZDNx'
    }
  ],
  "alg": "sha256",
  "merkle": [
    {
      "uniqueId": 17,
      "localId": 19,
      "count": 23,
      "initHash": b64'Hf0IgeqbL0m+FTTLpUWwsDGR8pvhUR1AlwvaXjQ0qGY=',
      "hashes": [ b64'HvWZOxKMfkSatRAygs8DJfnEEcN/G1BNi359NdIDxbQ=' , b64'HvWZOxKMfkSatRAygs8DJfnEEcN/G1BNi359NdIDxbQ=' ]
    },
    {
      "uniqueId": 34,
      "localId": 38,
      "count": 69,
      "initHash": b64'Hf0IgeqbL0m+FTTLpUWwsDGR8pvhUR1AlwvaXjQ0qGY=',
      "hashes": [ b64'9Zk7Eox+RJq1EDKCzwMl+cQRw38bUE2Lfn010gPFtB0=' , b64'9Zk7Eox+RJq1EDKCzwMl+cQRw38bUE2Lfn010gPFtB0=' , b64'mTsSjH5EmrUQMoLPAYX5xBHDfxtQTYt+fTXSA8W0Hf0=' , b64'mTsSjH5EmrUQMoLPAYX5xBHDfxtQTYt+fTXSA8W0Hf0=' , b64'0xKMfkSatRAygs8DJfnEEcN/G1BNi359NdIDxbQd/Qg=' ]
    },
    {
      "uniqueId": 51,
      "localId": 57,
      "count": 46,
      "initHash": b64'Hf0IgeqbL0m+FTTLpUWwsDGR8pvhUR1AlwvaXjQ0qGY=',
      "hashes": [ b64'0xKMfkSatRAygs8DJfnEEcN/G1BNi359NdIDxbQd/Qg=' ]
    }
  ],
  "name": "Example BMFF-hash object for fMP4"
}

```

A pseudo-code implementation of this algorithm follows.

```
offset = 0
While (offset < length of file)
    Starting at offset, locate the first byte of the first box that matches any entry in the
    exclusions array, call this first_excluded_byte
        If no such box is found, set first_excluded_byte = length of file
        Determine the length of that box, call this excluded_byte_count
            If no such box was found, set excluded_byte_count = 0
            To the hash, add all bytes between offset and first_excluded_byte minus one (inclusive)
            If first_excluded_byte < length of file and there exists a subset array within the
            exclusion that determined the value of first_excluded_byte
                set next_included_begin = first_excluded_byte
                For each entry in the subset array within the exclusion that determined the value of
                first_excluded_byte
                    Set next_excluded_begin = this subset array entry's offset field plus
                first_excluded_byte
                    If next_excluded_begin > next_included_begin
                        To the hash, add all bytes between next_included_begin and
                    next_excluded_begin minus one (inclusive)
                        Set next_included_begin = this subset array entry's length field plus
                    next_excluded_begin
                    If next_included_begin < first_excluded_byte + excluded_byte_count
                        To the hash, add all bytes between next_included_begin and first_excluded_byte +
                    excluded_byte_count minus one (inclusive)
                    Set offset = first_excluded_byte + excluded_byte_count
```

17.6.2. Exclusion list profiles

17.6.2.1. Basic profile

Typical untimed media (e.g., still photos) and timed media (e.g. videos with or without audio tracks, whether fragmented or not) need only include the mandatory exclusions listed in [Exclusion List Requirements](#).

NOTE

Editor's Note

Additional recommended exclusions for other types of media should be added here as needed.

17.7. Soft Binding

If a claim generator wishes to provide a soft binding for the asset's content, it shall be described using a soft binding assertion. The types of soft bindings which can be created and stored in such an assertion are described in [Section 8.3, “Soft Bindings”](#).

NOTE

The `url` value provides flexibility for hashing assets that may be represented in multiple chunks or portions, local or remote.

A Soft Binding assertion shall have a label of `c2pa.soft-binding`.

17.7.1. Schema and Example

The [CDDL Definition](#) for this type is:

```
;The data structure used to store one or more soft bindings across some or all of the
asset's content
soft-binding-map = {
    "alg": tstr, ; A string identifying the soft binding algorithm and version of that
algorithm used
        ; to compute the value, taken from the C2PA soft binding algorithm identifier
list. If this field is absent, the algorithm is taken from the `softbinding-alg` value of
the enclosing structure. If both are present, the field in this structure is used.
        ; If no value is present in any of these places, this structure is invalid;
there is no default.
    "blocks": [1* soft-binding-block-map],
    "pad": bytes, ; byte string used for filling up space
    ? "pad2": bytes, ; byte string used for filling up space
    ? "name": tstr .size (1..max-tstr-length), ; (optional) a human-readable description of
what this hash covers
    ? "alg-params": bstr, ; (optional) CBOR byte string describing parameters of the soft
binding algorithm.
        ; If this field is absent, the algorithm is taken from the
`softbinding-alg-params`
        ; value of the enclosing structure, if present."
    ? "url": uri, ; (optional) a file or http(s) URL to where the bytes that are being hashed
lived.
        ; This is useful for cases where the data lives in a different file chunk or
side-car
        ; than the claim.
}

soft-binding-block-map = {
    "scope": soft-binding-scope-map,
    "value": bstr, ; CBOR byte string describing, in algorithm specific format,
        ; the value of the soft binding computed over this block of digital content"
}
soft-binding-scope-map = {
    ? "extent": bstr, ;CBOR byte string describing, in algorithm specific format,
        ; the part of the digital content over which the soft binding value has
been computed"
    ? "timespan":soft-binding-timespan-map,
}
soft-binding-timespan-map = {
    "start": uint, ; Start of the time range (as milliseconds from media start) over which the
soft binding value has been computed.
    "end": uint, ; End of the time range (as milliseconds from media start) over which the
soft binding value has been computed.
}
```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```

{
  "url": 32("http://example.c2pa.org/media.mp4"),
  "alg": "phash",
  "blocks": [
    {
      "scope": {
        "extent": b64'c2NvcGUxCg==',
        "timespan": {
          "start": 0,
          "end": 133016
        }
      },
      "value": b64'dmFsdWUxPg=='
    },
    {
      "scope": {
        "extent": b64'c2NvcGUyCg==',
        "timespan": {
          "start": 133017,
          "end": 245009
        }
      },
      "value": b64'dmFsdWUyCg=='
    }
  ],
  "pad": h'00'
}

```

The soft binding algorithm used shall be present as the value of the `alg` field, and the blocks over which it was applied shall be listed in the `blocks` field. If the algorithm used requires any additional parameters, they should be present as the value of `alg-params`.

17.8. Cloud Data

There are use cases where storing the data for the assertion remotely, such as in the cloud, is better than embedded inside the asset, especially when the data is large. For any such cases, it is possible to use a special type of assertion that serves as a reference to that information. For privacy and reliability reasons, data referenced through a cloud data assertion shall be considered optional: their contents should not be retrieved as part of manifest validation. A validator may retrieve the contents later to serve an application-dependent need, such as further exploration of the provenance history.

