Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

C2PA Content Credentials: Manifests, Claims, Signatures

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · c2pa · content credentials · c2pa specification · content authenticity initiative · c2pa manifest

Original article: https://reliablecontext.com/articles/c2pa-content-credentials-explained-manifests-claims-signatures



In this article

1. The problem: edited and synthetic media without a record
2. Architecture: asset, manifest store, claim, assertions, signature
3. Hard bindings and soft bindings
4. Trust model: signers, trust lists and validation states
5. What validation proves, and what it does not
6. Worked example: a fictional press image
7. Relation to W3C PROV and Verifiable Credentials 2.0
8. Regulatory driver: EU AI Act Article 50(2)
9. Deployment checklist
10. What C2PA does not do for facts in text
11. Sources

The problem: edited and synthetic media without a record

Most images, audio and video reach their audience without a reliable history. Ordinary file metadata (EXIF or XMP fields for camera, time or location) can be rewritten without trace, and generative models now produce media that no camera recorded, at a volume no review desk can inspect.

The Coalition for Content Provenance and Authenticity (C2PA) publishes an open technical standard, branded Content Credentials, for establishing the origin and edits of digital content (C2PA). It is a standards body under the Linux Foundation's Joint Development Foundation, formed by Adobe, Arm, Intel, Microsoft and Truepic, and it unifies Adobe's Content Authenticity Initiative (CAI) with Project Origin, led by Microsoft and the BBC (CAI, How it works).

This article follows the current specification, version 2.5, dated October 2026 in its Version History appendix (C2PA Technical Specification 2.5, Appendix F.1). Its central distinction: a credential establishes who signed which statements about which bytes. It does not establish that the statements, or the scene, are true.

Architecture: asset, manifest store, claim, assertions, signature

All definitions below are from the glossary in Section 2 of version 2.5 (C2PA spec 2.5, §2).

- **Asset**: a file or stream containing digital content, asset metadata and, optionally, a C2PA Manifest.
- **Assertion**: a data structure representing a statement about the asset, either made by the signer or gathered at claim-generation time.
- **Claim**: a digitally signed, tamper-evident data structure that references a set of assertions and the information needed to represent the content binding; redactions are declared in it.
- **Claim signature**: the digital signature on the claim, made with the signer's private key.
- **C2PA Manifest**: one or more assertions (including content bindings), a single claim and a claim signature. A manifest can reference other manifests.
- **C2PA Manifest Store**: a collection of manifests, embedded in the asset or external to it.

- **Content Credential**: the preferred non-technical term for a C2PA Manifest.

The **claim generator** is the hardware or software that produces the claim and signature; the specification states that it is not the signer and does not take responsibility for the content of the claim. The **signer** holds the private key and is identified by the subject of its credential (C2PA spec 2.5, §2.1). The store is a JUMBF superbox (ISO 19566-5:2023); its last manifest is the **active manifest**, the one whose bindings are validated against the asset as it now exists (C2PA spec 2.5, §11.2). Earlier manifests remain, so each edit appends to the record.

Created and gathered assertions

The claim lists assertions in two arrays, and the distinction governs interpretation. `created_assertions` is mandatory, must include a hard-binding assertion in a standard manifest, and every entry is attributed to the signer. `gathered_assertions` holds assertions supplied by other components of the workflow; they are part of the claim but are explicitly not attributed to the signer, with human-entered information as the specification's example (C2PA spec 2.5, §10.2.2).

Ordinary metadata gets the same treatment. Metadata outside a signed `created_assertions` entry is not attributed to the signer even if the hard binding covers it: the specification's example is that the signer is not responsible for an EXIF GPS location unless it is restated in a `c2pa.metadata` assertion under `created_assertions` (C2PA spec 2.5, §12.2.2.9).

Actions and ingredients

Every standard manifest carries at least one `c2pa.actions` assertion. An asset created de novo (camera capture, File > New, or generative model output) must start with `c2pa.created` plus a `digitalSourceType` value; the specification's example for a video generated from a text prompt is the IPTC term `trainedAlgorithmicMedia`. An edited asset starts with `c2pa.opened`, which must reference an ingredient assertion (C2PA spec 2.5, §16.4.1.2).

