



vera.ai

vera.ai: VERification Assisted by Artificial Intelligence

# Content provenance and authentication for trusted content with C2PA

Hans Hoffmann, Antonio Arcidiacono, Mohamed Badr Taddist,  
Lalya Gaye, Lucille Verbaere – European Broadcasting Union (EBU)  
30 July 2025

This document reports on the emerging technology C2PA and its relevance to vera.ai, with technical descriptions and use cases.

Keywords: content provenance, verification, authenticity, C2PA



vera.ai is a Horizon Europe Research and Innovation Project co-financed by the European Union under Grant Agreement ID: 101070093, an Innovate UK grant 10039055 and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00245.  
The content of this document is © of the author(s) and respective referenced sources. For further information, visit [veraai.eu](https://veraai.eu).

## Glossary

---

Abbreviation	Meaning
AFP	Agence France-Presse
AI	Artificial Intelligence
C2PA	Coalition for Content Provenance and Authentication
CAI	Content Authenticity Initiative
CAs	Certificate Authorities
DASH	Dynamic Adaptive Streaming over HTTP
DBKF	Database of Known Fakes
EXIF	Exchangeable Image File Format
GenAI	Generative Artificial Intelligence
HLS	HTTP Live Streaming
IFCN	International Fact-Checking Network
IPTC	International Press Telecommunications Council
ISO	International Organization for Standardization
ISO-BMFF	ISO Base Media File Format
JUMBF	JPEG Universal Metadata Box Format
KSE	Keyframe Selection and Enhancement
L1-4	Level 1-4
ML	Machine Learning
MXF	Material Exchange Format
PoC	Proof of Concept
SMPTE	Society of Motion Picture and Television Engineers
TSAs	Time-stamp Authorities
TWG	Technical Working Group
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
W3C	World Wide Web Consortium



## Table of Contents

Executive Summary	5
1. Introduction	6
1.1 The Challenge of Synthetic Media in Detection and Authentication	6
1.2 C2PA as an Authentication Tool in the Verification Process	7
1.3 The Relevance of C2PA to vera.ai	8
2. What is C2PA?	8
2.1 How Does C2PA Work?	8
2.1.1 Content Credentials	8
2.1.2 C2PA AI Tags	11
2.1.3 C2PA: Training and Data Mining/Scraping	12
2.1.4 Soft Binding	13
2.2 How Does C2PA Interface with Users?	14
2.2.1 End-User Experience	14
2.2.2 Levels of Information Disclosure	16
3. Bringing C2PA into the World of Media	20
3.1 Standardisation Efforts	20
3.2 Governance Structure	21
3.3 Assessments of C2PA Impact by the BBC	22
3.4 Current Open Issues and Further Work	23
3.4.1 Streaming and Live Streaming	23
3.4.2 Interoperability	23
3.4.2.1 Open-Source Implementation	24
3.4.2.2 Industry Examples of C2PA Interoperability	28
3.4.3 Limitations and Further Work	31
4. Why Use C2PA?	32
4.1 To Increase Trust in Publishers' Content	32
4.2 To Label AI-Generated Content	36
5. C2PA in Practice	40
5.1 User Scenario: John the Fact-Checker	41
5.1.1 Happy Path	41
5.1.2 Halfway There	44

5.1.3 I know It Was There	48
5.1.4 You Can't See It	49
5.2 C2PA and the AFP IMATAG Proof of Concept	54
5.2.1 Authentication Workflow	55
5.2.2 Guarantees	56
5.2 C2PA and Audio	57
5.2.1 AI-Extended Audio Label	57
5.2.2 C2PA for Audio Translation and Dubbing	59
6. Potential Integrations of C2PA with vera.ai Tools	60
6.1 C2PA and the Database of Known Fakes	60
6.2 C2PA and the Keyframe Selection & Enhancement Service	65
6.3 C2PA and the Geolocalizer Tool	69
Conclusion	72
References	72

## Executive Summary

---

This report, developed under the vera.ai project, highlights the transformative potential of the Coalition for Content Provenance and Authentication (C2PA) in addressing the challenges of synthetic media, misinformation, and content manipulation. It shows how C2PA is providing an open standard for embedding cryptographically verifiable content credentials directly into digital media, enabling traceable provenance and enhancing trust in digital assets.

Key features of C2PA include soft binding (through e.g. watermarking and fingerprinting), AI-generated content labelling, and cross-platform interoperability supported by industry giants such as Adobe, Microsoft, Google, the EBU, and EBU Members such as public broadcasters BBC, CBC/Radio-Canada, WDR, DW and others. The adoption of C2PA ensures transparency in media workflows, empowers consumers to assess content authenticity, and addresses growing concerns around AI-generated media and their ethical use.

C2PA aligns with the goals of vera.ai by strengthening content integrity and traceability while countering the misuse of generative AI. It can become part of the same ecology as the fact-checking tools developed within the vera.ai project. Despite challenges such as limited support for live streaming and varying implementation standards, progress is being made towards broader adoption, including expected ISO certification by 2025/26.

Through use cases in journalism, publishing, and AI-generated content, C2PA is establishing itself as a cornerstone for verifiable digital provenance, fostering audience trust and combating media manipulation. This positions C2PA as a pivotal tool in ensuring transparency and accountability in today's digital ecosystem.

This report describes what C2PA is in further details: how it works “under the hood”, and how it can be used from a user point of view. This technical description is then followed by a description of the process of bringing up C2PA as an industry standard, including governance structure, impact, and current technical challenges. Concrete use cases of C2PA in a fact-checking and journalistic context are provided and illustrated by user scenarios, and potential implementations of C2PA in concert with vera.ai tools are presented in detail as proof of concepts.

# 1. Introduction

---

As disinformation continues to spread and damage society, using increasingly sophisticated methods of generation and distribution, there is an urgent need to develop tools to counter this modern plague. For this, two complementary approaches can be taken:

1. Equip journalists, fact-checkers and other media professionals with advanced tools to help them verify content in a reliable and efficient manner (the approach taken by the vera.ai project<sup>1</sup>).
2. Guarantee the authenticity of content originating from trusted sources.

The Coalition for Content Provenance and Authenticity (C2PA) takes the second approach by hardwiring credentials into content. We believe that tackling both approaches together, in unison, is essential in the fight against disinformation.

## 1.1 The Challenge of Synthetic Media in Detection and Authentication

---

Detection and authentication are valuable tools in the verification process that helps determine the truth or correctness of a piece of information. Detection involves identifying signs of manipulation or unusual patterns, using Machine Learning (ML) and forensic analysis, and results in the flagging of suspicious content. It is especially useful when provenance data is unavailable. Authentication focuses on confirming authenticity through, for example, cryptographic signatures, trusted metadata, and traceable provenance. These two approaches – besides other analytical tools in the verification toolbox – provide a robust defence against misinformation and manipulated media, especially if combined together.

Relatively recently, synthetic media content creation technologies have fundamentally transformed the disinformation landscape. Driven by Machine Learning and generative AI (Gen AI), the creation of hyper-realistic forgeries surpass traditional text-based fake news creation in both sophistication and deceptive power. Deepfake videos, AI-generated articles, and voice cloning tools produce content that appears strikingly realistic, amplifying their potential to manipulate public perception and influence critical decisions. This technological advancement represents a significant escalation in information warfare capabilities. Unlike earlier forms of misinformation, which required minimal technical skills and typically resulted in less convincing fabrications, synthetic media can now produce highly realistic forgeries of public figures, fabricate plausible news events, and create false evidence with unprecedented authenticity. The resulting content results in the misinformation of citizens and consequently poses substantial risks to democratic processes, financial markets, and social stability.

The task of detecting fake media is therefore becoming increasingly complex with the advancement of such technologies, posing difficulties for both seasoned professionals and AI detection systems in differentiating it from genuine content. Primary challenges involve the high volume of generated fake content, the nuanced nature of modifications applied to these contents, and the rapid progression of AI

---

<sup>1</sup> vera.ai is a research and development project focusing on disinformation analysis and AI-supported verification tools and services, under which this report is published: <https://www.veraai.eu/home> (last accessed on 18 July 2025)

methods that often surpass existing detection capabilities. Modern solutions (such as AI or Not<sup>2</sup>) exist, many of which utilise temporal features and frequency analysis, cutting-edge AI strategies, deep learning models, and examine discrepancies in aspects such as lighting, shadows, metadata, genuine characteristics of devices, surroundings, and so forth. The vera.ai project in particular has produced a significant body of innovative work<sup>3</sup> in the detection area – for example in the area of synthetic media detection and audio forensics – as well as other AI-based analytical tools that support fact-checkers and journalists in their daily work, such as information visualisation of disinformation networks, or resources to simplify, speed up and link together research processes.

With regards to authentication, the evolving sophistication of generative AI and manipulation techniques underscores the necessity of more universal, scalable, and interoperable solutions. This is where industry groups such as the C2PA coalition come into play. The focus of the C2PA standard is to support the authentication process, where trust in the media content can be already embedded at the point of creation or origination of this content. Developed through the collaboration of industry leaders such as Adobe, Microsoft, Intel, and users (*Project Origin*<sup>4</sup>, BBC, New York Times, CBC/Radio-Canada), C2PA aims to create a robust, unified approach to embedding metadata and cryptographic proofs directly into digital media. These provenance-driven standards ensure that content can be traced all the way back to its origin, accompanied by a verifiable chain of custody that highlights where changes have occurred. Integrating standards like C2PA with established tools into the media workflow for content verification not only strengthens the ecosystem but also paves the way for greater cross-platform adoption and interoperability for combating misinformation.

In this report written by the EBU (a partner in both the vera.ai project and in the C2PA coalition), we explore C2PA's pivotal role in addressing the challenges of content authentication and its potential to set a global standard for provenance and trust in digital media for broadcasters.

## 1.2 C2PA as an Authentication Tool in the Verification Process

---

Established in 2021, C2PA is building trust in digital media by developing an open technical standard to certify its origin and history. This standard addresses the challenge of misinformation through content credentials, a secure annotation system that documents the provenance of images, videos, and audio throughout their lifecycle. This technology captures essential information – such as capture device details, editing history, and distribution pathways – and makes this data accessible to end users on publishing platforms. The C2PA specification, first released in 2022, and now evolving to include AI-generated content, has garnered significant market interest and is supported by a growing membership of major tech firms, news organisations, and hardware manufacturers.

---

<sup>2</sup> AI or Not: <https://www.aiornot.com/> (last accessed on 18 July 2025)

<sup>3</sup> See the vera.ai Zenodo Community <https://zenodo.org/communities/veraai/records> (last accessed on 18 July 2025), which gathers publications accepted in leading conferences and journals, as well as documentation of the technologies and tools developed to support the fight against disinformation.

<sup>4</sup> Project Origin: <https://www.originproject.info/> (last accessed on 18 July 2025)

## 1.3 The Relevance of C2PA to vera.ai

---

Whilst the vera.ai project was already in progress, the C2PA initiative began to gain significant traction across the broader media and technology sectors, particularly in response to the growing need for trust, authenticity, and provenance in digital content. Recognising the strategic relevance of this emerging standard, the EBU (see also EBU Technical Committee statement on C2PA (EBU TC, 2024)) as a core partner in the vera.ai consortium acted swiftly to introduce C2PA and its underlying technical framework to the consortium. This included highlighting the collaborative, open governance model of C2PA, its alignment with industry-wide standardisation initiatives, and its early but promising implementations in real-world testbeds and pilot projects.

C2PA addresses many of the foundational challenges that vera.ai seeks to solve in a complementary manner, particularly around content integrity, traceability, and resilience against AI-generated disinformation. Given this alignment, the consortium decided to explore C2PA within vera.ai's broader landscape. This has led to the development of proofs-of-concepts (PoCs) mapping the specific requirements and functionalities of vera.ai tools to C2PA's evolving specification framework, contributing to C2PA working groups where appropriate, and ensuring that the mutual interests of the consortium are reflected in the direction of the standard's development. The potential integration of C2PA's architecture into the vera.ai ecosystem positions the project to benefit from an interoperable, industry-backed solution for secure metadata and verifiable media assets and consequently reinforcing vera.ai's long-term impact and sustainability.

## 2. What is C2PA?

---

Let us take a closer look “under the hood” of C2PA: how it is implemented and how users can interface with it.

### 2.1 How Does C2PA Work?

---

C2PA builds on the notion of embedding hardwired content credentials into digital media, in order to provide integral contextual and historical information about the content and make it verifiable by users upon consumption.

#### 2.1.1 Content Credentials

---

Content credentials are a collection of cryptographically verifiable metadata of a media asset. The website [contentcredentials.org](https://contentcredentials.org) provides guidance and hosts a variety of tools for users to work with these credentials.

The specification supports the following media types, which are including but not limited to:

- Image formats (such as JPEG, PNG, WEBP)
- Video file formats (such as MP4, AVI, MOV)
- Audio formats (MP3)

- Video on demand (VoD) streams format HLS<sup>5</sup> and DASH<sup>6</sup> have also been adopted recently through the implementation of ISO-BMFF format<sup>7</sup>, therefore adding support for authentic streaming.
- Fonts
- Portable Document Format (PDF)
- Specifications for live streaming and text-based assets such as HTML webpages or documents are still in progress.

Content credentials are a collection of manifests (see Figure 1) that provide contextual information about an asset's history at a specific point in time. Each manifest within the store may also contain links to additional ingredient manifests, indicating that an asset contributed to the creation of another asset.

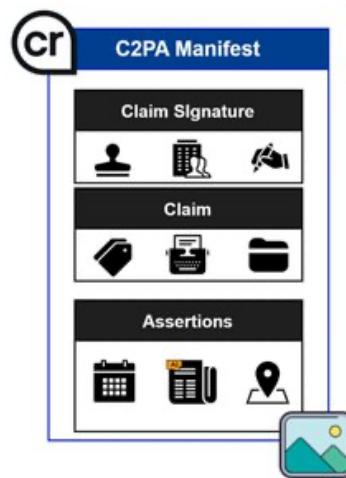


Figure 1: C2PA Manifest<sup>8</sup>.

## Assertions

*Assertions* are the provenance of information, contextual information about the creation and history of the asset. Example of assertions may include creative work information, date and time, GPS-location, EXIF/IPTC metadata, and most importantly the digital source type<sup>9</sup>: annotating how the assets were created: camera-captured, AI-generated or synthesised, screen-captured and so forth. The assertions refer only to the media asset file and are stored within the manifest itself.

---

<sup>5</sup> HTTP Live Streaming (HLS) is a protocol developed by Apple for streaming audio and video over HTTP. It is widely used for both live and on-demand content delivery.

<sup>6</sup> Dynamic Adaptive Streaming over HTTP (DASH), also known as MPEG-DASH, is an adaptive bitrate streaming technique that enables high-quality streaming of media content over the Internet, delivered from conventional HTTP web servers, similarly to HLS.

<sup>7</sup> The ISO base media file format (ISO-BMFF) is a container file format that defines a general structure for files that contain time-based multimedia data, such as video and audio.

<sup>8</sup> Image created by the EBU

<sup>9</sup> IPTC Digital Source Types: NewsCodes Scheme (Controlled Vocabulary).  
<https://cv.iptc.org/newscodes/digitalsourcetype/> (last accessed on 18 July 2025)

## Claim

The manifest's assertions are cryptographically hashed and bundled as a *Claim* structure then stored in the manifest. Note that the assertions (human readable) cannot be signed by the signature algorithm directly. They need to be passed through a cryptographic hashing algorithm prior to signing. This is what enables the security of the content credentials.

## Claim Signature

A *Claim Signature* is a temper-evident cryptographic signature that binds the digital identity of the signer to the claim about the content. The claim signature's structure additionally includes a certificate (credential) that identifies the signer.

## Generation

C2PA generation creates a tamper-evident record, called a C2PA Manifest, that travels with digital content. This manifest holds various pieces of information, cryptographically sealed to ensure its integrity. Each significant action on the content, like creation or editing, generates or updates this manifest, building a transparent history of the asset's lifecycle. The core process starts when digital content is created or modified with a C2PA-enabled tool. Provenance information is then collected and encoded into the C2PA *Manifest*, which is then cryptographically signed using a private key to ensure its authenticity and integrity. This signed manifest is then embedded directly within the content or linked externally, forming a tamper-evident connection.

## Validation

Content credentials validation requires content with C2PA metadata. The verifier recomputes the cryptographic hashes from the *Assertions* located in the *Manifest*, as well as the content to form *the Claim* then verifies it against the signature reporting any mismatch in the process. The verifier additionally validates the included certificate of the signer. The complete steps could be found under the validation section of the C2PA specifications<sup>10</sup>.

The official C2PA verifier is hosted at [contentcredentials.org/verify](https://contentcredentials.org/verify). A content consumer can provide the C2PA content itself or a link to validate it. The verifier runs on the client's side, so the content is not uploaded to any external server. Other verification tools exist such browser plugins, command line tools, and verification websites.

## Certificate Authorities (CAs)

In the C2PA ecosystem, the *Certificate Authorities* have the following roles:

- Issuance of certificates: CAs are trusted third-party organisations that issue digital certificates of type X.509<sup>11</sup> (Boeyen, 2008) to content creators and C2PA-enabled products.

---

<sup>10</sup> C2PA specifications for validation:

[https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA\\_Specification.html#validation-clause](https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html#validation-clause) (last accessed on 18 July 2025)

<sup>11</sup> C2PA specifications for certificates:



- Identity verification: They perform due diligence to verify the real-world identity of the entities requesting certificates.
- Establishing trust: By issuing these certificates, CAs help establish a chain of trust, allowing consumers to confidently verify the identity of the content creator and the integrity of the provenance data. The C2PA maintains a *C2PA Trust List*<sup>12</sup> of recognised CAs.