A Cloud Data assertion shall have a label of `c2pa.cloud-data`.

A Cloud Data assertion must not refer to an assertion with the label `c2pa.hash.data` or `c2pa.hash.bmff`.

17.8.1. Schema and Example

The [CDDL Definition](#) for this type is:

```

; Assertion that references the actual assertion stored in the cloud
cloud-data-map = {
  "label": tstr, ; label for the cloud-based assertion (eg.c2pa.actions)
  "size": size-type, ; Number of bytes of data
  "location": $hashed-ext-uri-map, ; a file or http(s) URL to where the bytes that are being
  hashed can be found
  "content_type": tstr, ; media/MIME type for the data
  ? "metadata": $assertion-metadata-map, ; additional information about the assertion
}

; size is minimum 1 in multiples of 1.0
size-type = int .ge 1

```

An example in CBOR Diagnostic Format (.cbordiag) is shown below:

```
{
  "label": "c2pa.thumbnail.claim.jpg",
  "size": 98765,
  "location": {
    "url": "https://some.storage.us/foo",
    "hash": b64'zP84FPSremIrAQHlhw+hRYQdZp/+KggnD0W8opXlIQQ='
  },
  "content_type": "application/jpeg"
}
```

To allow external references to content, this type uses the `hashed-ext-uri` type which is a variation on the `hashed-uri` type used elsewhere in this specification. The CDDL definition for `hashed-ext-uri` is:

```

; The data structure used to store a reference to an external URL and its hash. We use a
socket/plug here to allow hashed-ext-uri-map to be used in individual files without having
the map defined in the same file
$hashed-ext-uri-map /= {
  "url": url-regexp-type, ; http/https URI reference
  "alg": tstr .size (1..max-tstr-length), ; A string identifying the cryptographic hash
  algorithm used to compute all hashes in this claim, taken from the C2PA hash algorithm
  identifier list. Unlike alg fields in other types, this field is mandatory here.
  "hash": bstr, ; byte string containing the hash value
}

; with CBOR Head (#) and tail ($) are introduced in regexp, so not needed explicitly
url-regexp-type /= tstr .regexp "https?:\/\/[-a-zA-Z0-9@:%._\+\~#=]{2,256}\/\/[a-
z]{2,6}\b[-a-zA-Z0-9@:%_\+\.\~#\&/=]*"

```

NOTE In keeping with common practice, it is recommended that the `https` scheme be used to retrieve assertion data to protect the privacy of the data in transit, but `http` is also permitted because the data's integrity is protected by the `hash` field and this privacy may not be required in all circumstances. Authors of manifests with cloud-hosted data should choose the scheme to suit their needs.

17.9. Thumbnail

A **thumbnail** assertion provides an approximate visual representation of the asset at a specific event in the lifecycle of an asset. There are currently two specific events - ingredient import and claim creation - each using a unique label for the assertion.

For thumbnails created at claim creation time, the Thumbnail assertion shall have a label that starts with `c2pa.thumbnail.claim` and be followed by the [IANA registry](#) image type (e.g., `c2pa.thumbnail.claim.png`). For each of these types of thumbnails, there can be only one per claim.

When importing an ingredient (see [Section 9.3.2.2, “Adding Ingredients”](#)), it is preferable to reference that ingredient’s own manifest-stored thumbnail. However, some ingredients may not include a thumbnail assertion, or even a manifest. In that case, a new thumbnail of the ingredient should be generated, and a new thumbnail assertion in the active manifest created. The Thumbnail assertion shall have a label that starts with `c2pa.thumbnail.ingredient` and be followed by an underscore (_) (U+005F) then a unique ID such as a simple monotonically increasing integer and ending with the image type. For example, an ingredient thumbnail of type `jpeg` could have label `c2pa.thumbnail.ingredient_1.jpg`.

The data in a thumbnail assertion is the bits of a file (such as a raster image) in whatever format is desired by the claim generator. The Embedded File content type (ISO 19566-5:AMD-1), `bfdb`, shall be used to contain the thumbnail’s data.

17.10. Actions

An **actions** assertion provides information on edits and other actions taken that affect the asset’s content. There will be an array of actions - each action declaring *what* took place on the asset, *when* it took place, along with possible other information such as what software performed the action.

An Actions assertion shall have a label of `c2pa.actions`.

Actions are modelled after [XMP ResourceEvents](#).

For each action present in the array, the value of the `action` field shall be either a well-known common action name (`c2pa.resized`, `c2pa.edited`, etc.) or entity-specific action name (`adobe.ps.gaussian_blur`, etc.). The value of the `when` field should be the ISO 8601-compliant string of when the action was performed.

The set of defined names, prefixed with `c2pa.` are:

Action	Meaning	Source
<code>c2pa.converted</code>		XMP
<code>c2pa.copied</code>		XMP
<code>c2pa.created</code>		XMP

Action	Meaning	Source
c2pa.cropped		XMP
c2pa.edited		XMP
c2pa.filtered		XMP
c2pa.formatted		XMP
c2pa.version_updated		XMP
c2pa.printed		XMP
c2pa.published		XMP
c2pa.managed		XMP
c2pa.produced		XMP
c2pa.resized		XMP
c2pa.saved		XMP
c2pa.transcoded	A direct conversion of one encoding to another, including resolution scaling, bitrate adjustment and encoding format change. Does not include any adjustments that would affect the "editorial" meaning of the content.	C2PA
c2pa.repackaged	A conversion of one packaging or container format to another. Content may be repackaged without transcoding. Does not include any adjustments that would affect the "editorial" meaning of the content.	C2PA
c2pa.placed	Added/Placed an ingredient into the asset	C2PA
c2pa.unknown	Something happened, but the claim_generator can't specify what	C2PA

In addition, an action may include a **actors** key which can point to the identity of one or more human actor who performed the action as described in [Section 17.16, “Common Data Model: Actor”](#).

When using a **c2pa.transcoded**, **c2pa.repackaged**, or a **c2pa.placed** action, the **parameters** field shall contain the JUMBF URI to the related ingredient assertion.

17.10.1. Asset Renditions

Asset renditions are a common occurrence when distributing media on the internet. These renditions are often created for the purpose of delivering media to consumers in differing connectivity, screen resolution, and other environments. We can use the **actions** assertion to help consuming actors understand the intention of certain claim creators to create asset renditions.

The presence of only the **c2pa.transcoded** and/or **c2pa.repackaged** actions in a **c2pa.actions** assertion

provides a signal to the manifest consumer that the signer is asserting that no "editorial" changes have happened between the ingredient asset(s) and this one. Editorial changes are those that alter the intent and/or meaning of the content.

The additional presence of a single "parentOf" ingredient provides a further signal to the manifest consumer that the signer is asserting that the asset has been derived directly from that parent.