An **ingredient** (`c2pa.ingredient.v3`) is an asset used in creating, or otherwise related to, the current asset, with a reference to its own provenance data. Its `relationship` is `parentOf` (the current asset derives from it), `componentOf` (one part of a composition) or `inputTo` (input to a computational process such as an AI/ML model) (C2PA spec 2.5, §16.5). Ingredients turn provenance into a graph rather than a single label.

Hard bindings and soft bindings

A manifest must be tied to specific content, or it could be copied onto any file. The specification defines two kinds of **content binding** (C2PA spec 2.5, §12.2):

Property	Hard binding	Soft binding
Definition	One or more cryptographic hashes uniquely identifying the asset or a portion of it	A content identifier that is not statistically unique (fingerprint) or is embedded as an invisible watermark
Computed over	Raw bytes, with declared exclusions	Digital content, not raw bits

Property	Hard binding	Soft binding
A match shows	This manifest belongs to this asset and covered bytes are unmodified	Probably the same or derived content
Survives re-encoding	No, by design	Intended to
Per manifest	At most one	Zero or more
Labels	`c2pa.hash.data`, `c2pa.hash.boxes` (JPEG, PNG, GIF), `c2pa.hash.bmff.v3` (MP4), plus collection, segment and multi-asset variants	`c2pa.soft-binding`

A hard binding produces values that match only this asset, not even its renditions; a soft binding exists to identify derived assets and renditions, and the specification forbids using one as a hard binding (C2PA spec 2.5, §12.2.1, §12.2.3). The data hash covers all bytes except ordered, non-overlapping exclusion ranges (typically where the manifest store sits); the BMFF hash uses box exclusion lists and, for fragmented MP4, Merkle trees (C2PA spec 2.5, §16.3.1, §16.3.2). Both are forms of content addressing: change the bytes and the identifier changes.

Soft bindings address the commonest failure. An asset "can become separated from its C2PA Manifest due to removal or corruption of asset metadata"; a fingerprint or watermark then lets a validator find the manifest in a **manifest repository**. A credential discoverable this way is a **Durable Content Credential** (C2PA spec 2.5, §2.4).

Trust model: signers, trust lists and validation states

Trust rests on "the identity of the signer associated with the cryptographic signing key used to sign a claim"; with trusted time stamps, signatures can be re-validated indefinitely to show they were made while the credential was valid and unrevoked (C2PA spec 2.5, §1.3.1). Each COSE signature must carry exactly one identity credential (C2PA spec 2.5, §13.2). Version 2.0 (January 2024) restricted signing to X.509 certificates, shifted trust from identifying humans or organisations to identifying hardware and software components, and removed the earlier W3C Verifiable Credentials section (C2PA spec 2.5, Appendix F.6).

Trust lists. For the `c2pa-kp-claimSigning` Extended Key Usage, a validator's trust anchors must include the **C2PA Trust List**, a C2PA-managed list of X.509 anchors that issue certificates to hardware and software signers; a separate **C2PA TSA Trust List** covers time-stamping authorities. Validators should let users add anchors and may keep a private store of individually trusted signer credentials, populated only at the user's request (C2PA spec 2.5, §2.3.16-2.3.17, §13.4).

Conformance program. Opened in June 2025 and now at v0.2, the C2PA Conformance Program assesses generator products, validator products and certification authorities against the specification, a Certificate Policy and Security Requirements, and publishes a Conforming Products List alongside both trust lists (C2PA Conformance). Products asserting as themselves are advised to sign with a certificate issued through it (C2PA Human and Organizational Identity Recommendation).

Validation states are nested (C2PA spec 2.5, §13.3):

- **Well-Formed**: the manifest, its assertions and its ingredients meet the normative requirements.
- **Valid**: Well-Formed, unmodified since signing, signature validated and within its validity period, credential not revoked. The claim can then be attributed to the claim generator in `claim_generator_info`.
- **Trusted**: Valid, and the signing credential chains to a configured trust anchor.

An asset is Valid when its active manifest is Valid or Trusted and the bound content is unmodified.

What validation proves, and what it does not

The scope section quotes the C2PA guiding principles: C2PA specifications "DO NOT provide value judgements 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" (C2PA spec 2.5, §1.2). The consumer, outside the trust model, uses signer identity and other signals "to decide whether the assertions made about an asset are true" (C2PA spec 2.5, §13.1). The CAI likewise says Content Credentials are not intended to indicate prescriptively whether content is "real" (CAI, How it works).