### Timestamp Authorities (TSAs)

In the C2PA ecosystem, the *Timestamp Authorities* have the following roles:

- Time validation: TSAs provide cryptographically secure timestamps for digital signatures through the Time-stamp protocol (Zuccherato, 2001). This proves that a digital signature existed at a specific point in time, independent of the signer's system clock.
- Long-term validity: The timestamp ensures that the C2PA Manifest remains verifiable even if the signing certificate later expires or is revoked. A timestamp from a TSA attests that the signature was made while the certificate was still valid<sup>13</sup>.
- Combatting backdating or fraud: This independent time proof is vital for preventing the backdating of digital content or fraudulent claims about its creation time.

#### 2.1.2 C2PA AI Tags

---

C2PA can also be used to directly label AI-generated content. This is a critical functionality since typical postproduction processes may include AI (such as generative extend, colour space transforms, scaling, and edits). The provision is that assertions can state that the content is AI-generated and indicate exactly which AI model was used to generate it. This is done by including a so called “created” C2PA actions that denotes the creation of the asset at hand, a software agent that created the asset, the AI model that generated the image, and importantly, the digital source type attribute denoting the type of the created asset. For AI, this could be any of *trainedAlgorithmicMedia* or *compositeSynthetic* values (see Figure 2).

---

[https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA\\_Specification.html#x509\\_certificates](https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html#x509_certificates) (last accessed on 18 July 2025)

<sup>12</sup> C2PA Trust List

[https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA\\_Specification.html#\\_trust\\_lists](https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html#_trust_lists) (last accessed on 18 July 2025)

<sup>13</sup> C2PA specifications Creating a Claim:

[https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA\\_Specification.html#\\_time\\_stamps](https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html#_time_stamps) (last accessed on 18 July 2025)

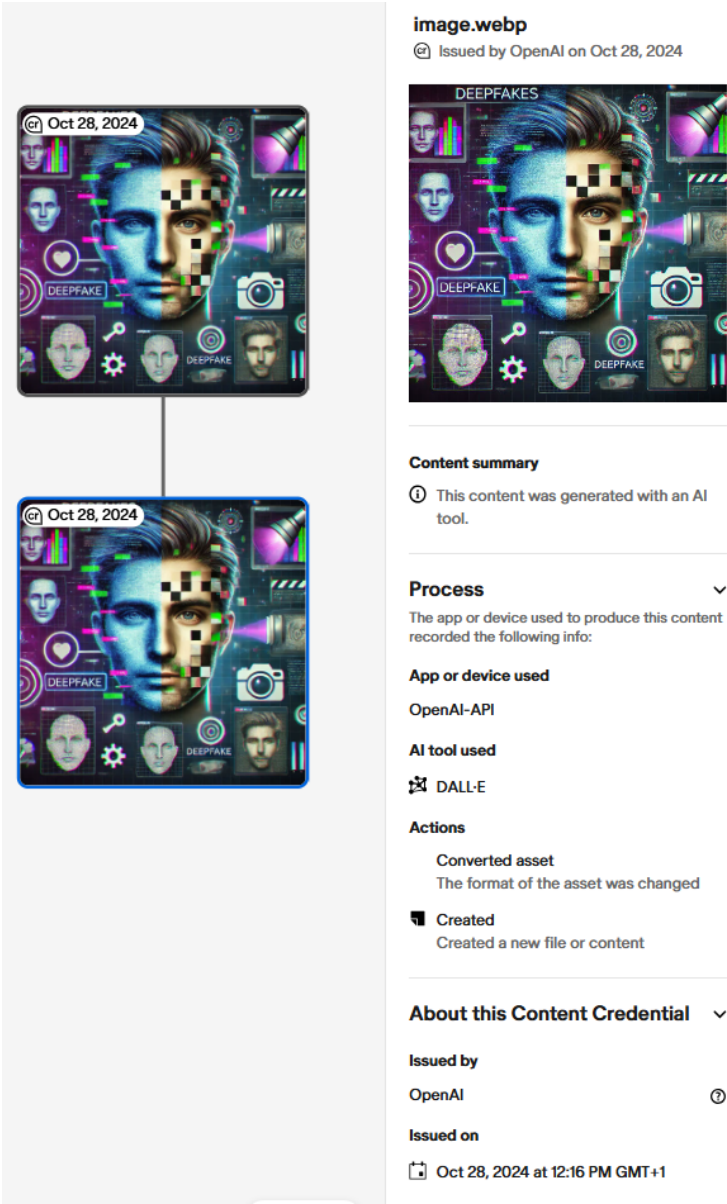


Figure 2: Example of C2PA-tagged AI content<sup>14</sup>.

2.1.3 C2PA: Training and Data Mining/Scraping

Up to version 1.4 of the C2PA specifications, C2PA supported a special type of assertion to enable the creator of content to provide manifest to consumer information about whether the asset may be used as part of a data mining or AI/ML training workflow. Data mining is also known as *data scraping*. This is expressed in a special assertion which includes a map of one or more training-mining-entries. Each entry describes whether its use is *allowed*, *notAllowed* or *constrained*.

<sup>14</sup> Image created by the EBU

It is worth noting that the training and data mining/scraping assertion have been deprecated since version 2.0, following some formation changes to the C2PA steering committee.

Table 1 describes the four possible training and data mining/scraping entries in C2PA version 1.4:

*Table 1 C2PA Training and data mining entries (C2PA v1.4)*

AI-Assertion entry	Description
c2pa.data_mining	Can any text or data content be extracted from the asset for purposes of determining "patterns, trends and correlations". This would include images containing text, where the text could be extracted via OCR.
c2pa.ai_inference	Can the asset be used as input to a trained AI/ML model for the purposes of inferring a result.
c2pa.ai_generative_training	Can the asset be used as training data for an AI/ML model that could produce derivative assets.
c2pa.ai_training	Can the asset be used as data to train non-generative AI/ML models, such as those used for classification, object detection, etc.

The value of constrained implies that permission is not unconditionally granted for this usage. Consumers of this content that wish to use the content in this way may need to contact the media rights holder (the signer) to get more info about the usage or obtain permission. In the absence of additional information, constrained must be treated as equivalent to *notAllowed*. More details on the constraints may be provided in the *constraints\_info* text field.

It should be noted that the media industry is currently debating provisions to regulate data scraping and unauthorised training of AI models. Here, C2PA could provide a useful technology and solution.

#### 2.1.4 Soft Binding

Soft bindings are described using soft binding assertions such as a fingerprint computed from digital content, or an invisible watermark embedded within the digital content. These soft bindings enable digital content to be matched even if the underlying bits differ. Since content can be subject to some modifications or changes this approach is different than hard binding. Additionally, the *soft-binding* links a stripped content back to the manifest from which it may have originated. This binding is implemented by using two supported techniques, watermarking and fingerprinting.

## Watermarking

Invisible watermarks are hidden information that is not observable by humans and may even be resilient to different processing steps along the media value chain. It embeds a small amount of information in content that can be decoded using a watermark detector. State-of-the-art watermarks can be impervious to alterations such as the cropping or rotation of images, or to the transcoding and packaging of video and audio. Importantly, the strength of a watermark is that it can survive rebroadcasting efforts like screenshotting, pictures of pictures, or re-recording of media, which effectively detach C2PA metadata.

## Fingerprinting

Fingerprinting is a technique to create a unique code based on image pixels, frames, or audio waveforms that can be computed and matched against other instances of the same content, even if there have been some alterations. C2PA clients can compute fingerprints of signed contents then store them separately (e.g., in a large database) along with their C2PA manifests. Upon verification of stripped content, the fingerprint can be re-computed on the fly and matched against a database of stored fingerprints. Unlike watermarking, fingerprinting does not require any prior information embedding to the content itself but only storing and updating fingerprints in the database.

## 2.2 How Does C2PA Interface with Users?

---

Now that we know how C2PA works, let us have a look at how end-users can interface with it, and the recommended approach for applications to surfacing it for their users.

### 2.2.1 End-User Experience

---

C2PA intends to provide clear guidance for implementers of provenance-enabled user experiences (UX). Developing these recommendations is an ongoing process that involves diverse stakeholders. The results will balance uniformity and familiarity with utility and flexibility for users across contexts, platforms, and devices. The Technical Working Group (TWG) responsible for UX intent is to present a comprehensive range of conventions for the user experience and evolve them based on feedback from both users, subject matter experts and implementors.

The two main goals in terms of user experience are:

1. Providing *assets creators* with a means to capture information and history about the content they are creating, and
2. Providing *asset consumers* with 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.

It is very important that user interfaces which are designed for the consumption of C2PA provenance are aware of the context of the asset.

The user groups targeted by C2PA are consumers, creators, publishers, and verifiers/investigators. The TWG of C2PA UX promotes the following “designing for trust” recommendations<sup>15</sup>.

### **Designing for Trust**

Rather than attempting to determine the veracity of an asset for a user, users themselves should be presented with the most prominent and/or comprehensive provenance information. As such, it is critical that users develop trust in the system itself, over the individual data presented. There is no design pattern that can guarantee to generate trustworthiness across multiple contexts, and while a degree of contextual customisation is anticipated, C2PA recommends all implementations adhere to the UX specification general principles.

### **Quality**

Implementations should be created using industry standards, robust user interface technologies.

### **Accessibility**

Implementations should adhere to accepted, current accessibility standards to ensure no users are excluded. For an example of such criteria, see the Web Content Accessibility Guidelines (WCAG)<sup>16</sup>.

### **Consistency**

Wherever suitable, UX patterns should match those outlined by the C2PA guidelines. In the case that this would break contextual paradigms of the platform, lean on precedent whether in the OS or app design. Users should not have to learn new paradigms or terminology in different contexts to access the information.

### **Summarised vs. Comprehensive**

Usually, a subset of the available information will be the most useful to a user in each context. A link to the display of full information should always be made available.

### **Linked**

Because the complete set of C2PA data for a given asset can be overwhelming to a user, some implementations may display only a fraction of available information. If applicable, a link to a more detailed display of information should be made available to allow the user to make a more informed judgement on the content’s trustworthiness.

---

<sup>15</sup> C2PA User Experience Guidance for Implementers: [https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html) (last accessed on 18 July 2025)

<sup>16</sup> W3C Web Content Accessibility Guidelines (WCAG) 2.1. <https://www.w3.org/TR/WCAG21/> (last accessed on 18 July 2025)

## 2.2.2 Levels of Information Disclosure

C2PA describes four different levels of progressive disclosure, as a guide the design of user interfaces (Figure 3).

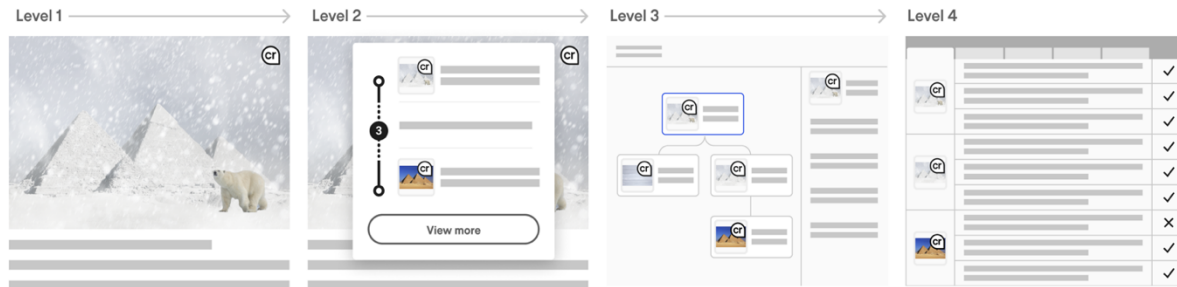


Figure 3: Disclosure levels<sup>17</sup>.

### Level 1 (L1)

L1 provides an indication that C2PA data is present and its cryptographic validation status (Figure 4).

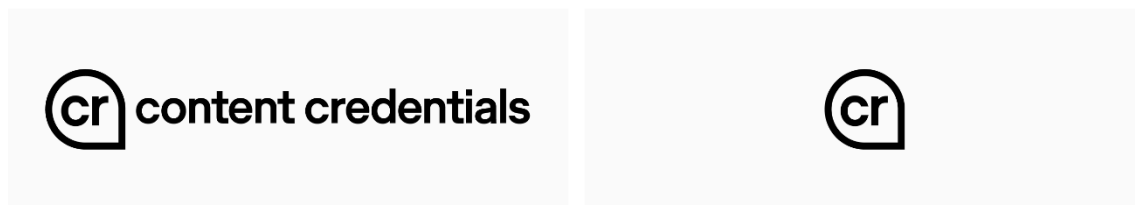


Figure 4: Content credentials extended (L) and pin (R) in L1.

### Level 2 (L2)

Level 2 provides a summary of C2PA data available for a given asset, which should provide enough information for the content, user, and context to allow the consumer to understand to a sufficient degree how the asset came to its current state (Figure 5).

<sup>17</sup> Source:

[https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_levels\\_of\\_information\\_disclosure](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_levels_of_information_disclosure) (last accessed on 18 July 2025)

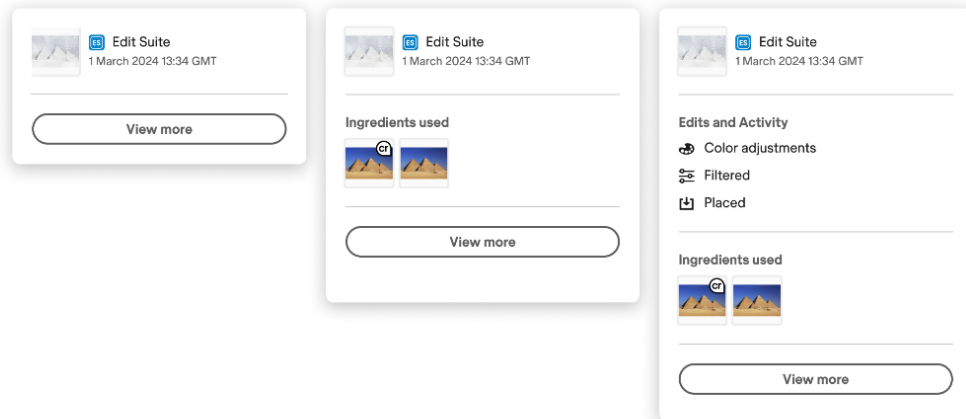


Figure 5: L2 provenance summary<sup>18</sup>.

### Level 3 (L3)

Level 3 provides a detailed display of all relevant provenance data. The relevance of certain items over others is contextual and determined by the UX implementer (Figure 6).

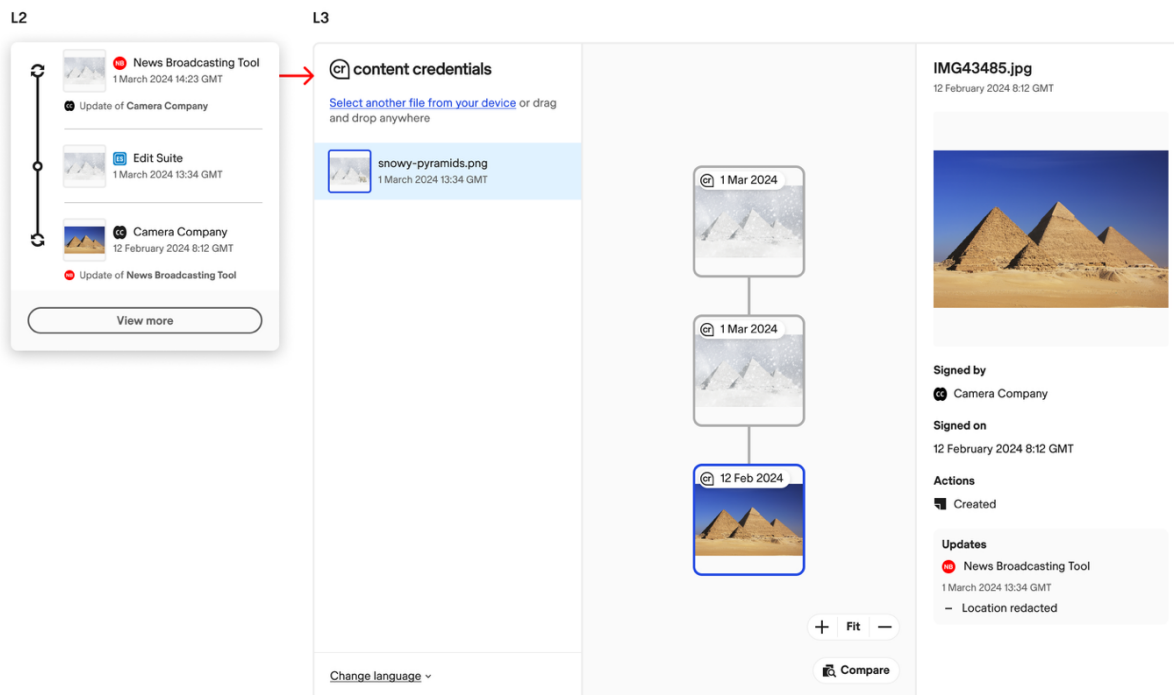


Figure 6: From L2 to L3<sup>19</sup>.