17.10.2. Schema and Example

The [CDDL Definition](#) for this type is:

```
actions-map = {
    "actions" : [1* action-items-map], ; list of actions
    ? "metadata": $assertion-metadata-map, ; additional information about the assertion
}

buuid = #6.37(bstr)

$action-choice /= "c2pa.converted"
$action-choice /= "c2pa.copied"
$action-choice /= "c2pa.created"
$action-choice /= "c2pa.cropped"
$action-choice /= "c2pa.edited"
$action-choice /= "c2pa.filtered"
$action-choice /= "c2pa.formatted"
$action-choice /= "c2pa.version_updated"
$action-choice /= "c2pa.printed"
$action-choice /= "c2pa.published"
$action-choice /= "c2pa.managed"
$action-choice /= "c2pa.produced"
$action-choice /= "c2pa.resized"
$action-choice /= "c2pa.saved"
$action-choice /= "c2pa.transcoded"
$action-choice /= "c2pa.repackaged"
$action-choice /= "c2pa.placed"

actor-map = {
    ? "credentials": [1* $hashed-uri-map / $hashed-ext-uri-map],
    ? "identifier": tstr .size (1..max-tstr-length), ; An identifier for a human actor
}

action-items-map = {
    "action": $action-choice,
    ? "when": tdate, ; Timestamp of when the action occurred.
    ? "softwareAgent": tstr .size (1..max-tstr-length), ;The software agent that performed the action.
    ? "changed": tstr .size (1..max-tstr-length), ; A semicolon-delimited list of the parts of the resource that were changed since the previous event history. If not present, presumed to be undefined. When tracking changes and the scope of the changed components is unknown, it should be assumed that anything might have changed.
    ? "instanceID": buuid, ; The value of the xmpMM:InstanceID property for the modified (output) resource
    ? "parameters": tstr .size (1..max-tstr-length), ; Additional description of the action
    ? "actors": [1* actor-map], ; An array of the creators that undertook this action
}
```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```
{  
  "actions": [  
    {  
      "action": "c2pa.filtered",  
      "when": 0("2020-02-11T09:00:00Z"),  
      "softwareAgent": "Joe's Photo Editor 2020 for Windows",  
      "changed": "change1,change2",  
      "instanceID": 37(h'ed610ae51f604002be3dbf0c589a2f1f'),  
      "actor": [  
        {  
          "credentials": [  
            {  
              "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-  
329BF39FA1E4/c2pa.credentials/Joe_Bloggs",  
              "alg": "sha256",  
              "hash": b64'hoOspQQ1lFTy/4Tp8Epx670E5QW5NwkNR+2b30KFXug='  
            }  
          ]  
        }  
      ]  
    },  
    "metadata": {  
      "reviewRating": [  
        {  
          "value": 1,  
          "explanation": "Content bindings did not validate"  
        }  
      ],  
      "dateTime": 0("2021-06-28T16:34:11.457Z")  
    }  
  }  
}
```

17.11. Ingredient

When assets are composed together, for example placing an image into a layer in Photoshop or an audio clip into a video in Premiere, it is important that information about any claim from the placed asset be recorded into the new asset to provide a way to understand the entire history of the new composed asset. This is also true when an existing asset is used to create a derived asset or asset rendition.

An Ingredient assertion shall have a label of `c2pa.ingredient`.

NOTE Since there will most likely be more than one ingredient assertion, the use of the monotonically increasing index in the label would be used (e.g., `c2pa.ingredient_1`, `c2pa.ingredient_2`).

17.11.1. Concept

The concept of ingredients in C2PA is modelled on the XMP Ingredient and Pantry model, as described in the [Partner Guide to XMP for Dynamic Media](#) and [Asset Relationships in XMP](#). That model relies on the fact that each asset used in

the construction of a document has, at the time of inclusion, at least one [unique identifier](#).

If the ingredient being added contains XMP, then asset's `xmpMM:DocumentID` becomes the `documentID` field while the `xmpMM:InstanceID` becomes the `instanceID`. However, if the ingredient being added does not have any associated XMP, then it may be possible for the XMP to be created and added to the ingredient itself and the identifiers used as described above.

When it is not possible or desirable to create XMP for an ingredient, then some other [unique identifier](#) for the ingredient shall be used instead for the required values. In this situation, the `instanceID` field shall contain the unique identifier and the `documentID` field shall not be present.

17.11.2. Relationship

When adding an ingredient, its relationship to the current asset shall be described. These are the possible values of the `relationship` field and their meanings:

Value	Meaning
<code>parentOf</code>	The current asset is a derived asset or asset rendition of this ingredient. This relationship value is also used with update manifests .
<code>componentOf</code>	The current asset is composed of multiple parts, this ingredient being one of them.

17.11.3. Title

The value of `dc:title` shall be a human-readable name for the ingredient, which may be taken either from the asset's XMP or the asset's name in a local or remote (e.g., cloud-based) filesystem. The [Media Type](#) of the ingredient shall be declared in `dc:format`.

17.11.4. Schema and Example

The [CDDL Definition](#) for this type is:

```
;Assertion that describes an ingredient used in the asset
ingredient-map = {
    "dc:title": tstr, ; name of the ingredient
    "dc:format": format-string, ; Media Type of the ingredient
    ? "documentID": tstr, ; value of the ingredient's `xmpMM:DocumentID`
    "instanceID": tstr, ; unique identifier, such as the value of the ingredient's
                      `xmpMM:InstanceID`
    "relationship": $relation-choice, ; The relationship of this ingredient to the asset it is
                      an ingredient of.
                           ; For example, if an ingredient with a 'parentOf'
relationship is added to
                           ; an asset, then the asserter is stating that the
current asset is a derived asset of the ingredient.
    ? "c2pa_manifest": $hashed-uri-map, ; hashed_uri reference to the C2PA Manifest of the
```

```

ingredient
  ? "thumbnail": $hashed-uri-map, ; hashed_uri reference to an ingredient thumbnail
  ? "validationStatus": $status-map ; validation status of the ingredient
  "metadata": $assertion-metadata-map ; additional information about the assertion
}

format-string = tstr .regexp "^\\w+/-+\\w+$"
; Assertion that describes an ingredient used in the asset
$relation-choice /= "parentOf"
$relation-choice /= "componentOf"

; Success codes
$status-code /= "claimSignature.validated"
$status-code /= "signingCredential.trusted"
$status-code /= "timeStamp.trusted"
$status-code /= "assertion.hashedURI.match"
$status-code /= "assertion.dataHash.match"
$status-code /= "assertion.bmffHash.match"
$status-code /= "assertion.accessible"