A Trusted manifest on a Valid asset establishes that a key chaining to an accepted anchor signed the claim; that the claim, its assertions and the hard-bound bytes are unchanged since; and that `created_assertions` are attributable to that signer. It does not establish:

- that the depicted scene happened or a caption is accurate;
- that a human stands behind the claim (the core model identifies machines);
- that the signer was honest: a dishonest signer with a valid certificate yields a valid credential;
- that unsigned metadata or `gathered_assertions` are correct;
- anything, once a manifest is stripped and no soft binding allows recovery. A missing credential is absence of evidence, not evidence of manipulation.

Worked example: a fictional press image

Halden Therapeutics AG is a fictional pharmaceutical company. Its communications team publishes a photograph of a new fill-finish line. The workflow is invented; the labels are real C2PA 2.5 labels.

Step	Manifest and signer (fictional)	Key assertions	Binding	What a validator can report
1. Capture	M1, camera's claim generator, manufacturer certificate	`c2pa.created`, `digitalSourceType` digital capture; `c2pa.thumbnail.claim`	`c2pa.hash.boxes`	This exact JPEG was declared captured by that device's credential
2. Edit	M2, editing service certificate	`c2pa.opened`, `c2pa.cropped`; ingredient `parentOf` M1	New hard binding	Crop declared; M1 re-validated as ingredient

Step	Manifest and signer (fictional)	Key assertions	Binding	What a validator can report
3. Publish	M3 (active), Halden's publishing system	`c2pa.opened`, `c2pa.resized`, `c2pa.watermarked.bound`, `c2pa.published`; `c2pa.soft-binding`; ingredient `parentOf` M2	Hard plus soft binding	Chain M3 → M2 → M1 intact; watermark supports lookup
Caption	None	"Halden's new line, Basel, March 2026"	None	Nothing: the caption is outside every signature

If a reseller re-encodes the image and drops the manifest store, the hard binding cannot match the new bytes. A surviving watermark lets a validator query a repository, recover M3 and present it as describing a related rendition, not the file in hand.

Had marketing instead generated an "artist's impression", the model provider's manifest would start with `c2pa.created` and `trainedAlgorithmicMedia`. Version 2.4 also added a `c2pa.ai-disclosure` assertion recording model type and, optionally, model name and identifier (C2PA spec 2.5, §16.8.2, Appendix F.2). Either way the credential records how the image was made, not whether the line looks like that.

Relation to W3C PROV and Verifiable Credentials 2.0

W3C PROV. PROV-DM (W3C Recommendation, 30 April 2013) defines provenance as "information about entities, activities, and people involved in producing a piece of data or thing", with entity, activity, agent and relations such as derivation as core concepts (PROV-DM; see W3C PROV). The mapping is direct: asset or ingredient to entity, action to activity, signer or claim generator to agent, `parentOf` to derivation. The layers differ. PROV-DM is a conceptual interchange model with no hashing or signature mechanism; C2PA is a container format with mandatory hard bindings and claim signatures, bound to particular files. A PROV graph describes a history; a manifest store makes one asset's history tamper-evident and attributable.

Verifiable Credentials 2.0. The VC Data Model v2.0 (W3C Recommendation, 15 May 2025) defines a verifiable credential as "a tamper-evident credential whose authorship can be cryptographically verified", within an issuer-holder-verifier ecosystem (VC Data Model 2.0). A VC makes claims about a subject; a C2PA claim is hard-bound to an asset's bytes. Since C2PA 2.0 dropped built-in VC support, VCs enter Content Credentials through the Creator Assertions Working Group (CAWG) of the Decentralized Identity Foundation, which the C2PA recommends for human and organisational identity (C2PA Identity Recommendation). The CAWG identity assertion 1.3 (ratified 17 August 2026) has a credential holder sign hashlinks to other assertions in the manifest; its "identity claims aggregation" type is a W3C verifiable credential, issued by an aggregator and secured with COSE, that binds verified signals such as websites or social accounts to the specific asset (CAWG Identity Assertion 1.3, §8.1).

Both models share one limit. VC 2.0 separates verification of authorship and integrity from validation against a verifier's business rules, whose means "vary widely and are outside the scope of this specification" (VC Data Model 2.0, §2 Terminology). Neither judges whether a statement is true.