<sup>18</sup> Source:  
[https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_manifest\\_summaries\\_depth](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_manifest_summaries_depth) (last accessed on 18 July 2025)

<sup>19</sup> Source:  
[https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_redactions\\_and\\_updates](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_redactions_and_updates) (last accessed on 18 July 2025)

Level 4 (L4)

For sophisticated forensic investigatory usage, a standalone tool capable of revealing all the granular detail of signatures and trust signals is recommended. In addition to these standard levels, there will be common tools available for those interested in a full forensic view of the provenance data. This would reveal all available C2PA data across all manifests for an asset, including signature details.

Display of C2PA for AI Content

Audiences have the right to know when and how AI influences the information they consume. Direct disclosure is recommended for all content where generative AI has been used, as part of its production process.

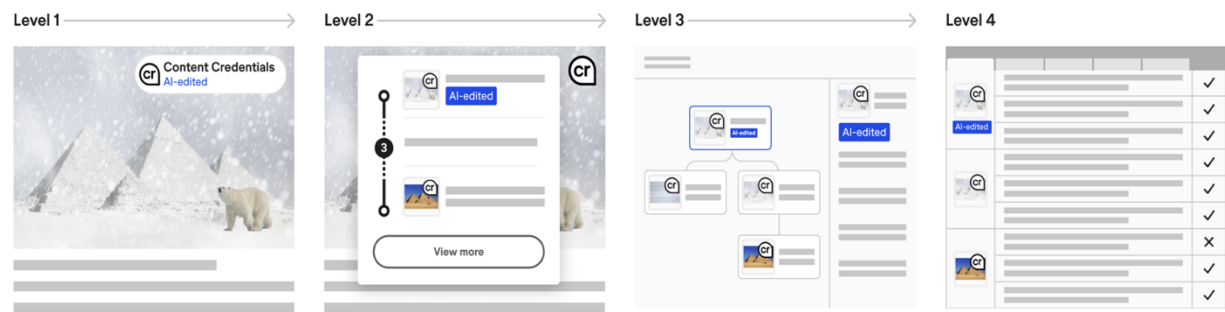


Figure 7: Progressive disclosure with label cards for AI-labelled content<sup>20</sup>.

Information contained in label cards should be considered important (Figure 7). Therefore, the recommendation is to place the cards high up in the hierarchy, just below the thumbnail and under the content summary section.

As the L3 assertion panel allows for more space, multiple label cards may be stacked on top of one another. Figure 8 shows a thumbnail view of a content alongside the generation time of the manifest, and most importantly an AI banner showcasing the AI nature of the edit.

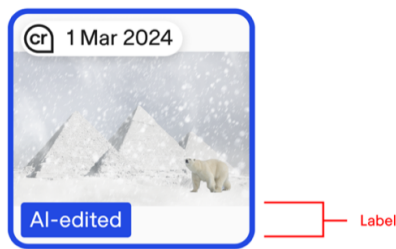


Figure 8: L3 Thumbnail with label<sup>21</sup>.

<sup>20</sup> Source: [https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_general\\_framework\\_appearance](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_general_framework_appearance) (last accessed on 18 July 2025)  
<sup>21</sup> Source: [https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_appearance\\_3](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_appearance_3) (last accessed on 18 July 2025)



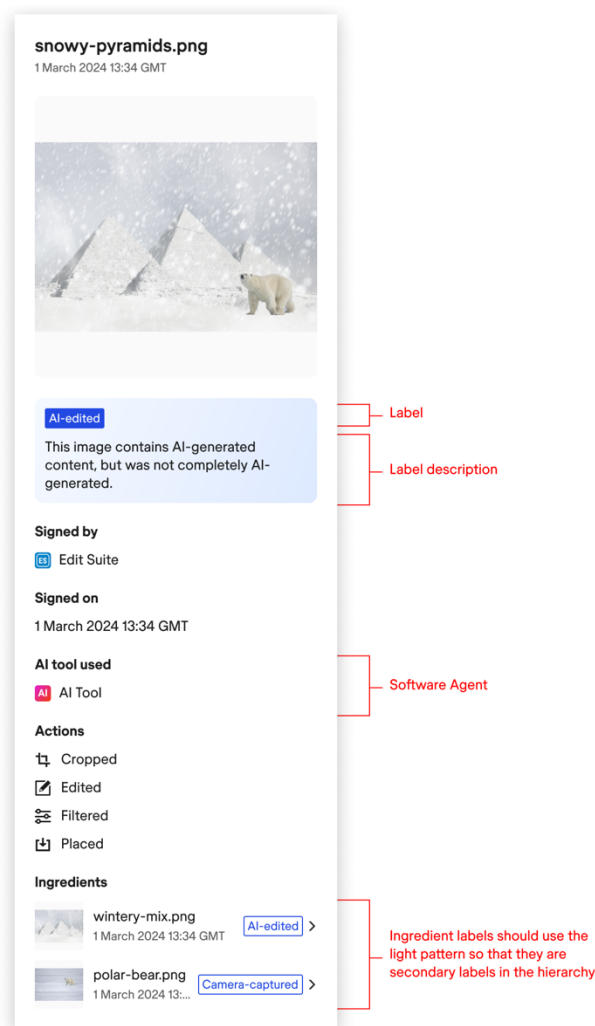


Figure 9: L3 manifest panel with AI label description<sup>22</sup>.

The manifest panel, as shown in Figure 9, displays the AI-edited/generated label along with its description. Labels and their descriptions are determined from the assertions of the C2PA metadata. The recommendations also suggest clear and concise formulated texts for the labels and descriptions, the software agent that generated the manifest attached to the displayed content. In case the content credentials include multiple ingredient manifests, these get also displayed with their labels as shown in Figure 10.

<sup>22</sup> Source: [https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_placement\\_3](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_placement_3) (last accessed on 18 July 2025)

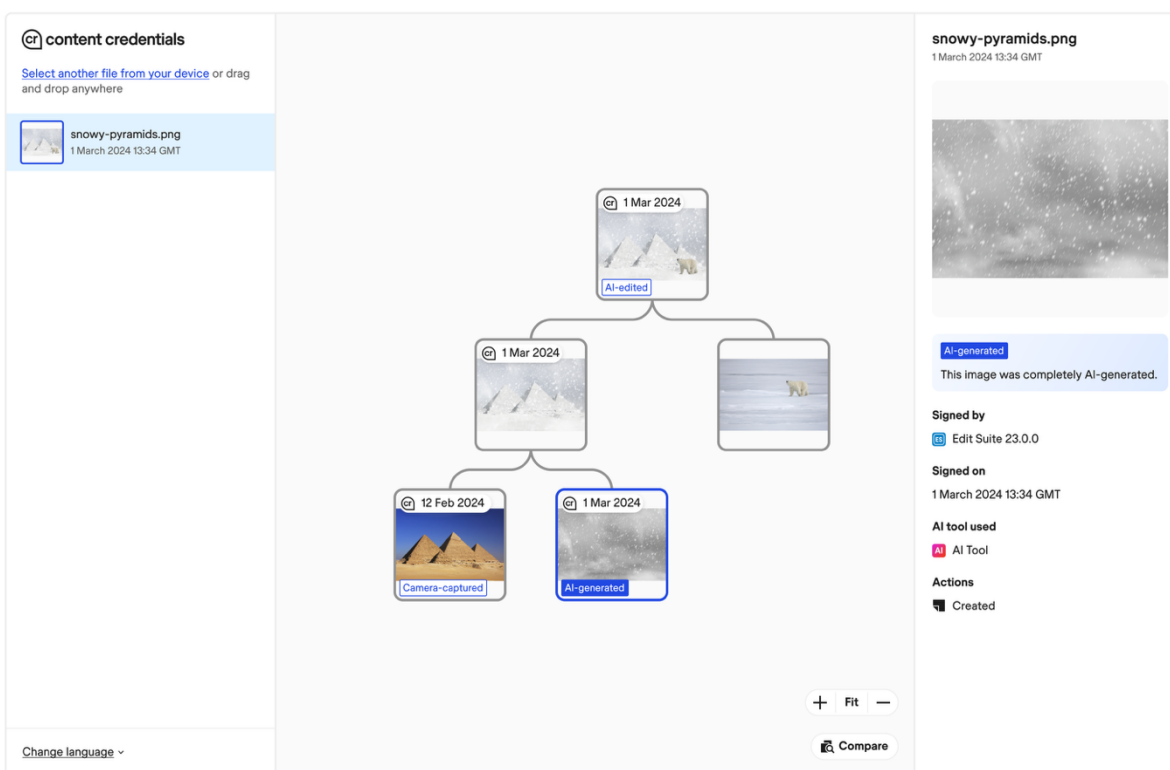


Figure 10. L3 tree view with labels<sup>23</sup>.

### 3. Bringing C2PA into the World of Media

Bringing up C2PA as a standard into the world of media requires a concerted effort from a number of key actors in the industry.

Within the EBU, we have for instance created a working group called *EBU C2PA workflows*<sup>24</sup> to not only promote C2PA but also open the floor for interesting discussions and allow for the participating members to share their experience and concerns about this technology.

This concerted effort includes joining forces to solve open technical challenges such as supporting live-streaming and enabling practical interoperability.

#### 3.1 Standardisation Efforts

The standardisation of the C2PA technology, also referred to as content credentials, is progressing through several different industry consortiums<sup>25</sup>, and accredited standardisation groups such as ISO's structured

<sup>23</sup> Source: [https://spec.c2pa.org/specifications/specifications/2.0/ux/UX\\_Recommendations.html#\\_interaction\\_3](https://spec.c2pa.org/specifications/specifications/2.0/ux/UX_Recommendations.html#_interaction_3) (last accessed on 18 July 2025)

<sup>24</sup> EBU C2PA Workflows group: [https://tech.ebu.ch/groups/c2pa\\_workflows](https://tech.ebu.ch/groups/c2pa_workflows) (last accessed on 18 July 2025)

<sup>25</sup> C2PA.org, IPTC.org

process. The preparation to align the C2PA specification with ISO format requirements began in March 2024 by the C2PA committee. Subsequently, in August 2024, ISO established a dedicated working group, identified as ISO TC 171/SC 2/WG 13, tasked with overseeing the standardisation of digital content authenticity technologies like C2PA. The group focuses on refining and validating the specification to ensure global applicability and compliance with ISO standards. By November 2024, C2PA was officially assigned the ISO number 22144. Member bodies that participate in the committee will be voting on its ratification. At the time of writing this document, the voting process was ongoing; after which, barring unforeseen issues, the standard will be ratified and prepared for publication.

Other accredited standardisation institutions are also looking into the impact of C2PA on the value chain. For instance, the application of C2PA in professional media workflows and the use of the MXF (Media Exchange Format) file format according to SMPTE 377<sup>26</sup> are currently being developed by the industry, and vera.ai's partner EBU has made a first input to the establishment of a *Study Group* in SMPTE<sup>27</sup>. It is also important to note that professional media ENG video camera manufacturers are active in this field and may show solutions soon. It is critical that the industry avoids the development of “quasi-standards” and that users express early on their requirements as well as formulate appropriate input to standardisation.

## 3.2 Governance Structure

---

The governance of the C2PA standardisation is designed to ensure openness, transparency, and robust stakeholder participation in the development and maintenance of technical specifications that address content authenticity and provenance in the digital ecosystem<sup>28</sup>.

C2PA operates as a project under the Joint Development Foundation (JDF)<sup>29</sup>, which is affiliated with the Linux Foundation. This affiliation provides a legally neutral and open collaborative framework, enabling multiple stakeholders to jointly create and evolve a royalty-free, openly available technical standard. The choice of the JDF governance structure reflects a commitment to transparent decision-making, industry-wide participation, and long-term sustainability.

The membership of C2PA includes a diverse group of organisations from the technology, media, hardware, and civil society sectors. Founding members include Adobe, Microsoft, Intel, BBC, Arm, and Truepic. These and other participating organisations form the core contributors to the specification process, while the governance framework remains open to broader industry engagement through general membership. Members contribute through both strategic input and technical implementation.

At the heart of the C2PA's development process is the Technical Working Group (TWG). This group is responsible for authoring and maintaining the core specifications. It brings together technical experts from member organisations to collaboratively define data models, cryptographic trust mechanisms, metadata

---

<sup>26</sup> <https://pub.smpte.org/latest/st377-1/st377-1-2019.pdf> (last accessed on 18 July 2025)

<sup>27</sup> Society of Motion Picture and Television Engineers (SMPTE): <https://www.smpte.org/> (last accessed on 18 July 2025)

<sup>28</sup> Please note that C2PA is not an accredited 'Standards Development Organization' (SDO) in the same way as SMPTE, ITU, ISO-IEC and ETSI.

<sup>29</sup> The Linux Foundation Project, Joint Development Foundation: <https://jointdevelopment.org/about/> (last accessed on 18 July 2025)

schemas, and interoperability rules. The TWG meets regularly and documents its proceedings to ensure transparency and traceability in decision-making.

The standard development process is consensus-driven, with voting mechanisms available when necessary. Proposals are typically reviewed and refined collaboratively until consensus is reached, reinforcing the commitment to openness and technical merit. All major decisions, including new features, changes, or version upgrades<sup>30</sup>, pass through a structured review and approval workflow.

The C2PA specification itself is published under a royalty-free licence, allowing free access and implementation by any interested party. It undergoes periodic public review cycles, where external stakeholders – including developers, researchers, civil society groups, and standards bodies – can submit feedback. This feedback is evaluated by the Technical Working Group and incorporated into future versions as appropriate.

In addition to internal governance, C2PA aligns with other key initiatives in the content authenticity space, including the Content Authenticity Initiative (CAI)<sup>31</sup>, Project Origin<sup>32</sup>, and relevant standards from W3C and IPTC. This alignment ensures interoperability across ecosystems and enhances the adoption of the standard across both proprietary and open-source platforms.

Finally, the overarching governance philosophy of C2PA is anchored in principles of open participation, transparency, security by design, and support for both human and machine-readable provenance. These principles are critical to building trust in digital content and are reflected throughout the organisation's processes and published materials.

### 3.3 Assessments of C2PA Impact by the BBC

---

The BBC has conducted extensive trials and user research to assess the potential impact of C2PA's content credentials on audience trust and content transparency. According to findings from BBC Research & Development and BBC News Labs, C2PA-backed provenance metadata significantly contributes to increasing audience confidence in digital content.

In a trial conducted with approximately 1,200 self-selecting users, BBC researchers found that 83% of participants reported increased trust in content after viewing associated provenance information via content credentials and 96% of users found the feature useful, and a similar share found it informative. These results held consistently across different types of content, including editorial images, user-generated photos, and stock photography, helping to reduce the trust gap between them (Monday, 2024).

In newsroom contexts, BBC News Verify began embedding content credentials into select visual reports from March 2024 onwards. This initiative allowed audiences to view when, where, and by whom content

---

<sup>30</sup> On 18 July 2025, the current version of C2PA is 2.2. The various versions mentioned throughout this document can be found on the C2PA Specifications menu here:

<https://spec.c2pa.org/specifications/specifications/2.2/index.html> (last accessed on 18 July 2025)

<sup>31</sup> Content Authenticity Initiative (CAI): <https://contentauthenticity.org/> (last accessed on 18 July 2025)

<sup>32</sup> Project Origin: <https://www.originproject.info/> (last accessed on 18 July 2025)

was created and edited, offering a layer of cryptographic transparency intended to support informed consumption of digital media (Halford, 2024).

BBC has emphasised that while provenance metadata does not and cannot confirm the truth of a piece of content, it can help audiences assess authenticity by showing the chain of custody and origin of media assets. This aligns with the editorial philosophy that if audiences *"know how something was made, they can better judge whether to trust it"* (Ellis, 2023).

To further these efforts, the BBC has hosted multi-stakeholder gatherings such as the Origin Media Provenance Summit<sup>33</sup> in October 2024, which brought together over 70 participants from 30 international organisations to share implementation insights and develop coordinated strategies.

While the BBC views C2PA as a promising tool in the fight against disinformation and media manipulation, it also acknowledges its limitations. The technology cannot prevent bad actors from removing metadata or creating deceptive content outside of authenticated systems. Nevertheless, BBC trials indicate that content credentials are a valuable step forward in increasing transparency and audience agency in the digital information space.

### 3.4 Current Open Issues and Further Work

---

The development of C2PA, as with other cutting-edge technologies, involves technical obstacles that require collaboration, especially regarding live-streaming and interoperability, and still requires further work.

#### 3.4.1 Streaming and Live Streaming

---

Initial C2PA (non-live) streaming has been demonstrated by several broadcasters, including NHK<sup>34</sup> and the EBU<sup>35</sup>. The broadcasting of C2PA enabled live streams is not yet formally supported by the C2PA specification. A task force in C2PA has been formed to tackle the technical requirements of live streaming technology. Many issues are being discussed and requirements for live streaming are under consideration.

The live stream taskforce is looking forward to reinforcing the authenticity and provenance of streams as well as protecting content creators' work from unauthorised use and distribution.

#### 3.4.2 Interoperability

---

The world of digital technologies and media is a diverse one, with its occasional incompatibilities. Essential to the C2PA mission is ensuring its interoperability across platforms, operative systems, tools and formats. For this purpose, both open-source and proprietary solutions are being explored.