; Failure codes
$status-code /= "claim.missing"
$status-code /= "claim.multiple"
$status-code /= "claim.hardBindings.missing"
$status-code /= "ingredient.hashedURI.mismatch"
$status-code /= "claimSignature.missing"
$status-code /= "claimSignature.mismatch"
$status-code /= "manifest.multipleParents"
$status-code /= "manifest.update.invalid"
$status-code /= "manifest.update.wrongParents"
$status-code /= "signingCredential.untrusted"
$status-code /= "signingCredential.invalid"
$status-code /= "signingCredential.revoked"
$status-code /= "signingCredential.expired"
$status-code /= "timeStamp.mismatch"
$status-code /= "timeStamp.untrusted"
$status-code /= "timeStamp.outsideValidity"
$status-code /= "assertion.hashedURI.mismatch"
$status-code /= "assertion.missing"
$status-code /= "assertion.undeclared"
$status-code /= "assertion.inaccessible"
$status-code /= "assertion.notRedacted"
$status-code /= "assertion.selfRedacted"
$status-code /= "assertion.action.redacted"
$status-code /= "assertion.dataHash.mismatch"
$status-code /= "assertion.bmffHash.mismatch"
$status-code /= "assertion.cloud-data.hardBinding"
$status-code /= "assertion.cloud-data.actions"
$status-code /= "algorithm.unsupported"

status-map = {
  "code": $status-code, ; A label-formatted string that describes the status
  ? "url": url-regexp-type, ; JUMBF URI reference
  ? "explanation": tstr .size (1..max-tstr-length), ; A human readable string explaining the
  status
}

```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```
{
  "dc:title": "image 1.jpg",
  "dc:format": "image/jpeg",
  "documentID": "uuid:87d51599-286e-43b2-9478-88c79f49c347",
  "instanceID": "uuid:7b57930e-2f23-47fc-affe-0400d70b738d",
  "c2pa_manifest": {
    "url": "self#jumbf=c2pa/urn:uuid:5E7B01FC-4932-4BAB-AB32-D4F12A8AA322",
    "hash": b64'1kjJT0108b71c195UxgfHD3eDgk9VrCedW8n3fYTRMk='
  },
  "thumbnail": {
    "url": "self#jumbf=cp2a/urn:uuid:F9168C5E-CEB2-4faa-B6BF-329BF39FA1E4/c2pa.thumbnail.ingredient_1.jpg",
    "hash": b64'UjRAYWiAq4lfCRDmksWAlDJN/XtHHFFwMWymsZsm3j8='
  },
  "relationship": "parentOf",
  "metadata": {
    "reviewRating": [
      {
        "value": 5,
        "explanation": "Content bindings validated"
      }
    ],
    "dateTime": o("2021-06-28T16:49:32.874Z")
  }
}
```

17.11.5. Thumbnails

When adding an ingredient, it may be useful to also include a thumbnail of the ingredient to help establish the state of the ingredient at the time of import. For that purpose, an ingredient import thumbnail should be added as described in [Section 17.11.5, “Thumbnails”](#) and referenced herein via a URI reference.

17.11.6. Existing manifests

If the ingredient has an existing C2PA Manifest Store, then all manifests in the store (both standard and updates) shall be copied into the C2PA Manifest Store for the asset. The [URI reference](#) to the ingredient’s active [C2PA Manifest](#) shall be stored as the value of `c2pa_manifest`. In addition, it is recommended that the C2PA validator validate the ingredient’s active C2PA manifest and document the validation status in the `validationStatus` field. When present, the value of the `validationStatus` field shall contain at least one entry in the array.

A `validationStatus` object consists of a `code` value that describes the validation status of a specific part of the manifest and an optional `url` field with a JUMBF URI reference to that element in the manifest. Depending on the code, the `url` could be to a claim, a claim signature or a specific assertion. It may also contain an optional description of the validation status if there is a need for additional human readable explanation.

17.11.6.1. Success codes

Value	Meaning	url Usage
<code>claimSignature.validated</code>	The claim signature referenced in the ingredient’s claim validated	C2PA Claim Signature Box

Value	Meaning	url Usage
<code>signingCredential.trusted</code>	The signing credential is listed on the validator's trust list.	C2PA Claim Signature Box
<code>timeStamp.trusted</code>	The time-stamp credential is listed on the validator's trust list.	C2PA Claim Signature Box
<code>assertion.hashedURI.match</code>	The hash of the referenced assertion in the ingredient's manifest matches the corresponding hash in the assertion's hashed URI in the claim.	C2PA Assertion
<code>assertion.dataHash.match</code>	Hash of a byte range of the asset matches the hash declared in the data hash assertion.	C2PA Assertion
<code>assertion.bmffHash.match</code>	Hash of a box-based asset matches the hash declared in the BMFF hash assertion.	C2PA Assertion
<code>assertion.accessible</code>	A non-embedded (remote) assertion was accessible at the time of validation.	C2PA Assertion

17.11.6.2. Failure codes

Value	Meaning	url Usage
<code>claim.missing</code>	The referenced claim in the ingredient's manifest cannot be found.	C2PA Claim Box
<code>claim.multiple</code>	More than one claim box is present in the manifest.	C2PA Claim Box
<code>claim.hardBindings.missing</code>	No hard bindings are present in the claim.	C2PA Claim Box
<code>ingredient.hashedURI.mismatch</code>	The hash of the referenced ingredient claim in the manifest does not match the corresponding hash in the ingredient's hashed URI in the claim.	C2PA Assertion
<code>claimSignature.missing</code>	The claim signature referenced in the ingredient's claim cannot be found in its manifest.	C2PA Claim Signature Box
<code>claimSignature.mismatch</code>	The claim signature referenced in the ingredient's claim failed to validate.	C2PA Claim Signature Box
<code>manifest.multipleParents</code>	The manifest has more than one ingredient whose <code>relationship</code> is <code>parentOf</code> .	C2PA Claim Box
<code>manifest.update.invalid</code>	The manifest is an update manifest, but it contains hard binding or actions assertions.	C2PA Claim Box
<code>manifest.update.wrongParents</code>	The manifest is an update manifest, but contains either zero or multiple <code>parentOf</code> ingredients.	C2PA Claim Box

Value	Meaning	url Usage
<code>signingCredential.untrusted</code>	The signing credential is not listed on the validator's trust list.	C2PA Claim Signature Box
<code>signingCredential.invalid</code>	The signing credential is not valid for signing.	C2PA Claim Signature Box
<code>signingCredential.revoked</code>	Signing credential has been revoked by the issuer.	C2PA Claim Signature Box
<code>signingCredential.expired</code>	Signing credential has expired.	C2PA Claim Signature Box
<code>timeStamp.mismatch</code>	The time-stamp does not correspond to the contents of the claim.	C2PA Claim Signature Box
<code>timeStamp.untrusted</code>	The time-stamp credential is not listed on the validator's trust list.	C2PA Claim Signature Box
<code>timeStamp.outsideValidity</code>	The signed time-stamp attribute in the signature falls outside the validity window of the signing certificate or the TSA's certificate.	C2PA Claim Signature Box
<code>assertion.hashedURI.mismatch</code>	The hash of the referenced assertion in the manifest does not match the corresponding hash in the assertion's hashed URI in the claim.	C2PA Assertion
<code>assertion.missing</code>	An assertion listed in the ingredient's claim is missing from the ingredient's manifest.	C2PA Claim Box
<code>assertion.undeclared</code>	An assertion was found in the ingredient's manifest that was not explicitly declared in the ingredient's claim.	C2PA Claim Box or C2PA Assertion
<code>assertion.inaccessible</code>	A non-embedded (remote) assertion was inaccessible at the time of validation.	C2PA Assertion
<code>assertion.notRedacted</code>	An assertion was declared as redacted in the ingredient's claim but is still present in the ingredient's manifest.	C2PA Assertion
<code>assertion.selfRedacted</code>	An assertion was declared as redacted by its own claim.	C2PA Claim Box
<code>assertion.action.redacted</code>	An <code>action</code> assertion was redacted when the ingredient's claim was created.	C2PA Assertion
<code>assertion.dataHash.mismatch</code>	Hash of a byte range of the asset does not match the hash declared in the data hash assertion.	C2PA Assertion
<code>assertion.bmffHash.mismatch</code>	Hash of a box-based asset does not match the hash declared in the BMFF hash assertion.	C2PA Assertion
<code>assertion.cloud-data.hardBinding</code>	A hard binding assertion is in a cloud data assertion.	C2PA Assertion