Regulatory driver: EU AI Act Article 50(2)

This section is educational, not legal advice. In the EU, Article 50(2) of the AI Act requires providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text to ensure outputs are "marked in a machine-readable format and detectable as artificially generated or manipulated", with solutions effective, interoperable, robust and reliable as far as technically feasible; assistive functions for standard editing are exempt (Regulation (EU) 2024/1689, Art. 50(2)). Recital 133 names "watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints" (Regulation (EU) 2024/1689, recital 133). These correspond to C2PA's soft bindings, manifests and hard bindings, but the Act names no standard.

Article 50 applies from 2 August 2026. The Digital Omnibus on AI, Regulation (EU) 2026/1744, adds Article 111(4): providers whose generative systems were placed on the market before 2 August 2026 must comply with Article 50(2) by 2 December 2026 (Regulation (EU) 2026/1744, Article 1, point (39) and recital 38). Commission guidelines on Article 50 state that compliance may be demonstrated through a Code of Practice on Transparency of AI-generated Content or equivalently adequate alternative means (European Commission, Guidelines on transparency obligations). Whether a given C2PA implementation meets Article 50(2) is a question for that framework, not one the specification answers.

Deployment checklist

1. Choose a claim-signing certificate that your audience's validators accept; check the C2PA Trust List and conforming-products status.
2. Match the hard binding to the format: box hash for JPEG and PNG, `c2pa.hash.bmff.v3` for MP4, data hash otherwise.
3. Put facts you stand behind in `created_assertions`; leave unverified input in `gathered_assertions`.
4. Time-stamp signatures via a TSA on the C2PA TSA Trust List.
5. Add soft bindings and a manifest repository where channels strip metadata.
6. Record ingredients for composites and AI inputs.
7. Display Well-Formed, Valid and Trusted distinctly; never present "no credential" as "fake".

What C2PA does not do for facts in text

C2PA 2.5 binds manifests to unstructured and structured text (C2PA spec 2.5, §12.2.2.4-12.2.2.5), proving who signed a document and that its bytes are unchanged. It does not link an individual statement (a dose, a date, a revenue figure) to the source supporting it. Statement-level provenance requires each value to carry its own source reference, as in signed RDF exports with graph evidence. Reliable Context is a system of record built around values that carry their sources.

Sources

1. Content Credentials: C2PA Technical Specification, version 2.5. Coalition for Content Provenance and Authenticity, October 2026. https://spec.c2pa.org/specifications/specifications/2.5/specs/C2PA_Specification.html
2. C2PA home page. Coalition for Content Provenance and Authenticity. https://c2pa.org/
3. C2PA Conformance Program. Coalition for Content Provenance and Authenticity. https://c2pa.org/conformance/
4. Human and Organizational Identity Recommendation (2.4). Coalition for Content Provenance and Authenticity, 2026. https://spec.c2pa.org/specifications/specifications/2.4/identity/identity.html
5. How it works. Content Authenticity Initiative. https://contentauthenticity.org/how-it-works
6. Identity Assertion, version 1.3. Creator Assertions Working Group, Decentralized Identity Foundation, 17 August 2026. https://cawg.io/identity/1.3/
7. Verifiable Credentials Data Model v2.0, W3C Recommendation. W3C, 15 May 2025. https://www.w3.org/TR/vc-data-model-2.0/
8. PROV-DM: The PROV Data Model, W3C Recommendation. W3C, 30 April 2013. https://www.w3.org/TR/prov-dm/
9. Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union, 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
10. Regulation (EU) 2026/1744 (Digital Omnibus on AI). Official Journal of the European Union, 24 July 2026. https://eur-lex.europa.eu/eli/reg/2026/1744/oj
11. Guidelines on transparency obligations for providers and deployers of certain AI systems (policy page). European Commission, last updated 6 August 2026. https://digital-strategy.ec.europa.eu/en/policies/guidelines-transparency-ai-generated-content

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://c2pa.org
2. https://contentauthenticity.org/how-it-works
3. https://spec.c2pa.org/specifications/specifications/2.5/specs/C2PA_Specification.html
4. https://c2pa.org/conformance
5. https://spec.c2pa.org/specifications/specifications/2.4/identity/identity.html
6. https://www.w3.org/TR/prov-dm
7. https://www.w3.org/TR/vc-data-model-2.0
8. https://cawg.io/identity/1.3
9. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
10. https://eur-lex.europa.eu/eli/reg/2026/1744/oj
11. https://digital-strategy.ec.europa.eu/en/policies/guidelines-transparency-ai-generated-content

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.