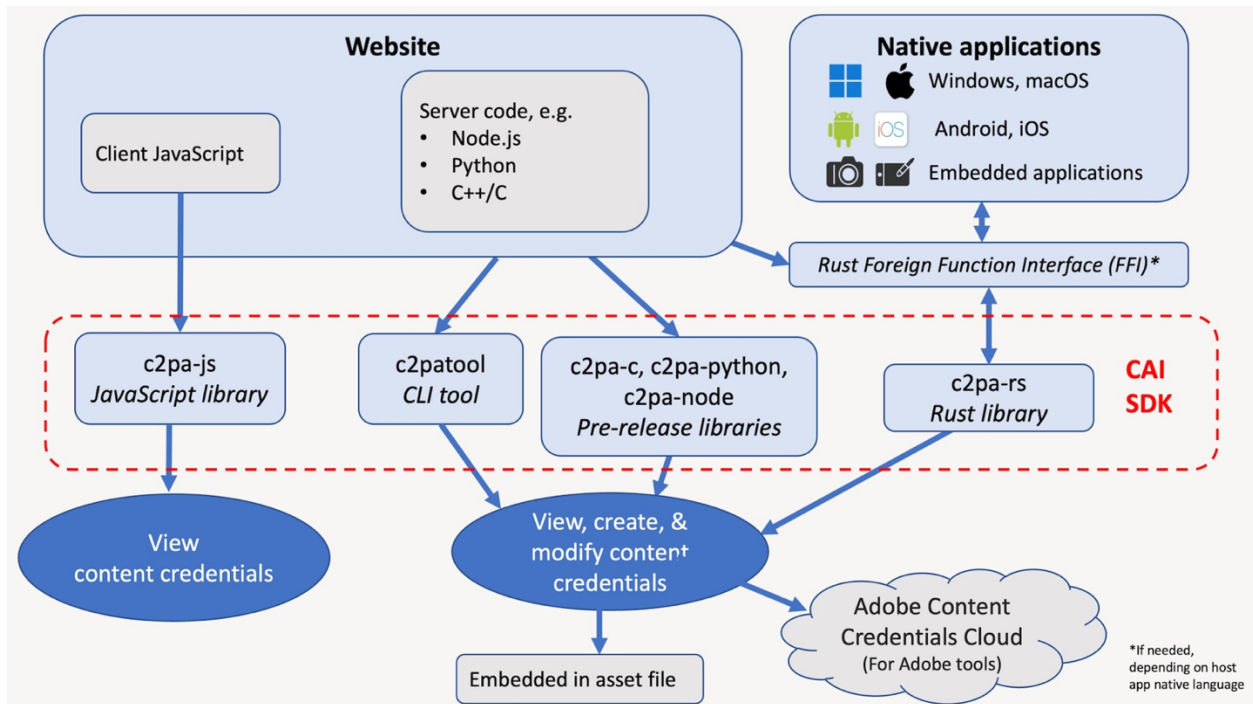
---

<sup>33</sup> BBC Origin Media Provenance Summit: <https://www.bbc.co.uk/rd/articles/2024-10-media-provenance-verification-news> (last accessed on 18 July 2025)

<sup>34</sup> NHK STRL Open House 2025: <https://www.nhk.or.jp/strl/english/open2025/tenji/7/index.html> (last accessed on 18 July 2025)

<sup>35</sup> C2PA End-to-End Workflow Demonstrated at EBU @ IBC 2024 (2024) <https://tech.ebu.ch/publications/c2pa-end-to-end-workflow-demonstrated-ebu-ibc-2024> (last accessed on 18 July 2025)

## 3.4.2.1 Open-Source Implementation

Figure 11: CAI open-source Software Development Kit<sup>36</sup>.

The Content Authenticity Initiative (CAI)<sup>37</sup> has launched a public initiative to implement the C2PA specifications<sup>38</sup> in multiple programming languages for different types of devices such as browsers, web and phone apps, and embedded devices, e.g. cameras (Figure 11). At the time of writing this document, the implementation is up to version 1.4 of the specification and community joint effort is being made to develop implementation for version 2.1 of the C2PA specification<sup>39</sup>. Third party software vendors such as *Truepic*<sup>40</sup> have already achieved a version 2 implementation.

For now, beta versions of Adobe products (*Premiere Pro*<sup>41</sup>, *Photoshop*<sup>42</sup>, *Express*<sup>43</sup>) have adopted C2PA and are relying on the opensource CAI's implementation of the specification.

<sup>36</sup> CAI open-source SDK: <https://opensource.contentauthenticity.org/docs/introduction> (last accessed on 18 July 2025)

<sup>37</sup> Introducing the Content Authenticity Initiative (CAI): <https://contentauthenticity.org/blog/test> (last accessed in July 2025)

<sup>38</sup> CAI offers an open-source implementation of C2PA <https://opensource.contentauthenticity.org/docs/getting-started/> (last accessed in July 2025)

<sup>39</sup> C2PA specifications version 2.1: <https://spec.c2pa.org/specifications/specifications/2.1/index.html> (last accessed on 18 July 2025)

<sup>40</sup> Truepic: <https://www.truepic.com/> (last accessed on 18 July 2025)

<sup>41</sup> Adobe Premiere Pro: <https://www.adobe.com/products/premiere.html> (last accessed on 18 July 2025)

<sup>42</sup> Adobe Photoshop: <https://www.adobe.com/products/photoshop.html> (last accessed on 18 July 2025)

<sup>43</sup> Adobe Express: <https://www.adobe.com/express/> (last accessed on 18 July 2025)

Google has also been working on an implementation of C2PA version 2.0+ for YouTube and Google Image Search<sup>44</sup>. Leica released the M11-P model<sup>45</sup> with its C2PA implementation<sup>46</sup>. The content credentials are cross device compatible and could be read by Adobe Premiere Pro and the open-source tool *c2patool*<sup>47</sup>.

C2PA interoperability considers Reading and Writing of different implementation libraries as shown in Table 2.

Table 2 Implementation libraries

Implementations/ Capabilities	Claim generation	Verification	Supported versions	Media formats
CAI open source	✓	✓	1.4	Image, audio, video, fMP4
Truepic library	✓	✓	1.4 to 2.1	Image, audio, video, fMP4
Leica	✓	—	1.x	Image
Microsoft	✓	✓	1.4	Image
Google	—	✓	2.0	Images, video

To facilitate interoperability between claim generators and validators, a claim generator declares which version of the specification it is using to generate the claim.

<sup>44</sup> Google implementation of C2PA:

\* Image metadata in Google Images: How C2PA metadata can appear in Google Search results. In: Google Search Central Documentation. <https://developers.google.com/search/docs/appearance/structured-data/image-license-metadata#c2pa-metadata> (last accessed on 18 July 2025)

\* Added information on how C2PA meta can appear in Google Search results. In: Google Search Central Documentation. <https://developers.google.com/search/updates#added-information-on-how-c2pa-metadata-can-appear-in-search> (last accessed on 18 July 2025)

\* Building trust on YouTube: ‘Captured with a camera’ disclosure. In: YouTube Help. <https://support.google.com/youtube/answer/15446725?hl=en> (last accessed on 18 July 2025)

<sup>45</sup> Leica M11-P, C2PA enabled camera: <https://leica-camera.com/en-int/photography/cameras/m/m11-p-black> (last accessed on 18 July 2025)

<sup>46</sup> Leica C2PA implementation: <https://contentauthenticity.org/blog/leica-launches-worlds-first-camera-with-content-credentials> (last accessed on 18 July 2025)

<sup>47</sup> c2patool, Content Authenticity Initiative, opensource programme implementing C2PA: <https://crates.io/crates/c2patool> (last accessed on 18 July 2025)



*“When a claim generator declares that it is using a version of the specification, it is declaring that the active manifest of the asset is produced in accordance with that version of the specification and thus does not contain any deprecated constructs listed under that version of the specification”<sup>48</sup>*

Some verifiers may be capable of parsing and verifying multiple versions of the specification.

*“A validator shall be compatible with at least one version of the specification but may be compatible with additional versions. A validator that is compatible with a specific version of the specification shall support all non-deprecated constructs listed for that version. If the validator encounters a manifest that uses constructs from a version of the specification that the validator does not support (either because they are deprecated or unknown), it may ignore the deprecated construct and process the rest of manifest as if that construct were not present.”<sup>48</sup>*

For example, a *Truepic* verifier can be configured for all different versions (up to version 2.1) of the specification. It will thus verify images signed by a Leica camera, by CAI open-source tools, or by the Microsoft C2PA implementation. The figures below illustrate different content credentials and their verification status by Truepic: Microsoft-generated (Figure 12), Adobe-generated (Figure 13), Leica-generated (Figure 14), and Truepic-generated (Figure 15).

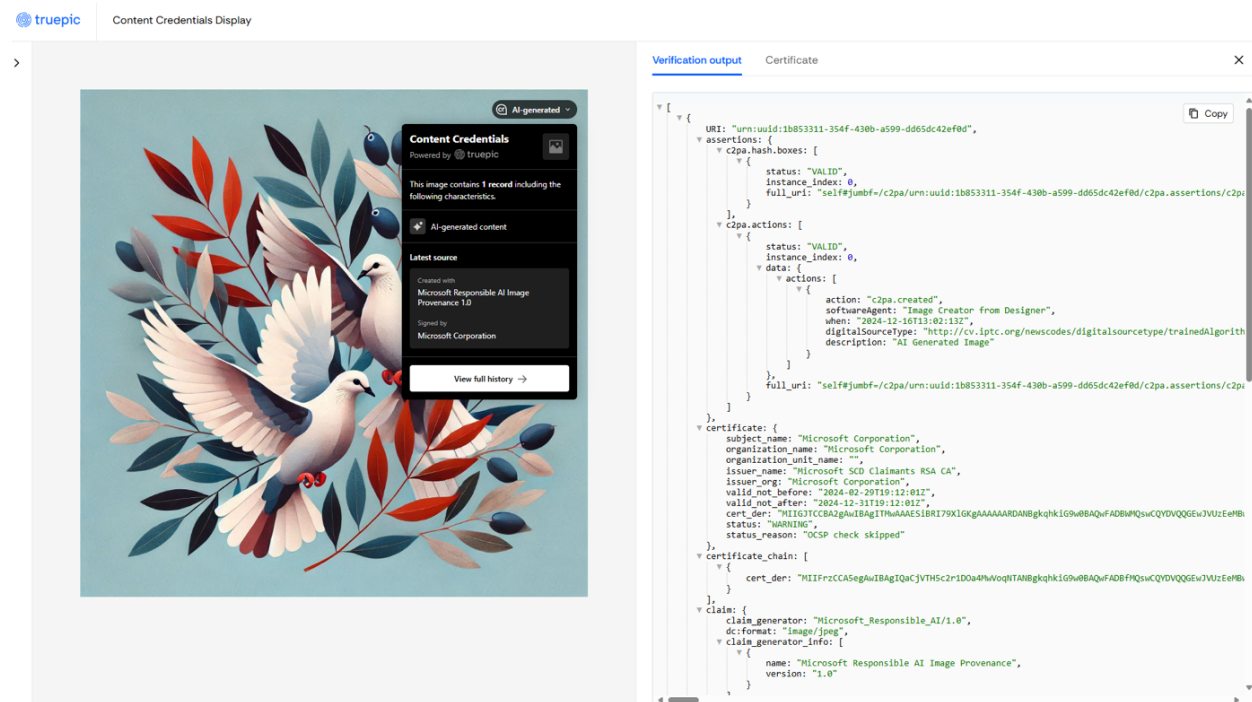


Figure 12: Microsoft-generated content credentials v1.4, verified by the Truepic verifier.

<sup>48</sup> C2PA Compatibility:

[https://spec.c2pa.org/specifications/specifications/2.1/specs/C2PA\\_Specification.html#\\_compatibility](https://spec.c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html#_compatibility) (last accessed on 18 July 2025)



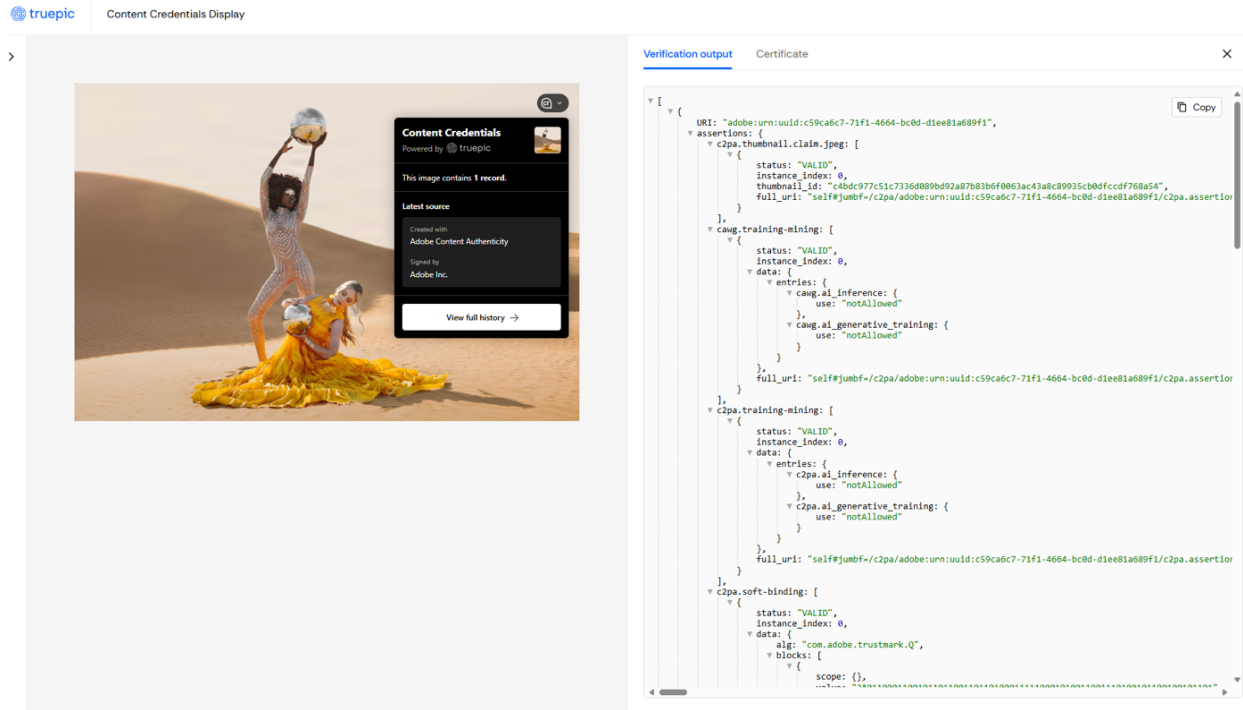


Figure 13: Adobe-generated content credentials v1.4, verified by the Truepic verifier.

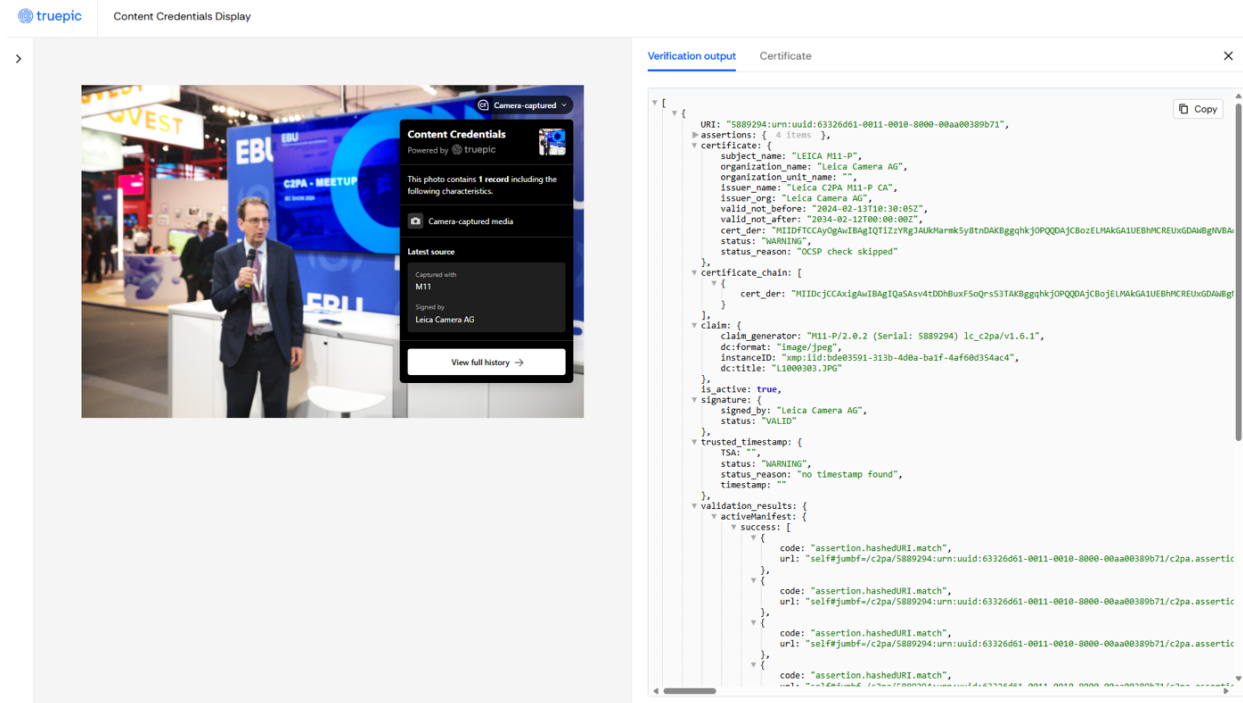


Figure 14: Leica camera-generated content credentials v1, verified by the Truepic verifier.