Value	Meaning	url Usage
<code>assertion.cloud-data.actions</code>	An update manifest contains a cloud data assertion referencing an actions assertion.	C2PA Assertion
<code>algorithm.unsupported</code>	The value of an <code>alg</code> header, or other header that specifies an algorithm used to compute the value of another field, is unknown or unsupported.	C2PA Claim Box or C2PA Assertion

17.12. Depthmap

A Depthmap assertion provides a 3D description of the scene being captured by a camera. A Depthmap assertion may contain a pre-computed [depth map](#), or data which can later be used to compute a depth map by downstream ingestion or viewing software (e.g., left/right stereo images).

All Depthmap assertions shall have a label that starts with `c2pa.depthmap` and be followed by a third section that identifies the type of depth map.

C2PA Depthmap assertions shall be captured optically, not inferred from a single 2D image via, for example, a machine learning model.

17.12.1. GDepth Depthmap

A GDepth depth map assertion leverages the well-established [GDepth format](#) to encode a pre-computed depth map.

A GDepth Depthmap assertion shall have a label of `c2pa.depthmap.GDepth`.

The schema for the data stored in this assertion should always mirror the schema at <https://developers.google.com/depthmap-metadata/reference>.

NOTE

There is no need to worry about splitting up the GDepth data when it grows beyond 64KB, as that limit existed in XMP to accommodate APP1 segment size limitations.

17.12.2. Schema and Example

The [CDDL Definition](#) for this type is:

```

; Assertion that encodes a GDepth-formatted 3D depth map of the captured scene
depthmap-gdepth-map = {
    "GDepth:Format": format-type, ; The format that describes how to convert the depthmap data
    into a valid float-point depthmap. Current valid values are 'RangeInverse' and 'RangeLinear'
    "GDepth:Near": float, ; The near value of the depthmap in depth units
    "GDepth:Far": float, ; The far value of the depthmap in depth units
    "GDepth:Mime": mime-type, ; The mime type for the base64 string describing the depth
    image content, e.g. 'image/jpeg' or 'image/png',
    "GDepth:Data": base64-string-type, ; The base64 encoded depth image. Please see GDepth
    encoding page at developers.google.com. The depthmap will be stretched-to-fit the
    corresponding color image
    ? "GDepth:Units": unit-type, ; The units of the depthmap, e.g. 'm' for meters or 'mm'
    for millimeters
    ? "GDepth:MeasureType": depth-meas-type, ; The type of depth measurement. Current valid
    values are 'OpticalAxis' and 'OpticRay'
    ? "GDepth:ConfidenceMime": confidence-mime-type, ; The mime type for the base64 string
    describing the confidence image content, e.g. 'image/png'.",
    ? "GDepth:Confidence": base64-string-type, ; The base64 encoded confidence image. Please
    see GDepth encoding page at developers.google.com. The confidence map should have the same
    size as the depthmap
    ? "GDepth:Manufacturer": tstr .size (1..max-tstr-length), ; The manufacturer of the
    device that created this depthmap
    ? "GDepth:Model": tstr .size (1..max-tstr-length), ; The model of the device that created
    this depthmap
    ? "GDepth:Software": tstr .size (1..max-tstr-length), ; The software that created this
    depthmap
    ? "GDepth:ImageWidth": float, ; The width in pixels of the original color image associated
    to this depthmap. This is NOT the depthmap width. If present, apps must update this property
    when scaling, cropping or rotating the color image. Clients use this property to verify the
    integrity of the depthmap w.r.t. the color image
    ? "GDepth:ImageHeight": float, ; The height in pixels of the original color image
    associated to this depthmap. This is NOT the depthmap height. If present, apps must update
    this property when scaling, cropping or rotating the color image. Clients use this property
    to verify the integrity of the depthmap w.r.t. the color image
    ? "metadata": $assertion-metadata-map, ; additional information about the assertion
}

base64-string-type = tstr

$mime-choice /= "image/jpeg"
$mime-choice /= "image/png"

mime-type = $mime-choice .default "image/jpeg"
confidence-mime-type = $mime-choice .default "image/png"

$format-choice /= "RangeInverse"
$format-choice /= "RangeLinear"

format-type = $format-choice .default "RangeInverse"

; Unit can be meter represented as "m" or could be millimeter represented as "mm"
$unit-choice /= "m"
$unit-choice /= "mm"
unit-type = $unit-choice .default "m"

$depth-meas-choice /= "OpticalAxis"
$depth-meas-choice /= "OpticRay"
depth-meas-type = $depth-meas-choice .default "OpticalAxis"

```

An example in CBOR Diagnostic Format ([.cbordiag](#)) is shown below:

```
{
  "GDepth:Format": "RangeInverse",
  "GDepth:Near": 29.3,
  "GDepth:Far": 878.7,
  "GDepth:Mime": "image/jpeg",
  "GDepth:Data": "hoOspQQ1lFTy/4Tp8Epx670E5QW5NwkNR+2b30KFXug=",
  "GDepth:Units": "mm",
  "GDepth:MeasureType": "OpticalAxis",
  "GDepth:ConfidenceMime": "image/png",
  "GDepth:Confidence": "acdbpQQ1lFTy/4Tp8Epx670E5QW5NwkNR+2b30KFXug=",
  "GDepth:Manufacturer": "CameraCompany",
  "GDepth:Model": "CameraCompany Shooter S1",
  "GDepth:Software": "Truepic Foresight Firmware for QC QRD8250 v0.01",
  "GDepth:ImageWidth": 32.2,
  "GDepth:ImageHeight": 43.6
}
```

As defined by the GDepth specification, the following fields shall be present in all GDepth depth map assertions

- GDepth:Format
- GDepth:Near
- GDepth:Far
- GDepth:Mime
- GDepth:Data

17.13. Exif Information

The **Exif** assertion can be used to ensure that Exif information, for example about the capture device, is added to the asset in a way that can be validated cryptographically. It is preferable to copy this information from the **Exif** block or from the [EXIF namespace](#) of the XMP block. Information should be copied to JSON-LD using the XMP field names specified in the [Exif 2.32 metadata for XMP](#) specification. When copying the information from XMP, the data shall be re-serialized according to the rules of the [JSON-LD serialization of XMP](#).

An Exif Information assertion shall have a label of **stds.exif**.

Any property from the [latest version of the Exif specification \(currently 2.32\)](#) can be added to the stds.exif assertion, with the exception of the MakerNote (37500, 0x927c) field, the contents of which are vendor-specific.