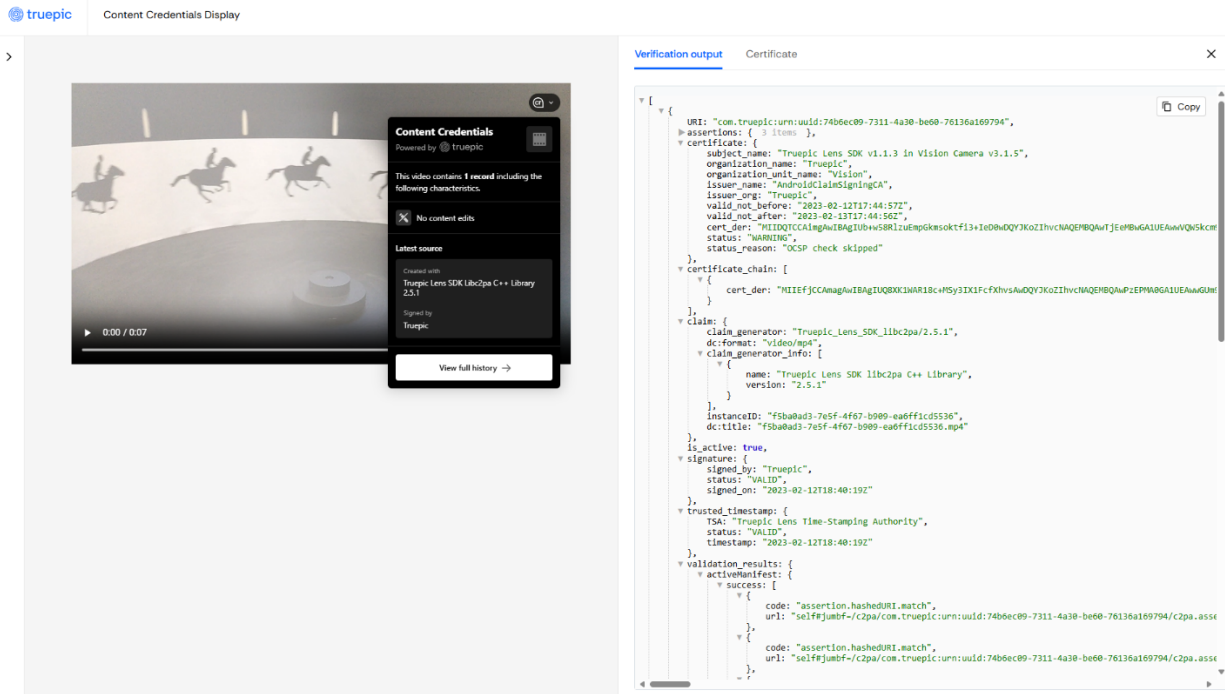


Figure 15: Truepic-signed content credentials v1.4 verified by the Truepic verifier.

### 3.4.2.2 Industry Examples of C2PA Interoperability

In parallel with efforts happening in open source, manufacturers are also working on C2PA interoperability, both in hardware and software.


#### Leica and Adobe Premiere Pro

Leica was one of the first hardware implementations for C2PA, initially commercialised with its M11-P camera product. The first product release<sup>49</sup> implemented version 1 of the C2PA specifications. Leica continuously updates its C2PA implementation to stay up to date with the C2PA specifications and fix implementation issues and bugs. Figure 16 shows a C2PA manifest created and signed by a M11-P camera. The *claim\_generator* field in the manifest describes the camera model and its serial number as well as Leica's C2PA software implementation used for signing.


<sup>49</sup> First C2PA compatible camera: <https://leica-camera.com/en-int/photography/content-credentials> (last accessed on 18 July 2025)

```
"manifests": {
  "5889294:urn:uuid:63326d61-0011-0010-8000-00aa00389b71": {
    "claim": {
      "dc:title": "L1000303.JPG",
      "dc:format": "image/jpeg",
      "instanceID": "xmp:iid:bde03591-313b-4d0a-ba1f-4af60d354ac4",
      "claim_generator": "M11-P/2.0.2 (Serial: 5889294) lc_c2pa/v1.6.1",
      "claim_generator_info": null,
      "signature": "self#jumbf=c2pa.signature",
      "assertions": [...],
      "alg": "sha256"
    },
    "assertion_store": {"stds.exif": ...},
    "signature": {
      "alg": "es256",
      "issuer": "Leica Camera AG"
    }
  }
},
}
```

Figure 16: C2PA Claim Generator definition for the Leica M11-P.




© Sep 6, 2024



© Leica Camera AG

### Generated Video

© Issued by Adobe Inc. on Sep 6, 2024




No thumbnail available


### Process

The app or device used to produce this content recorded the following info:


**App or device used**

 Adobe Premiere Pro (Beta) 24.6.0

**Ingredients**





D12EDBE8.JPG



© Leica Camera AG

### About this Content Credential

**Issued by**

 Adobe Inc. 

**Issued on**


 Sep 6, 2024 at 11:51 AM GMT+2

Figure 17: Video content credentials generated with Adobe Premiere Pro – Active manifest.

Leica images with content credentials can be imported and used in Adobe Premiere Pro (Figure 17). The content credentials of the generated video have a reference to the manifest of the Leica image (Figure 18).

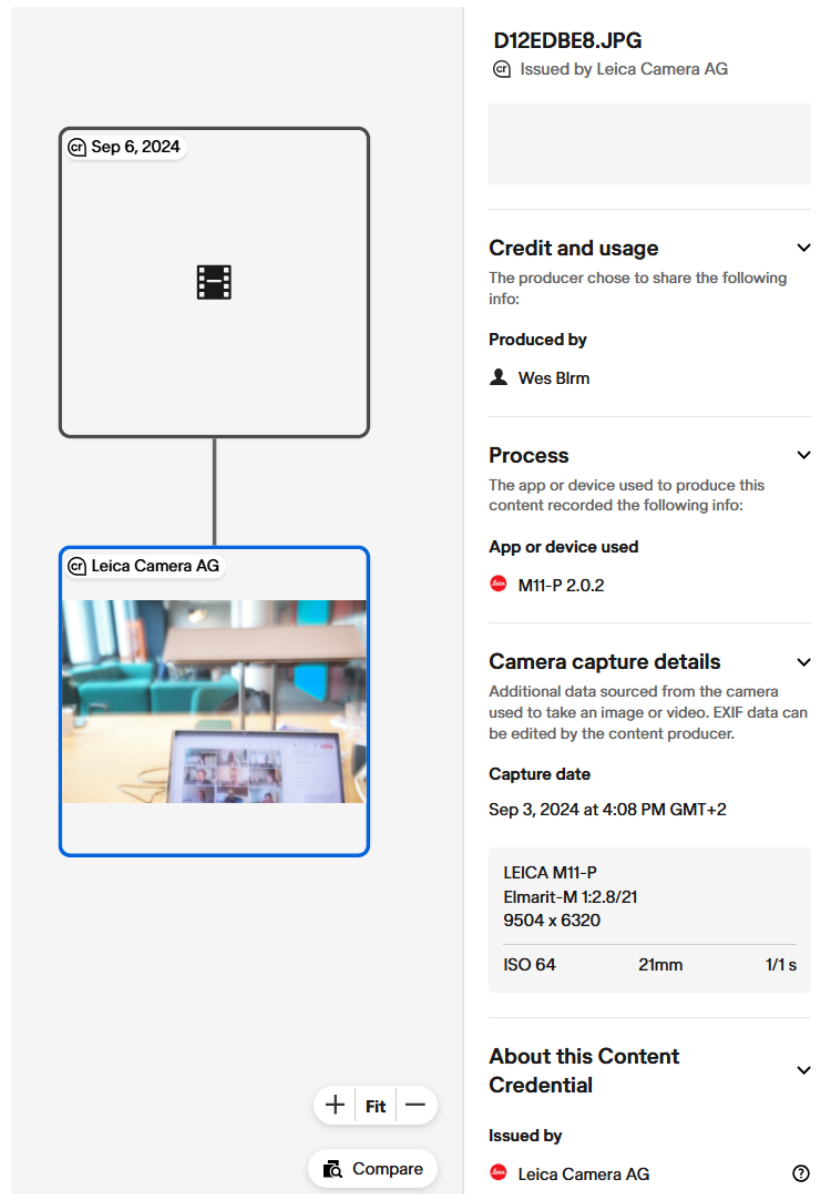


Figure 18: Video content credentials generated with Adobe Premiere Pro – Leica manifest.

### Format Interoperability: Image, Audio and Video

Adobe Premiere Pro supports both reading C2PA credentials from different formats and writing a new manifest for the generated output video (Figure 19).

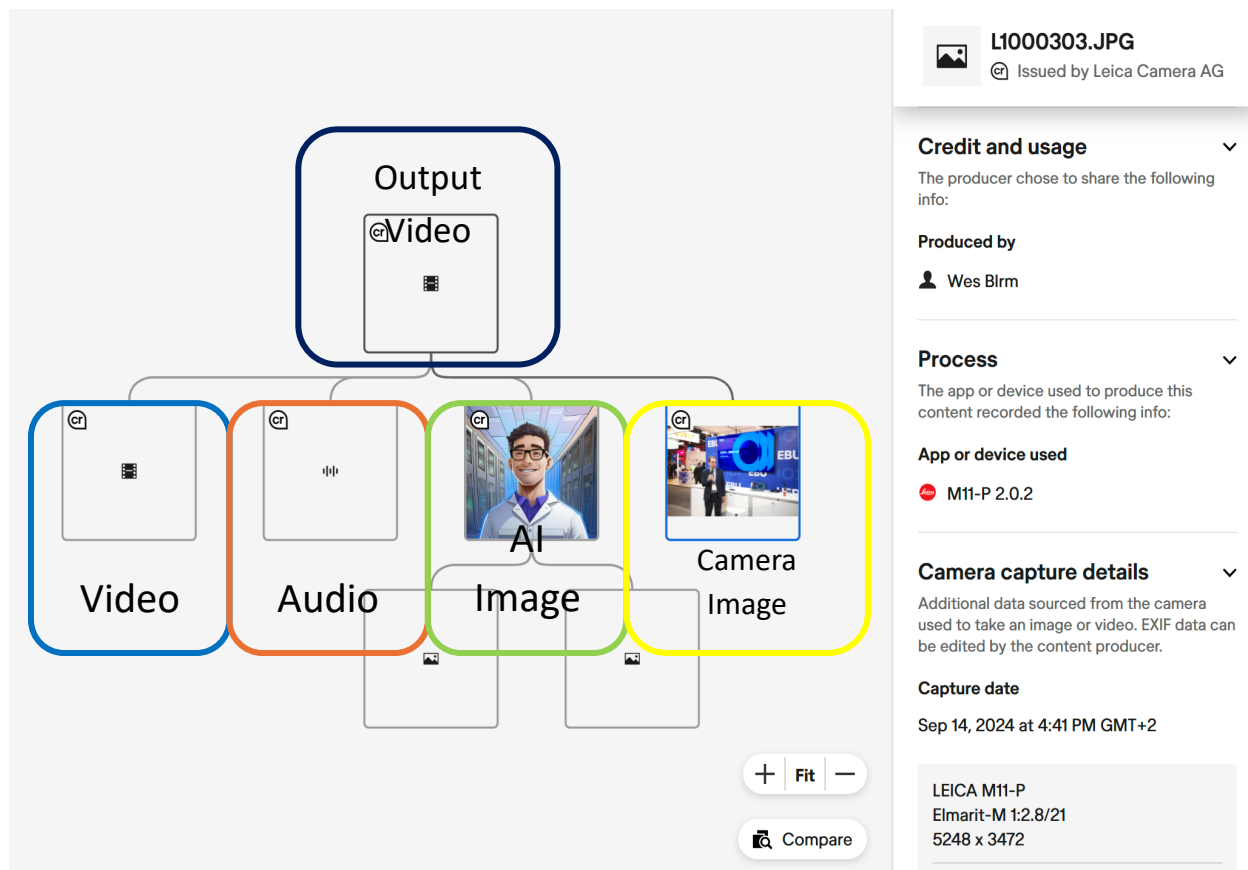


Figure 19: Adobe Premiere Pro-generated video with content credentials.

### 3.4.3 Limitations and Further Work

It is important to recognise that C2PA's effectiveness hinges on content credentials being consistently present and resistant to tampering: these can still be lost, stripped, or even applied by malicious actors. While mitigations exist against stripping content credentials such as watermarking and fingerprinting, it is noteworthy that many proprietary such solutions often lacking open-source scrutiny due to security reasons. Furthermore, user education is essential: trust cannot depend solely on the presence of credentials, but rather on the reliability of the credential's source, often verified against a recognised and well-established trust lists of identifiable organisations or hardware/software content generators.

While The C2PA specification still lacks support for important content types such as text-based content, live streaming, and professional media file exchange, work is under way in the various standardisation bodies. In line with these standardisation efforts, it is critical that plug-fests and interoperability tests accompany these activities. For example, the EBU has extensive experience in MXF interoperability testing and would volunteer to conduct C2PA tests once the MXF standard is C2PA-ready – although this would happen after the end of the vera.ai project.

Open-source implementations supporting dated versions of the specifications up to e.g. v1.4 by CAI, may break compatibility with third party software vendors that have already implemented newer versions of

the specifications (e.g. v2.1 by Truepic), or with C2PA products within the ecosystem that are using a proprietary implementation of v2.1 of the specification, such as Google, YouTube, LinkedIn or TikTok.

The trust model of C2PA requires that claim generators outputting content credentials are securely implemented. They may provide in some cases proof of correctness of the metadata information they include in the manifests, such as timestamps, GPS coordinates, or hardware used. Different levels of assurance may be provided, subject to the adoption of various security mechanisms and protocols in the claim generators environments. C2PA should provide guidance in the future about how to attain these assurance levels. C2PA verifiers will be equipped with a list of such trusted claim generators to present the content credentials as trusted or of unknown provenance, also taking into account assurance levels. The C2PA committee should follow up with policy documentation that governs the management of the trust list.

The existence of an open-source, configurable and complete verifier solution is necessary for C2PA consumers to validate and display provenance metadata. A more sophisticated verifier is a must, especially for fact-checkers and forensic researchers. Currently, the official C2PA verifier [contentcredentials.org/verify](https://contentcredentials.org/verify) only displays created assertions by the software and hardware agents but lacks in showing custom metadata (e.g., news publishers, fact-checkers). A C2PA verifier for news publishers is currently under development by IPTC<sup>50</sup>.

Finally, the question of how C2PA will help manage the data scraping questions has to be re-addressed since the standard has evolved.

## 4. Why Use C2PA?

---

Using C2PA as part of a media workflow offers a number of benefits, in particular helping audiences and consumers know what content to trust, thanks to content credentials and the labelling of AI-generated content.

### 4.1 To Increase Trust in Publishers' Content

---

Prior to publishing content, publishers can considerably improve audience and consumers' trust in their content by attaching content credentials to image, video, or audio assets.

The signature in the content credentials makes the asset and the provenance information tamper-evident, meaning that the slightest change to the content or its manifest will break the signature, and hence detected upon validation. Additionally, the content credentials contain a digital certificate identifying the signer, in this case the publisher of the content. This enables mutual trust between publisher and audience in different ways:

- The user can verify the signature of the content prior to consuming it using a C2PA compatible verifier. The identity of the signer is authenticated against a list of trusted certificates. The verification of the certificate is conducted by querying the Certificate Authority that issued the

---

<sup>50</sup> <https://iptc.freshcode.org/> (last accessed on 18 July 2025)

certificate. Furthermore, if timestamp exists in the content credentials, it is verified against a Timestamp Authority.

- If all checks pass successfully, the user can trust that the consumed content is as the publishers intended it to be, and that no tampering with the content occurred in-between the time of publication and the time of consumption. The user hence verifies the origin of the content and can trust it.
- The publisher can protect its reputation by only claiming responsibility for the content published with valid content credentials. No adversary can impersonate a publisher and share reputation damaging content on the internet on behalf of it

In an article about violence outbreaks in Haiti in March 2024 (Astier, 2024), the BBC re-publishes a modified video that it found on social media. They were transparent about the modifications they made to the video: removal of fake gunshots and quality enhancements. They also provided several provenance facts such as the platform from where the video first appeared. They also provide a real location where they believe, after several fact -checks, the video was taken: Grand Cemetery in Port-au-Prince, Haiti (Figures 20 and 21).

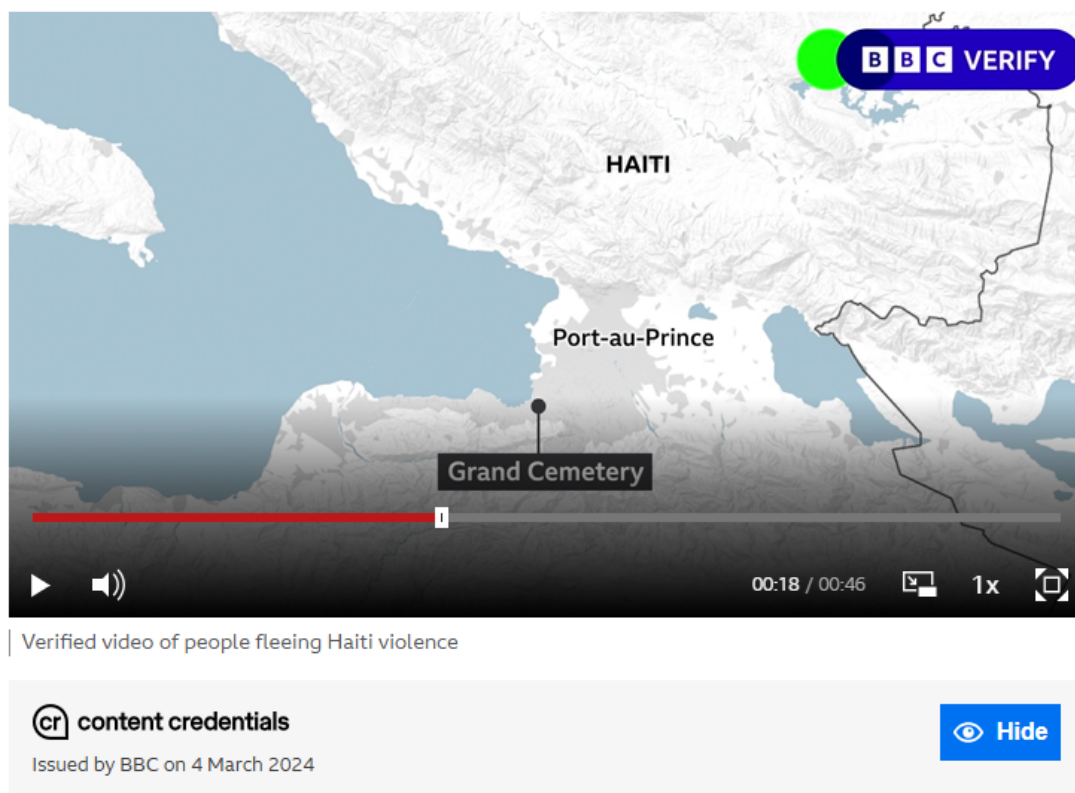




Figure 20: View of the signed BBC video with the content credentials banner under the video player.




 **content credentials**


Issued by BBC on 4 March 2024
 



 **Hide**


### About this video

The video shows people fleeing through the Grand Cemetery in Port-au-Prince, Haiti. (We have muted the audio as it had been misleadingly edited by a social media user to include sounds of gunfire taken from an earlier video). Then it shows the aftermath of an incident at Cabaret police station, north of the capital.

 **Posted on**  
TikTok

 **Created**  
6 March 2024

 **Location**  
18.534108, -72.344307  View map

 **Edits**  
Superficial edits were made to this content to improve technical quality, in line with editorial guidelines.

### Verification checks

Completed by BBC Verify

The video from the cemetery was first uploaded to TikTok on Saturday 2 March at 1750G (1250 local).

We've matched the gravestones, a large tree and other landmarks to existing images found Google Maps for the Grand Cemetery. The direction of the shadows observed indicates it was filmed in the morning.

The audio on the video is not verified - much of the gunfire noise has been edited in from an older video not filmed in Haiti.

The second video, of the police station, is from a Facebook post published on Sunday evening 3 March. We found no earlier versions of this video online.

Signage in the video matches that of the police Cabaret police station, both in an online image search and from a UN press release from 2018 when the station was inaugurated.

The police station lies about 20km north of the capital, Port-au-Prince.

(Location coordinates 18.736593, -72.417569)

Figure 21: Summary display of content credentials.

The fact-checks are also included in the content credentials of the video. This protects both the integrity and authenticity of the commentary about the video. The claims in Figure 22 are extracted from the C2PA metadata of the video. These claims were previously signed by the fact checking team at BBC. Any change to these claims from within the content credentials will render them invalid. It is important that the displayer acts as a verifier and verifies the content credentials before displaying them to the user.

```

{
  "@context": "http://schema.org",
  "@type": "ClaimReview",
  "author": {
    "@type": "Organization",
    "name": "BBC Verify"
  },
  "claimReviewed": "The video shows people fleeing through the Grand Cemetery in Port-au-Prince, Haiti. (We have muted the audio as it had been misleadingly edited by a social media user to include sounds of gunfire taken from an earlier video). Then it shows the aftermath of an incident at Cabaret police station, north of the capital.",
  "datePublished": "4 March 2024",
  "itemReviewed": {

```



```
"@type": "ImageObject",  
"caption": "Verified video of people fleeing Haiti violence"  
},  
"reviewBody": "The video from the cemetery was first uploaded to TikTok on Saturday 2 March at 1750G  
(1250 local).
```

We've matched the gravestones, a large tree and other landmarks to existing images found Google Maps for the Grand Cemetery. The direction of the shadows observed indicates it was filmed in the morning.

The audio on the video is not verified - much of the gunfire noise has been edited in from an older video not filmed in Haiti.

The second video, of the police station, is from a Facebook post published on Sunday evening 3 March. We found no earlier versions of this video online.

Signage in the video matches that of the police Cabaret police station, both in an online image search and from a UN press release from 2018 when the station was inaugurated.

The police station lies about 20km north of the capital, Port-au-Prince.

(Location coordinates 18.736593, -72.417569)

Correction 28 March: We constantly review our verification. In the original video you could hear gunfire - it has since been drawn to our attention that most of this was misleadingly edited in from an older unrelated video by a social media user. In line with our commitment to transparency, we have reviewed the video and muted the relevant passage of audio and amended our captions.",

```
"reviewRating": {  
  "@type": "Rating",  
  "alternateName": "Verified",  
  "bestRating": "5",  
  "ratingValue": "5",  
  "worstRating": "1"  
}  
}
```

Figure 22: Extracted BBC ClaimReview as a JSON from the content credentials of the fact-checked video.

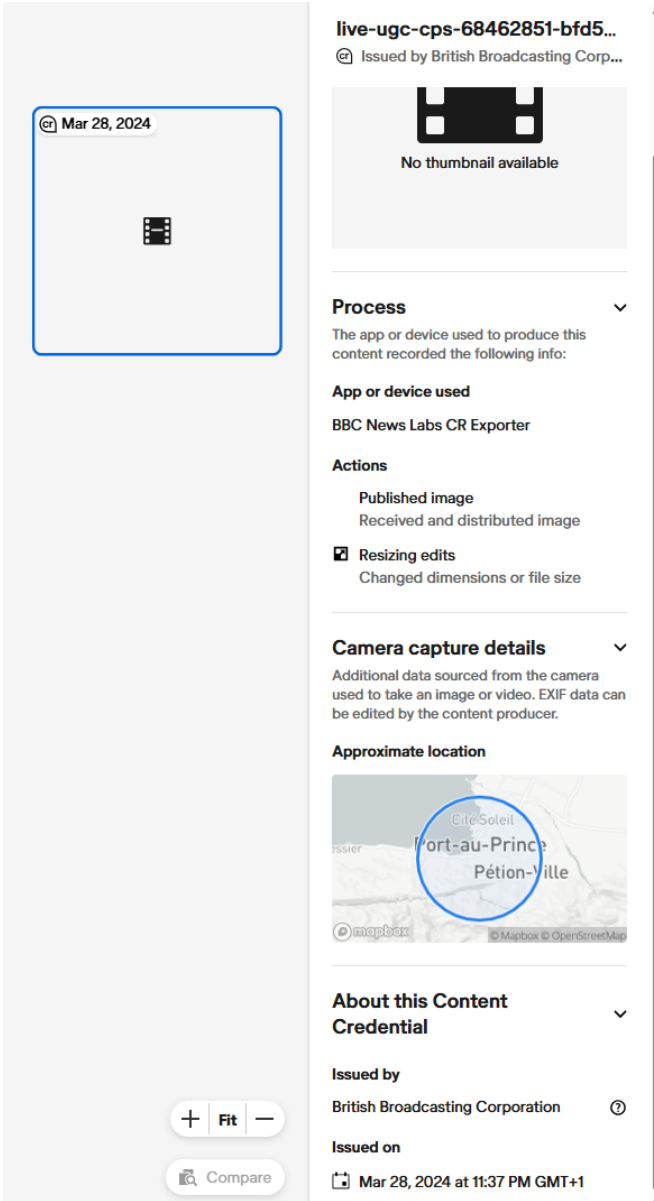


Figure 23: Level 3 display and verification of the BBC video in the C2PA official Content Credentials verifier<sup>51</sup>.

The BBC also provides an L3 display that shows the time the content credentials have been generated and the signer’s name (BBC), as well as a summary of the editorial actions (Figure 23).

## 4.2 To Label AI-Generated Content

Creating AI content responsibly is an important issue that both internet and AI companies are facing. Being able to prove whether a content has been totally AI-generated, simply AI-edited, or AI-retouched is

<sup>51</sup> Official C2PA verifier hosted at <https://contentcredentials.org/verify> (last accessed on 18 July 2025)

important both for users that consume the content, and for AI companies in identifying whether their AI models are being mistakenly trained on synthetic assets.

To tackle this problem, C2PA has developed a solution for this use case. It keeps track of the provenance information of an asset. The moment an AI asset has been created the *GenAI* server automatically attaches content credentials indicating the generative AI model used to output the asset, date and time of creation, and most importantly the digital source type, i.e. the nature of the asset's source.

C2PA uses the following IPTC-defined digital source types as labels for AI (Table 3).

Table 3 IPTC digital source types

Name	Label	Description
Created using Generative AI	digsrctype:trainedAlgorithmicMedia	Digital media created algorithmically using an Artificial Intelligence model trained on captured content <sup>52</sup>
Composite of elements	digsrctype:composite	Mix or composite of several elements, any of which may or may not be generative AI <sup>53</sup>
Composite including generative AI elements	digsrctype:compositeSynthetic	Mix or composite of several elements, at least one of which is generative AI <sup>54</sup>

In the following use case, we use *Adobe Firefly*<sup>55</sup> as a generative tool to create an artificial image then inspect its content credentials. Other tools such as *Microsoft Designer*<sup>56</sup>, *DALL·E*<sup>57</sup>, and *Adobe Photoshop Suite* have also implemented C2PA in a similar way. We then discuss the display and technical assertions added by C2PA.

### AI Image Generation

1. Prepare a real image: it will be used on a composition basis and provided to Firefly. For the sake of this example, we will use a peaceful image of Geneva on a sunny day (Figure 24).

<sup>52</sup> trainedAlgorithmicMedia: <http://cv.iptc.org/newscodes/digitalsourcetype/trainedAlgorithmicMedia> (last accessed on 18 July 2025)

<sup>53</sup> composite: <http://cv.iptc.org/newscodes/digitalsourcetype/composite> (last accessed on 18 July 2025)

<sup>54</sup> compositeSynthetic: <http://cv.iptc.org/newscodes/digitalsourcetype/compositeSynthetic> (last accessed on 18 July 2025)

<sup>55</sup> Adobe Firefly: <https://www.adobe.com/products/firefly.html> (last accessed on 18 July 2025)

<sup>56</sup> Microsoft Designer: <https://designer.microsoft.com/> (last accessed on 18 July 2025)

<sup>57</sup> DALL·E: <https://openai.com/index/dall-e-2/> (last accessed on 18 July 2025)



Figure 24: A peaceful image of Geneva, Switzerland<sup>58</sup>, used as input to Adobe Firefly.

2. Upload image to Adobe Firefly and then type a text prompt and select image styles for the output AI image. We use this prompt: *“Image of city nature with lava and volcano eruption in background and plane crash”*. As for the style of the image, we select *SynthWave* (see Figure 25).

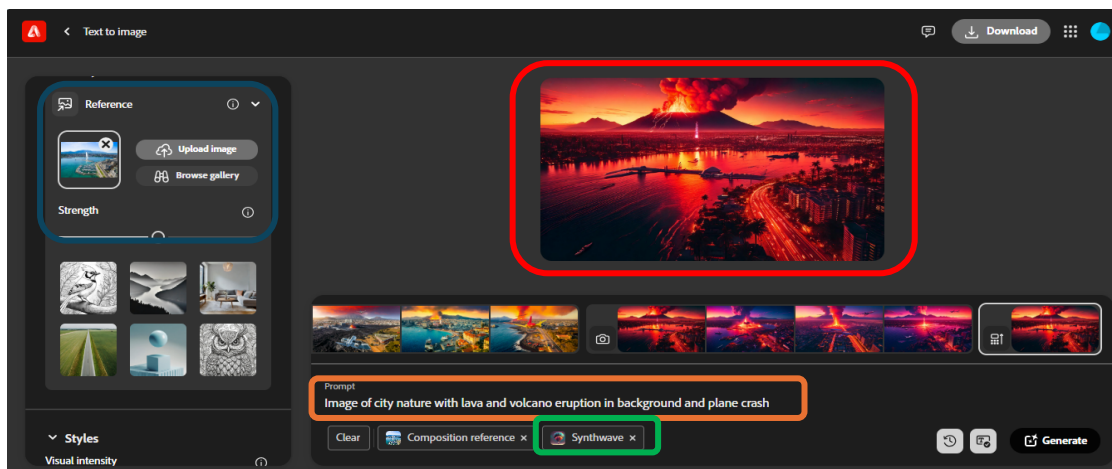


Figure 25: Adobe Firefly panel of generated images given a reference image, a prompt, and style.

3. Hit *Generate* and *Download AI image*: Adobe Firefly automatically generates the content credentials and embeds them as metadata into the image file.

<sup>58</sup> Source: ‘Aerial view of Leman lake - Geneva city in Switzerland’ by Samuel Borges (1 June 2016), iStockphoto, Photo ID: 528959560. Licensed to the EBU.

4. We now have generated a chaotic version of Geneva (Figure 26).



Figure 26: Firefly AI generated image “Chaotic Geneva”.

## Inspection

We can now inspect the C2PA AI image in a C2PA verifier:

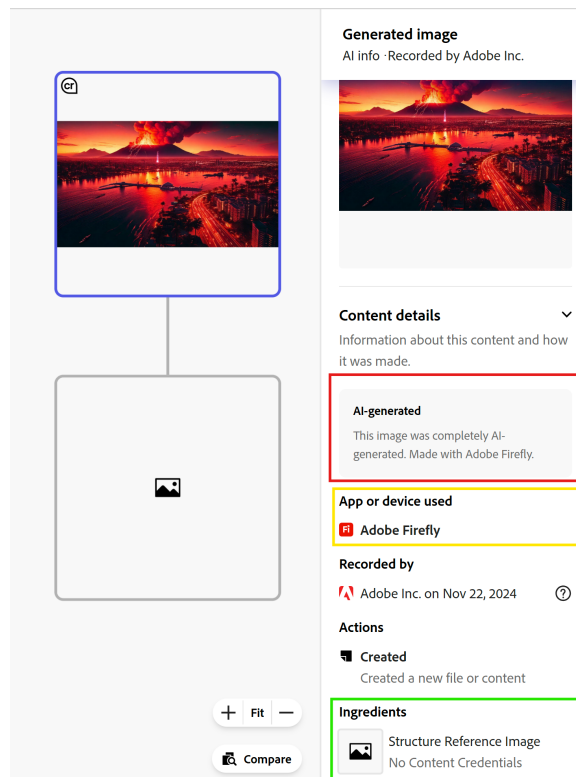


Figure 27: Verifier view of “Chaotic Geneva”.



In the red box in Figure 27, the content summary states that this image was created using an AI tool, with the yellow box specifying what AI tool (model) was used to generate the image. The green box specifies that an image was used as input to the AI model for structure inference, that is the image of “Peaceful Geneva”. Adobe Firefly does not include thumbnail of the input image for privacy reasons.

Figure 28 shows in detail the inspection of the C2PA AI image.

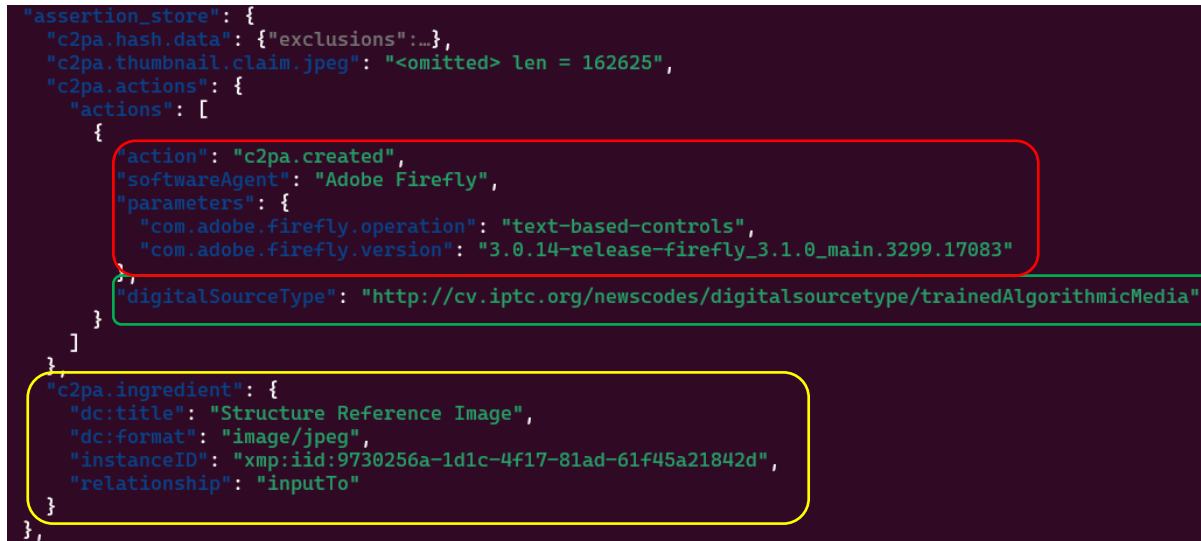


Figure 28: Snapshot of the C2PA assertion of “Chaotic Geneva”.

The assertion included in the C2PA metadata has the *c2pa.created* action label indicating that the content credentials were attached to a created asset. In addition, Firefly’s claim generator includes extra parameters that identify the model used for generation precisely, the type of operation and version of the trained model.

- In red: the model that generated the image
- In green: the IPTC digital source type label, here *trainedAlgorithmicMedia*, which, as documented by the IPTC table, is defined as “digital media created algorithmically using an Artificial Intelligence model trained on captured content”
- In yellow: the structure reference image that was used as input (no data of input image)

Through this process, end-users and AI company alike are therefore able to identify whether C2PA-signed images are AI-generated or not.