This assertion should be used to assert the precise location associated with the content using the Exif GPS properties (those starting **exif:GPS**). For broad information about the location, see [Section 17.14, “IPTC Photo Metadata”](#).

An example of an Exif assertion:

```
{
  "@context": {
    "dc": "http://purl.org/dc/elements/1.1/",
    "exifEX": "http://cipa.jp/exif/2.32/",
    "exif": "http://ns.adobe.com/exif/1.0/",
    "tiff": "http://ns.adobe.com/tiff/1.0/",
    "xmp": "http://ns.adobe.com/xap/1.0/",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  },
  "exif:GPSVersionID": "2.2.0.0",
  "exif:GPSLatitude": "39,21.102N",
  "exif:GPSLongitude": "74,26.5737W",
  "exif:GPSAltitudeRef": 0,
  "exif:GPSAltitude": "100963/29890",
  "exif:GPSTimeStamp": "2019-09-22T18:22:57Z",
  "exif:GPSSpeedRef": "K",
  "exif:GPSSpeed": "4009/161323",
  "exif:GPSImgDirectionRef": "T",
  "exif:GPSImgDirection": "296140/911",
  "exif:GPSDestBearingRef": "T",
  "exif:GPSDestBearing": "296140/911",
  "exif:GPSHPositioningError": "13244/2207",
  "exif:ExposureTime": "1/100",
  "exif:FNumber": 4.0,
  "exif:ColorSpace": 1,
  "exif:DigitalZoomRatio": 2.0,
  "tiff:Make": "CameraCompany",
  "tiff:Model": "Shooter S1",
  "exifEX:LensMake": "CameraCompany",
  "exifEX:LensModel": "17.0-35.0 mm",
  "exifEX:LensSpecification": { "@set": [ 1.55, 4.2, 1.6, 2.4 ] }
}
```

Although the redaction process works in such a way that only an entire assertion can be redacted (see [Section 5.7, “Redaction of Assertions”](#)), the use of an `update manifest` enables only partial redaction by removing the original and then placing the new, reduced, versions in the update manifest.

NOTE

For example, an assertion containing both location data and camera information which may need to have the location data redacted which could be done through an update manifest with a new `stds.exif` assertion with only the camera information.

17.14. IPTC Photo Metadata

The [International Press Telecommunications Council](#) defines a standard set of descriptive, administrative and rights metadata typically used by photographers, distributors, news organizations, archivists, and developers. Early versions of the standard used IPTC’s binary [Information Interchange Model \(IPTC-IIM\)](#) format and later versions introduced XMP support for all IIM fields, plus a new set of fields which can only be expressed in XMP. Together, these are called the [IPTC Photo Metadata Standard](#).

The IPTC Photo Metadata assertion can be used to ensure that IPTC Photo Metadata, for example describing ownership, rights and descriptive metadata about an image, is added to the asset in a way that can be validated

cryptographically.

It is preferable to copy this information from the IPTC-IIM or XMP block. Information should be copied to JSON-LD using the XMP field names specified in the [Exif 2.32 metadata for XMP](#) specification. When copying the information from XMP, the data shall be re-serialized according to the rules of the [JSON-LD serialization of XMP](#).

IPTC Photo Metadata properties can be copied to the C2PA assertion store as an assertion with the label `stds.iptc.photo-metadata`. The data should be in JSON-LD format using the XMP field names and structures specified in the [IPTC Photo Metadata Specification](#).

Any property from the [IPTC Photo Metadata Standard](#) can be added to the `stds.iptc.photo-metadata` assertion.

In particular, the Location Created property can be used to assert the broad location associated with an asset. To assert the precise GPS location associated with an asset, the [Exif information](#) assertion should be used.

An example of an IPTC Photo Metadata assertion including location information:

```
{
  "@context": {
    "Iptc4xmpCore": "http://iptc.org/std/Iptc4xmpCore/1.0/xmlns/",
    "Iptc4xmpExt": "http://iptc.org/std/Iptc4xmpExt/2008-02-29/",
    "dc": "http://purl.org/dc/elements/1.1/",
    "photoshop": "http://ns.adobe.com/photoshop/1.0/",
    "plus": "http://ns.useplus.org/ldf/xmp/1.0/",
    "xmp": "http://ns.adobe.com/xap/1.0/",
    "xmpRights": "http://ns.adobe.com/xap/1.0/rights/"
  },
  "photoshop:DateCreated": "Aug 31, 2021", ①
  "dc:creator": [ "Julie Smith" ], ②
  "dc:rights": "Copyright (C) 2021 Example Photo Agency. All Rights Reserved." ③
  "photoshop:Credit": [ "Julie Smith/Example Photo Agency via Example Distributor" ], ④
  "plus:licensor": [ ⑤
    {
      "plus:LicensorName": "Example Photo Agency",
      "plus:LicensorURL": "http://examplephotoagency.com/images/"
    }
  ],
  "xmpRights:WebStatement": "http://examplephotoagency.com/terms.html", ⑥
  "xmpRights:UsageTerms": [
    "Not for online publication. Germany OUT" ⑦
  ],
  "Iptc4xmpExt:LocationCreated": { ⑧
    "Iptc4xmpExt:City": "San Francisco"
  },
  "Iptc4xmpExt:PersonInImage": [ ⑨
    "Erika Fictional"
  ]
}
```

① Date Created

② Creator

③ Copyright Notice

④ Credit Line

⑤ Licensor

⑥ Web Statement of Rights

⑦ Rights Usage Terms

⑧ Location Created

⑨ Person Shown in the Image

Although the redaction process works in such a way that only an entire assertion can be redacted (see [Section 5.7, “Redaction of Assertions”](#)), the use of an [update manifest](#) enables only partial redaction by removing the original and then placing the new, reduced, versions in the update manifest.

NOTE

For example, an assertion containing location properties, creator name and other information may need to have the creator name redacted which could be done through an update manifest with a new [stds.iptc.photo-metadata](#) assertion with the other information still present.

17.15. Use of Schema.org

[Schema.org](#) is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet. As such, they provide numerous JSON-based grammars that one may wish to include as an assertion. When using a schema.org type as a C2PA assertion, we require a full JSON-LD serialisation. As such, the top-level [@context](#) field must have a value of <http://schema.org>.

All Schema.org assertions shall have a label that starts with [schema.org](#) and be followed by the name of the schema that is being used. For example, [schema.org.ImageObject](#) can be used to include something extra about an image. Since schemas on [Schema.org](#) are not versioned as described [here](#), no version indicator is used in their labels.

17.15.1. Claim Review

A schema that is used by the publishing community for web sites has [recently also been introduced for images](#). A [ClaimReview](#) is used as 'A fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed).'

A ClaimReview assertion shall have a label of [stds.schema-org.ClaimReview](#).

Inside the assertion (which, as described previously, is serialised as JSON-LD), the top-level [@type](#) field should be set to a value of [ClaimReview](#).

The "Claim" in ClaimReview refers to a "claim made (or reported) in some creative work", not the C2PA claim block. ClaimReview assertions should be used as a way to mark the current asset as a review of the claims of another work, not any specific part of the C2PA manifest. To provide a review of a specific assertion, use an [assertion metadata assertion](#).