## 5. C2PA in Practice

The use cases below illustrate the many ways in which C2PA can prove useful as part of a journalistic practice – as illustrated by a user scenario of a journalist using C2PA to verify images, and by use cases of C2PA protecting professional photography and authenticating audio work.

## 5.1 User Scenario: John the Fact-Checker

---

John is a fact-checker journalist that works at AuthenticNews Corp. His company receives multiple media content from press agencies, news organisations, and a public repository where his team can receive content directly from whistleblowers.

Like every journalist out there, John and his fact checking team are very worried about digesting synthetic media into their workflow and thereby contributing to the spread of misinformation. Basically, the problem boils down to answering the following questions: How was this content produced, and by whom?

Content credentials, or C2PA metadata, is one way to verify the authenticity of provenance information of the content. Let us look at a few scenarios.

### 5.1.1 Happy Path

---

John receives the following image<sup>59</sup> (Figure 29).



Figure 29: Image received by John for inspection.

---

<sup>59</sup> Photograph: 'Voters cast their ballots during general elections in Mexico City on Sunday' by Matias Delacroix, The Associated Press. In (Thomson Reuters, 2024)

John wants to be sure that the image was camera-captured (the “how”) and be sure of the source of the image (the “who”). Luckily, he finds that the image contains content credentials, by inspecting it in a C2PA-compliant verifier such as [contentcredentials.org/verify](https://contentcredentials.org/verify).

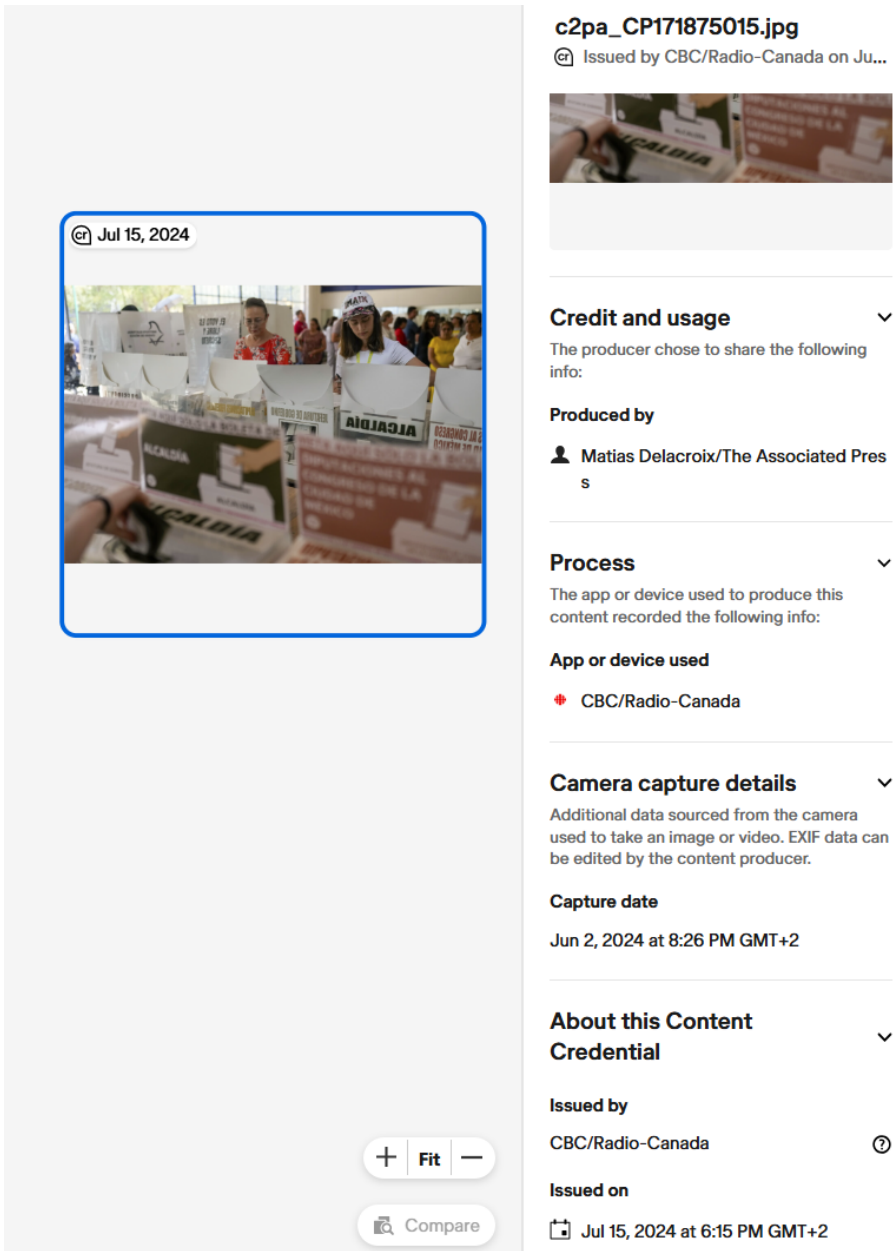


Figure 30: Verified C2PA metadata of the image.

Since the content credentials (see Figure 30) are valid and John and his team trust the “who” (CBC/Radio Canada), John can be assured that the image he is viewing is authentic. He can also rest assured that the image is not synthetic, and that the image capture date is 2 June 2024, according to CBC/Radio-Canada. However, the picture alone lacks context, and he is afraid that it might be misplaced, not belonging to the article that he had received (which was sent by a party that is not necessarily trusted).



He decides to dig deeper into a verification process and verify the C2PA directly, using the opensource command line tool: *c2patool*.

```
"com.truepic:urn:uuid:f3585d1d-5c10-443d-a75e-a3461e28ae44": {
  "claim": {
    "dc:title": "c2pa_CP171875015.jpg",
    "dc:format": "image/jpeg",
    "instanceID": "a4e8b9aa-fd3f-49b4-80d3-acdeb8f8b787",
    "claim_generator": "CBC/Radio-Canada libc2pa/3.8.19",
    "claim_generator_info": [
      {
        "name": "CBC/Radio-Canada",
        "icon": {
          "url": "self#jumbf=c2pa.databases/c2pa.data",
          "hash": "QiKbz/LNqTiDHfprSMZjVpVJI8jqIBxnpwEdN7Q09xs="
        }
      }
    ],
    "signature": "self#jumbf=c2pa/com.truepic:urn:uuid:f3585d1d-5c10-443d-a75e-a3461e28ae44/c2pa.signature",
    "assertions": [...],
    "alg": "sha256"
  },
  "assertion_store": {
    "stds.schema-org.CreativeWork": {
      "@context": "https://schema.org",
      "@type": "CreativeWork",
      "author": [
        {
          "@type": "Person",
          "name": "Matias Delacroix/The Associated Press"
        }
      ]
    },
    "com.truepic.libc2pa": {
      "lib_name": "Truepic C2PA C++ Library",
      "lib_version": "3.8.19",
      "target_spec_version": "1.4",
      "git_hash": "v3.8.19"
    },
    "stds.exif": {
      "exif:DateTimeOriginal": "2024-06-02T18:26:30Z",
      "@context": {
        "exif": "http://ns.adobe.com/exif/1.0/"
      }
    },
    "c2pa.hash.data": {"exclusions": "..."},
    "cbc.custom.assertions": {
      "title": "Mexico Election",
      "description": "Voters cast their ballots during general elections in Mexico City, Sunday, June 2, 2024.",
      "identifier": "CP171875015"
    },
    "c2pa.thumbnail.claim.jpeg": "<omitted> len = 135496"
  },
  "signature": {
    "alg": "es256",
    "issuer": "CBC/Radio-Canada",
    "time": "2024-07-15T16:15:07+00:00"
  }
}
```

Figure 31: Extracted content credentials for the image published by CBC/Radio-Canada.

This new view provides him with in-depth details about the content credentials (Figure 31). He can even read now that CBC/Radio-Canada added a tamper-proof custom assertion about the context of the image, which is labelled under *cbc.custom.assertions*:

*"Voters cast their ballots during general elections in Mexico City, Sunday, June 2, 2024."*

He now can conclude the following:

- CBC-Radio Canada asserted on the 15 July 2024 that:
  - The photo is camera-captured
  - The photo was taken on the 2 June 2024
  - The custom assertion is authentic and is bound to the photo
- If he trusts CBC/Radio-Canada, then he can trust these claims

### 5.1.2 Halfway There

---

John and his team now receive the following image of sharks from an unknown source (Figure 32). He wants to find out if the image is real.



Figure 32: A cropped image of two sharks, a turtle, and a human – up for inspection.

He suspects that the image might have been AI-generated, but he would like to prove it. His teammate Dan disagrees and argues that the image is real.

John: *“That is so silly, no human can get this close to nurse sharks without safety equipment and stay alive”*

Dan: *“Why not? My cousins and I used to have fun with them all the time during summer”*

John: *“Do you want to bet on that? Loser pays for lunch”*

Dan: *“Bet, bring your proof!”*

John scratches his head and gets to work. Unfortunately, content credentials are just metadata under the hood, therefore like any other kind of metadata they can be stripped away from the assets that they authenticate.

He assumes that the original image has provenance information, so he thinks of the following strategy:

1. Look on the web for similar images.
2. Find an image that is visually like the one he has at hand, but with valid content credentials.
3. Come to a conclusion about his findings.

To this end, John decides to try the *search by image* functionality in *contentcredentials.org/verify*. He submits the photo and as expected, no content credentials are included with the image (Figure 33).

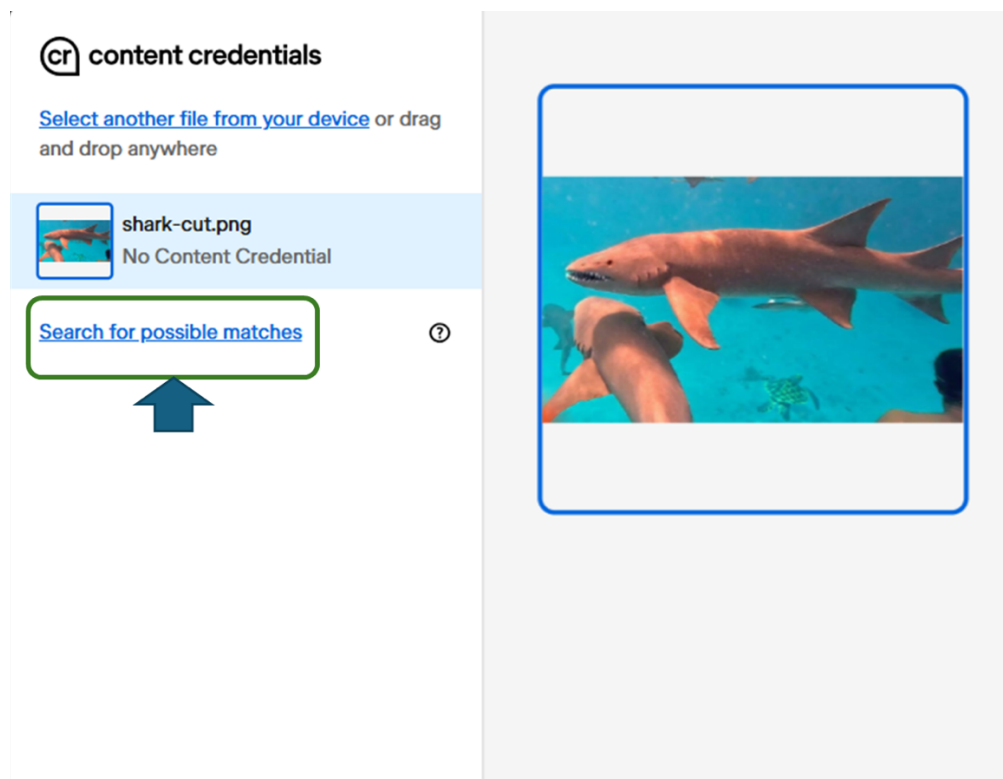


Figure 33: The inspected image has no content credentials.

He decides to search through the publicly available content credentials database for possible matching images. Reverse image searching relies heavily on fingerprinting techniques alongside databases optimisation to efficiently look for visually similar images in a large database.

The image search returns the following results (Figure 34):

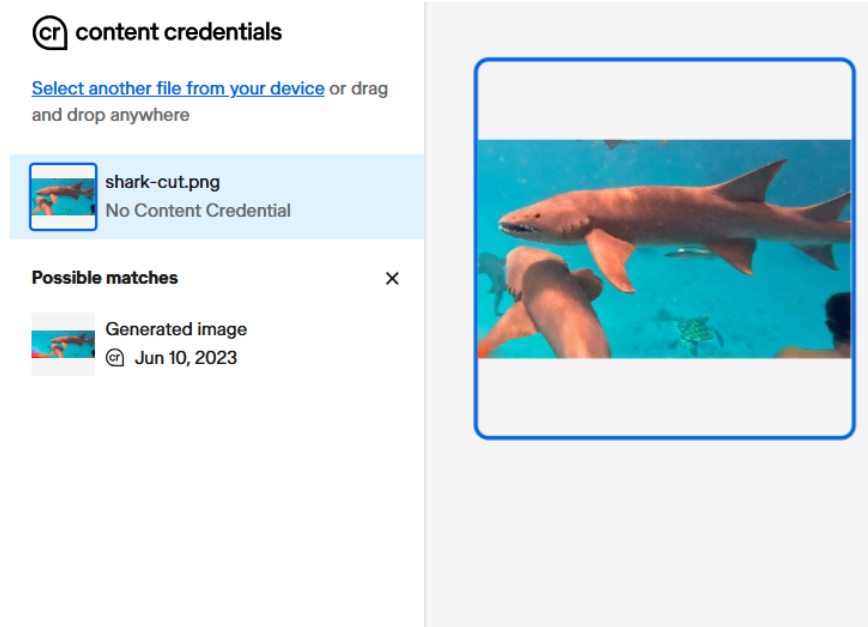


Figure 34: Visual matches of the image in the content credentials verifier database.

BINGO, there is indeed a potential match in the database. John is familiar with forensic fingerprinting, so he is aware that a possible false-positive match may be returned, which is a known limitation of the fingerprinting technique. The false positive rate is related to the fingerprinting algorithm used by the reverse-image search service and should typically be low. He therefore performs a manual visual check on the thumbnail of the asset.



Figure 35: Thumbnail of the matching image.

Even if the result image (Figure 35) is not exactly the same, the content is rather similar. John can still recognise the three sharks, the turtle, the human presence, and a visual watermark that says “Adobe Firefly (Beta)”.

His heart starts beating fast as he continues to inspect the content credentials of the submitted image (Figure 36).

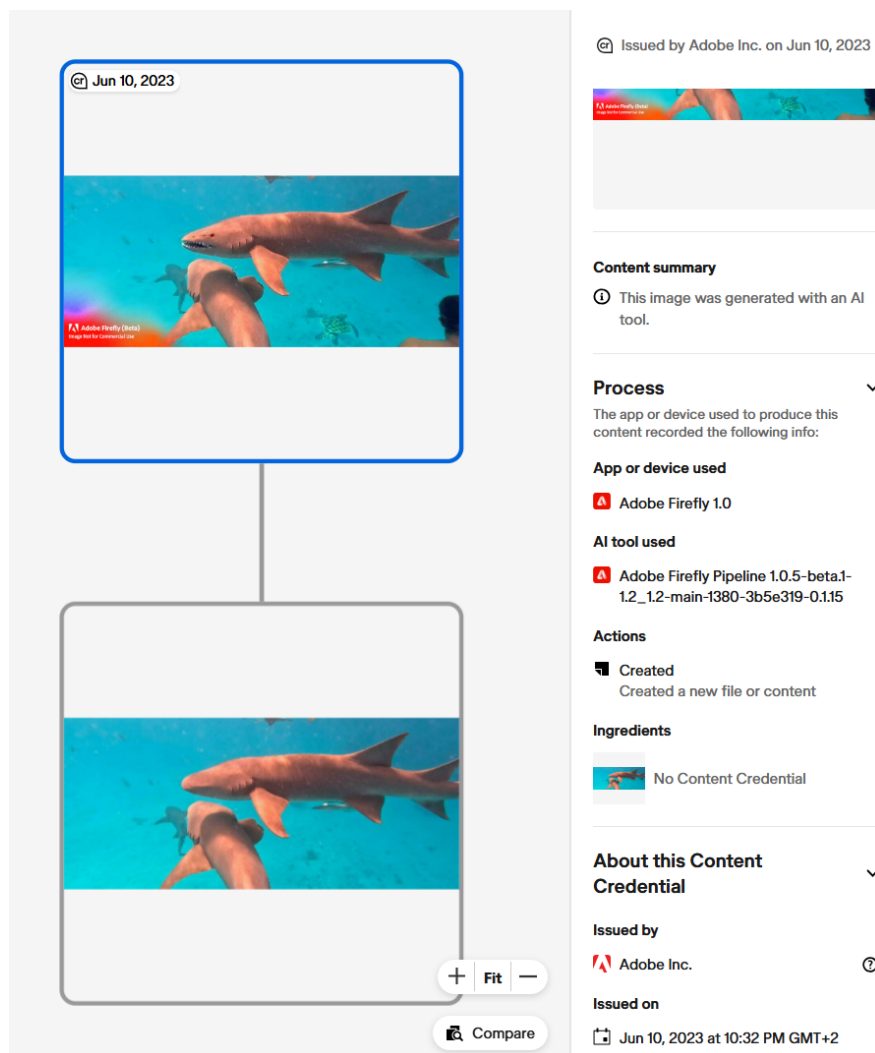


Figure 36: Content credentials of the matching image.