NOTE

An example: an image infographic has been assembled that fact-checks some other claim that is currently in public distribution. The image infographic includes a ClaimReview assertion which references the text of the claim its reviewing, a fact-checking rating of the claim, and zero to many links to other works (e.g., articles, videos, images) that are making that claim.

The reviewRating property inside the ClaimReview assertion is of type [Rating](#). Depending on how your organisation does fact-checking reviews, you may have a numerical rating, a true or false, or some other textual description of how the asset has reviewed a claim. <https://www.claimreviewproject.com/user-guide> provides a good guide on how to use the ClaimReview vocabulary.

A partial [JSON Schema](#) for this type is:

This schema does not validate all aspects of the referenced schema.org types (just the top-level ClaimReview and Review properties). Please refer to <https://schema.org> for the details on referenced types, and <https://json-ld.org/> for how to reference them in the assertion.

```
{  
  "$schema": "http://json-schema.org/draft-07/schema",  
  "$id": "http://ns.c2pa.org/claim_review/v1",  
  "type": "object",  
  "description": "Assertion that describes a fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed).",  
  "examples": [  
    {  
      "@context": "http://schema.org",  
      "@type": "ClaimReview",  
      "claimReviewed": "The world is flat",  
      "reviewRating": {  
        "@type": "Rating",  
        "ratingValue": "1",  
        "bestRating": "5",  
        "worstRating": "1",  
        "ratingExplanation": "The world is not flat",  
        "alternateName": "False"  
      },  
      "itemReviewed": {  
        "@type": "CreativeWork",  
        "author": {  
          "@type": "Person",  
          "name": "A N Other"  
        },  
        "headline": "Earth: Flat."  
      }  
    }  
  ],  
  "anyOf": [  
    {  
      "$ref": "https://json.schemastore.org/schema-org-thing.json"  
    }  
  ]  
}
```

```

        },
    ],
    "properties": {
        "claimReviewed": {
            "type": "string",
            "minLength": 1,
            "description": "A short summary of the specific claims reviewed in a ClaimReview."
        },
        "reviewAspect": {
            "type": "string",
            "minLength": 1,
            "description": "From http://schema.org/Review: This Review or Rating is relevant to this part or facet of the itemReviewed."
        },
        "reviewBody": {
            "type": "string",
            "minLength": 1,
            "description": "From http://schema.org/Review: The actual body of the review."
        },
        "reviewRating": {
            "description": "From http://schema.org/Review: The rating given in this review. Note that reviews can themselves be rated. The reviewRating applies to rating given by the review. The aggregateRating property applies to the review itself, as a creative work.",
            "anyOf": [
                {
                    "$ref": "https://json.schemastore.org/jsonld#"
                }
            ]
        },
        "itemReviewed": {
            "description": "From http://schema.org/Review: The item that is being reviewed/rated.",
            "anyOf": [
                {
                    "$ref": "https://json.schemastore.org/jsonld#"
                }
            ]
        }
    }
}

```

17.15.2. Creative Work

Schema.org provides a well-known and well-deployed set of types and metadata fields. One of the core types is [CreativeWork](#), which is intended to describe any representation of creative effort. This assertion allows an asserter to provide various pieces of information about the asset, including who they are, and the date/time of publication.

A Creative Work assertion shall have a label of [stds.schema-org.CreativeWork](#).

Inside the assertion (which, as described previously, is serialised as JSON-LD), the top-level `@type` field should be set to a value of [CreativeWork](#). The JSON-LD document's root subject is the bound asset of the claim that this assertion is part of.

A partial JSON Schema for this type is:

NOTE

This schema does not validate all aspects of the schema.org types, a valid CreativeWork assertion is free to use any types or vocabulary specified by schema.org.

```
{  
  "$schema": "http://json-schema.org/draft-07/schema",  
  "$id": "http://ns.c2pa.org/creative_work/v1",  
  "type": "object",  
  "description": "Assertion that describes the most generic kind of creative work, including books, movies, photographs, software programs, etc.",  
  "examples": [  
    {  
      "@context": [  
        "http://schema.org/",  
        {  
          "credential": null  
        }  
      ],  
      "@type": "CreativeWork",  
      "datePublished": "2021-05-20T23:02:36+00:00",  
      "publisher": {  
        "name": "BBC News",  
        "publishingPrinciples": "https://www.bbc.co.uk/news/help-41670342",  
        "logo": "https://m.files.bbci.co.uk/modules/bbc-morph-news-waf-page-meta/5.1.0/bbc_news_logo.png",  
        "parentOrganization": {  
          "name": "BBC",  
          "legalName": "British Broadcasting Corporation"  
        }  
      },  
      "url": "https://www.bbc.co.uk/news/av/world-europe-57194011",  
      "identifier": "p09j7vzy",  
      "producer": {  
        "identifier": "https://en.wikipedia.org/wiki/Joe_Bloggs",  
        "name": "Joe Bloggs",  
        "credential": [  
          {  
            "url": "self#jumbf=c2pa/urn:uuid:F9168C5E-CEB2-4faa-B6BF-  
329BF39FA1E4/c2pa.credentials/Joe_Bloggs",  
            "alg": "sha256",  
            "hash": "Auxjtmax46cC2N3Y9aFmB09Jfay8LEwJWzBUTZ0sUM8gA"  
          }  
        ]  
      },  
      "copyrightHolder": {  
        "name": "BBC",  
        "legalName": "British Broadcasting Corporation"  
      },  
      "copyrightYear": 2021,  
      "copyrightNotice": "Copyright © 2021 BBC."  
    }  
  ],  
  "anyOf": [  
    {  
      "$ref": "https://json.schemastore.org/schema-org-thing.json"  
    }  
  ],  
  "properties": {  
    "datePublished": {  
      "type": "string",  
      "minLength": 1,  
      "maxLength": 256  
    }  
  }  
}
```

```

    "format": "date-time",
    "description": "From http://schema.org/CreativeWork: Date of first
broadcast/publication."
},
"publisher": {
    "description": "From http://schema.org/CreativeWork: The publisher of the creative
work.",
    "$ref": "#/definitions/personOrgOrUri"
},
"identifier": {
    "$ref": "#/definitions/credential"
},
"producer": {
    "description": "From http://schema.org/CreativeWork: The person or organization who
produced the work (e.g. music album, movie, tv/radio series etc.).",
    "$ref": "#/definitions/personOrgOrUri"
},
"creator": {
    "description": "From http://schema.org/CreativeWork: The creator/author of this
CreativeWork.",
    "$ref": "#/definitions/personOrgOrUri"
},
"editor": {
    "description": "From http://schema.org/CreativeWork: Specifies the Person who edited
the CreativeWork.",
    "anyOf": [
        {
            "$ref": "#/definitions/person"
        },
        {
            "$ref": "#/definitions/uri"
        }
    ]
},
"contributor": {
    "description": "From http://schema.org/CreativeWork: A secondary contributor to the
CreativeWork or Event.",
    "$ref": "#/definitions/personOrgOrUri"
},
"copyrightHolder": {
    "description": "From http://schema.org/CreativeWork: The party holding the legal
copyright to the CreativeWork.",
    "$ref": "#/definitions/personOrgOrUri"
},
"copyrightNotice": {
    "type": "string",
    "minLength": 1,
    "description": "From http://schema.org/CreativeWork: Text of a notice appropriate for
describing the copyright aspects of this Creative Work, ideally indicating the owner of the
copyright for the Work."
},
"copyrightYear": {
    "type": "integer",
    "description": "From http://schema.org/CreativeWork: The year during which the claimed
copyright for the CreativeWork was first asserted."
}
},
"definitions": {
    "credential": {
        "type": "array",
        "description": "Verifiable Credentials (VC) associated with this actor.",
        "minItems": 1,
        "items": {
            "type": "object"
        }
    }
}

```

```

  "oneOf": [
    {
      "$ref": "http://ns.c2pa.org/hashed-uri/v1"
    },
    {
      "$ref": "http://ns.c2pa.org/hashed-ext-uri/v1"
    }
  ]
},
"organization": {
  "type": "object",
  "description": "From http://schema.org/Organization: An organization such as a school, NGO, corporation, club, etc.",
  "properties": {
    "name": {
      "type": "string",
      "minLength": 1,
      "description": "The name of the Organization."
    },
    "legalName": {
      "type": "string",
      "minLength": 1,
      "description": "From http://schema.org/Organization: The official name of the organization, e.g. the registered company name."
    },
    "publishingPrinciples": {
      "type": "string",
      "minLength": 1,
      "format": "uri",
      "description": "From http://schema.org/Organization: The publishingPrinciples property indicates (typically via URL) a document describing the editorial principles of an Organization (or individual e.g. a Person writing a blog) that relate to their activities as a publisher, e.g. ethics or diversity policies."
    },
    "logo": {
      "type": "string",
      "minLength": 1,
      "format": "uri",
      "description": "From http://schema.org/Organization: An associated logo."
    },
    "parentOrganization": {
      "description": "From http://schema.org/Organization: The larger organization that this organization is a subOrganization of, if any.",
      "$ref": "#/definitions/organization"
    },
    "identifier": {
      "$ref": "#/definitions/credential"
    },
    "credential": {
      "$ref": "#/definitions/credential"
    }
  }
},
"person": {
  "type": "object",
  "description": "From http://schema.org/Organization: An organization such as a school, NGO, corporation, club, etc.",
  "properties": {
    "name": {
      "type": "string",
      "minLength": 1,
      "description": "The name of the Person."
    }
  }
}

```

```

},
"givenName": {
  "type": "string",
  "minLength": 1,
  "description": "From http://schema.org/Person: Given name. In the U.S., the first
name of a Person."
},
"familyName": {
  "type": "string",
  "minLength": 1,
  "description": "Family name. In the U.S., the last name of a Person."
},
"identifier": {
  "$ref": "#/definitions/credential"
},
"credential": {
  "$ref": "#/definitions/credential"
}
},
"uri": {
  "type": "string",
  "minLength": 1,
  "format": "uri"
},
"identifier": {
  "type": "string",
  "minLength": 1,
  "description": "The identifier property represents any kind of identifier for any kind
of Thing (in this case, an identifier for this CreativeWork)."
},
"personOrgOrUri": {
  "anyOf": [
    {
      "$ref": "#/definitions/organization"
    },
    {
      "$ref": "#/definitions/person"
    },
    {
      "$ref": "#/definitions/uri"
    }
  ]
}
}
}

```

It is possible to associate identifying information about an actor with their specific role in the Creative Work as described [below](#).

17.16. Common Data Model: Actor

An `actor` object references a particular person or organisation. A valid actor object shall have either:

1. an `identifier` field containing an identifier, or,

NOTE

Since this field is a URL, it is possible to use various identity schemes such as [OpenID](#), [WebID](#) or [ISNI](#).

2. a **credentials** or **credential** field containing at least one **hashed_uri** to a [W3C Verifiable Credential](#) in the [Credential Store](#) that is associated with the actor, or,

NOTE

Most references to a W3C Verifiable Credentials are via a **credentials** field, the Creative Work assertion uses a **credential** field in order to align with the [schema.org](#) model of singular field names.

3. both (in which case, the **id** field in the **credentialSubject** W3C Verifiable Credential object should match the **identifier** field of the actor)

Chapter 18. Open Topics

There are a number of open topics that have not yet been addressed in this public review draft specification. They are briefly described below, organized by the relevant section of the specification.

18.1. Assertions

- Support for marking a previous claim or assertion as "retracted", a feature commonly used in the news publication industry.

18.2. Binding to Content

- Add support for the binding of live video streams, 3D formats and audio formats.
- Addressing the case where servers make real-time modifications to content streamed to a client, potentially altering the cryptographic hash used to bind ISO BMFF content to a manifest.
- Determining whether the box exclusion list used for hashing a BMFF-formatted asset in the fragmented MP4 case needs to support having each fragment have its own exclusion list (e.g. if the subset to hash must be fragment-specific).

18.3. Trust Model

- Supporting additional credential types for use in signing manifests.

18.4. Validation

- Design for how a video player can communicate with the media validator to indicate that a discontinuity in video playback is expected, perhaps due to seek, fast forward, or other types of "trick play".

18.5. User Experience

- Working with the W3C and browser vendors to define a standardized model for exposing the provenance data to user agents.

Chapter 19. Patent Policy

The C2PA has adopted an open standard patent policy via W3C's Patent Mode (2004):

Licensing Commitment. For materials other than source code or datasets developed by the Working Group, each Working Group Participant agrees to make available any of its Essential Claims, as defined in the W3C Patent Policy (available at <http://www.w3.org/Consortium/Patent-Policy-20040205>), under the W3C RF licensing requirements Section 5 (<http://www.w3.org/Consortium/Patent-Policy-20040205>), in Approved Deliverables adopted by that Working Group as if that Approved Deliverable was a W3C Recommendation. Source code developed by the Working Group is subject to the license set forth in the Working Group charter.

For Exclusion. Prior to the adoption of a Draft Deliverable as an Approved Deliverable, a Working Group Participant may exclude Essential Claims from its licensing commitments under this agreement by providing written notice of that intent to the Working Group chair (“Exclusion Notice”). The Exclusion Notice for issued patents and published applications must include the patent number(s) or title and application number(s), as the case may be, for each of the issued patent(s) or pending patent application(s) that the Working Group Participant wishes to exclude from the licensing commitment set forth in Section 1 of this patent policy. If an issued patent or pending patent application that may contain Essential Claims is not set forth in the Exclusion Notice, those Essential Claims shall continue to be subject to the licensing commitments under this agreement. The Exclusion Notice for unpublished patent applications must provide either: (i) the text of the filed application; or (ii) identification of the specific part(s) of the Draft Deliverable whose implementation makes the excluded claim an Essential Claim. If (ii) is chosen, the effect of the exclusion will be limited to the identified part(s) of the Draft Deliverable. The Working Group Chair will publish Exclusion Notices