*"I knew it!", John shouts, as the team is gathering around his desk to get a closer look.*

*John: "I have found a similar image to the one submitted to us. The image was indeed edited using generative AI on June 10th, 2023. The content and AI labelling have been signed by Adobe and the signature is valid!"*

Furthermore, the C2PA manifest of the retrieved images also includes the provenance chain. In fact, there is an authentic (signed) thumbnail of the original image, prior to the AI edits, included in the metadata (Figure 37).



Figure 37: The original image prior to AI edits

Lessons learned, even if a content has been stripped, fingerprinting can still be used to retrieve similar content that has C2PA provenance information (this would however require a wider adoption of C2PA technology in reverse image searches such as Google Images and TinEye).

### 5.1.3 I know It Was There

---

After a free lunch, John gets the following image (Figure 38):

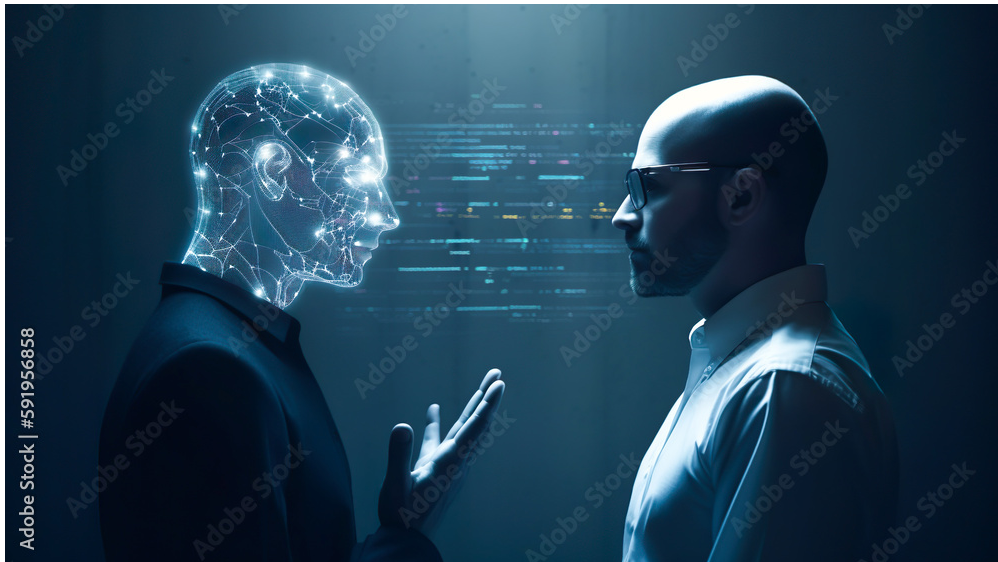


Figure 38: An image without C2PA manifest<sup>60</sup>.

John checks if the image has any C2PA metadata attached. He wants to be fast, so he uses the *c2patool* to read the manifests.

He runs the following script (Figure 39):

---

<sup>60</sup> Source: <https://stock.adobe.com/591956858>. Licensed to the EBU.



```
> c2patool --info machine.jpg
Information for machine.jpg
Provenance URI = self#jumbf=/c2pa/adobe:urn:uuid:a37a6826-acbc-473b-8db6-75ad70459901/c2pa.claim
No C2PA Manifests. (file size = 457621)
```

Figure 39: Content credentials information status of the image.

The tool detects the *Provenance URI* metadata header that points to the location of the C2PA manifest of the image:

self#jumbf=/c2pa/adobe:urn:uuid:a37a6826-acbc-473b-8db6-75ad70459901/c2pa.claim

The *Provenance URI* suggests that the image file (self) used to have in its JUMBF<sup>61</sup> section, a manifest with the UUID *adobe:urn:uuid:a37a6826-acbc-473b-8db6-75ad70459901*. The claim was likely created by Adobe, since it is in the namespace of the identifier, but John is careful not to blindly trust that the image is coming from them, since the *Provenance URI* field is not authenticated and can be manipulated. It only serves as a loose indicator that the C2PA metadata existed in the image but may have been stripped.

John still needs to recover the manifest somehow and perform the validation checks as suggested by the C2PA specification, using, for example, c2patool.

#### 5.1.4 You Can't See It

John receives the following image (Figure 39). He supposes that it was captured by a professional photographer. However, he wants to be sure of it before passing it on to his friends at the editorial news team at AuthenticNews Corp.



Figure 39: Lithuania Independence Day celebration<sup>62</sup> (C2PA stripped).

<sup>61</sup> A box-based serialisation format to embed any type of metadata in any JPEG file format

<sup>62</sup> Photograph by Andrius Aleksandravicius: <https://www.afoto.eu/> (last accessed on 18 July 2025)

Unfortunately, despite a thorough inspection, he cannot find any hint of a matching image that has C2PA credentials. As he is about to give up, he remembers that C2PA supports another type of binding in addition to the hard binding: soft binding!

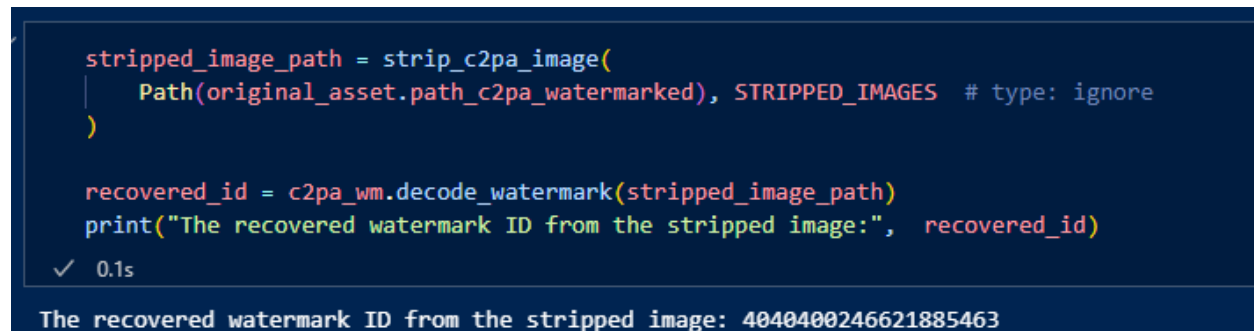
C2PA soft binding relies on watermarking to embed a *payload*, which is usually a reference to the C2PA manifest of the asset. This allows recovery in case:

- the metadata was stripped, or
- the asset has endured non-editorial modifications that do not change the meaning of the content – only its underlying byte representation, causing the hard binding to break.

John takes the image and uploads it to the *find-my-c2pa*<sup>63</sup> image service that detects the watermark and fetches the corresponding manifest from the database.

John is aware that in some cases, the watermark can be broken if the encoding algorithm is not robust enough, or even spoofed if it is insecure. John decides, nevertheless, to shoot his shot: he uploads the image to the *find-my-c2pa* service. He keeps nodding his head feeling concerned that watermarking in C2PA, if not used carefully, can breach user privacy. He understands that this is a powerful technique for content distribution against intellectual property theft, so he carries on.

Once the image is uploaded, the *find-my-c2pa* service tries to decode the watermark from the stripped image and return the database ID related to the photo (Figure 40). The service has a list of several decoding algorithms that it runs until it detects the watermark, otherwise it returns a *no watermark found* response.



```

stripped_image_path = strip_c2pa_image(
    Path(original_asset.path_c2pa_watermarked), STRIPPED_IMAGES # type: ignore
)

recovered_id = c2pa_wm.decode_watermark(stripped_image_path)
print("The recovered watermark ID from the stripped image:", recovered_id)
✓ 0.1s

The recovered watermark ID from the stripped image: 4040400246621885463

```

Figure 40: A snippet of code recovering the watermark from the image.

Assuming that the ID has been detected, the service can either query its database or alternatively retrieve extra information from the ID and/or image itself, to know what custom database to query. It is up to the application developers to design the database as they see fit for the application, just bearing in mind that the watermark payloads are of limited capacities – usually between 40 and 100 bits, depending on the watermarking algorithm and on the type of the media asset.

A simple example database can look as follows (see Figures 41 and 42):

<sup>63</sup> find-my-c2pa is a proof-of-concept service developed by the EBU for this report.



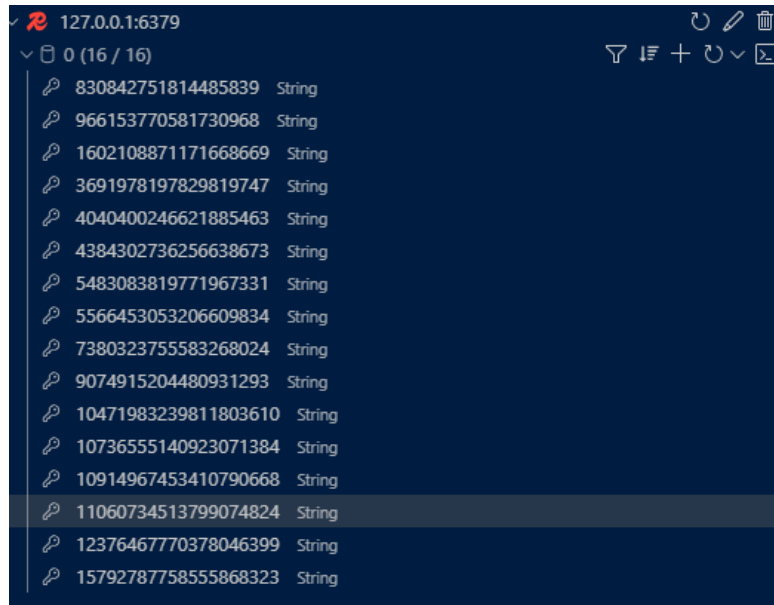


Figure 41: A key-value store database of C2PA assets with IDs as keys.

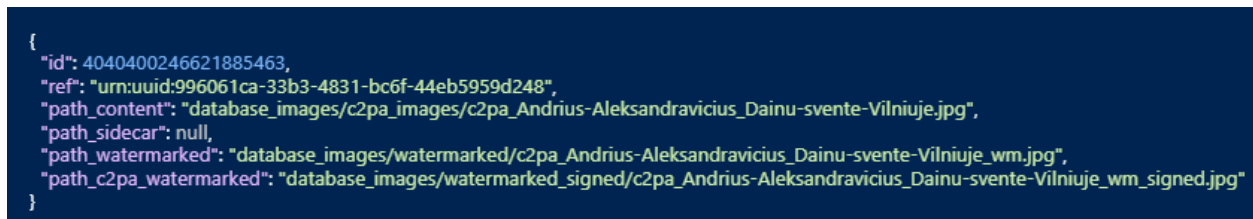


Figure 42: A representation of a C2PA asset in the database.

Each entry in the database is a C2PA asset that has:

- a *unique identifier* which corresponds to the embedded watermark
- a *manifest ID* that identifies the C2PA metadata
- *path\_content*: a path to the C2PA signed asset
- *path\_sidecar*<sup>64</sup>: a path to the C2PA sidecar bundle
- *path\_watermarked*: a path to the watermarked asset
- *path\_c2pa\_watermarked*: a path to the watermarked asset with a soft-binding assertion added to its C2PA credentials.

<sup>64</sup> *path\_sidecar* could be *null* because the C2PA metadata is attached to the image in *path\_watermarked*. C2PA metadata could be packaged separately into a sidecar.



Figure 43: Side-by-side display of the recovered watermarked image.

When inspected, the recovered image turns out to have the C2PA-defined watermarked action (Figure 43). This was inserted by the watermarking actor (e.g. *find-my-c2pa* or an organisation that uses watermarking) prior to the web publication of the image by the photographer.

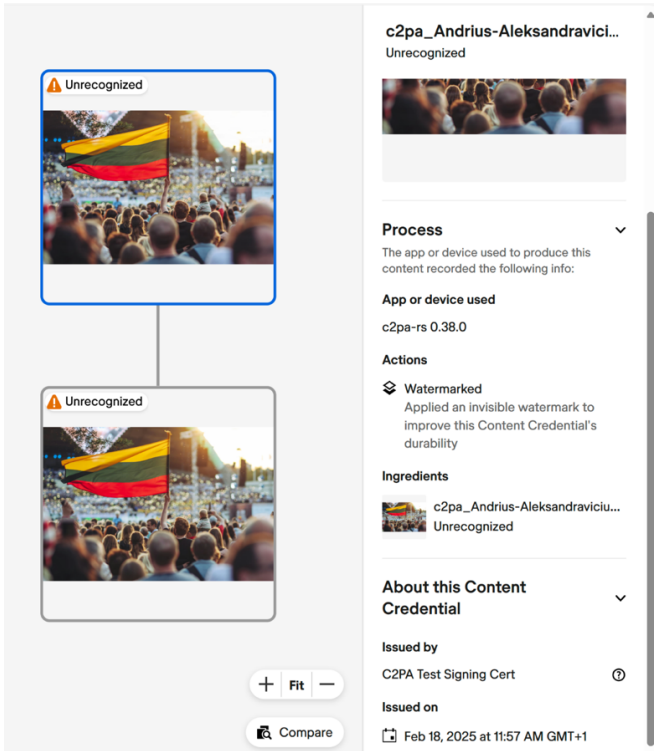


Figure 44: C2PA manifest of the watermarked image<sup>65</sup>.

<sup>65</sup> The watermarked image was signed using a Test C2PA certificate which is not trusted by the verifier, hence the warning message in orange.

John inspects the image in more detail, using the *c2patool* yet again (Figure 44).

```
{
  "active_manifest": "urn:uuid:7c00ee31-f775-4395-9724-15110aef325a",
  "manifests": {
    "urn:uuid:7c00ee31-f775-4395-9724-15110aef325a": {
      "claim": {"dc:title": "..."},
      "assertion_store": {
        "c2pa.soft-binding": {
          "alg": "com.adobe.trustmark.Q",
          "blocks": [
            {
              "scope": {},
              "value": "2*11100000010010011000101001111000001010010011000101010000010111"
            }
          ]
        },
        "c2pa.hash.data": {"exclusions": "..."},
        "c2pa.thumbnail.claim.jpeg": "<omitted> len = 337407",
        "c2pa.ingredient": {"dc:title": "..."},
        "c2pa.actions": {
          "actions": [
            {
              "action": "c2pa.watermarked"
            }
          ]
        },
        "signature": {"alg": "..."}
      },
      "urn:uuid:996061ca-33b3-4831-bc6f-44eb5959d248": {"claim": "..."}
    }
  },
  "validation_status": [...]
```

Figure 45: Parsed C2PA metadata of a watermarked image.

C2PA manifests created for watermarked content are treated in a special way by the C2PA specifications<sup>66</sup>. The C2PA manifest of any watermarked content must additionally contain two components:

1. **c2pa.watermarked (yellow box)**: action describing that the content was watermarked by the signer of the manifest
2. **c2pa.soft-binding (orange box)**: a special assertion within the C2PA manifests that enables binding the C2PA manifest to the watermarked content.
  - *alg*: algorithm that was used for the watermarking, e.g. this image was watermarked using Trustmark
  - *value*: embedded watermark ID in binary format
3. **urn:uuid:996061ca-33b3-4831-bc6f-44eb5959d248 (red box)**: identifier of the ingredient manifest that contains information (provenance information) about the creation of the image by the photographer, which corresponds to the *“ref”* attribute an image item in the database.

<sup>66</sup> C2PA Guidance for signing watermarked content:

[https://spec.c2pa.org/specifications/specifications/2.2/guidance/Guidance.html#\\_watermarks\\_and\\_soft\\_binding\\_assertions](https://spec.c2pa.org/specifications/specifications/2.2/guidance/Guidance.html#_watermarks_and_soft_binding_assertions) (last access on 18 July 2025)

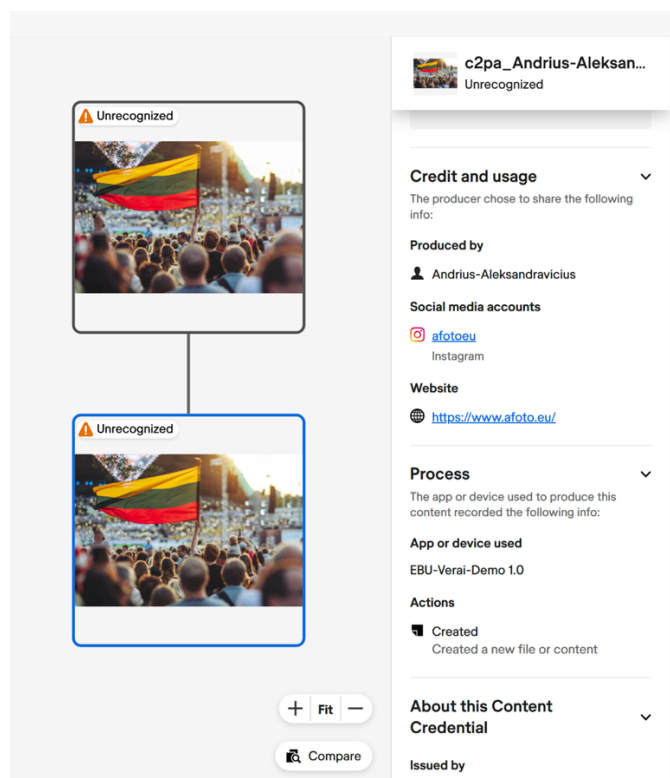


Figure 46: Display of the photographer's manifest urn:uuid:996061ca-33b3-4831-bc6f-44eb5959d248.

Thanks to the soft-binding mechanism, John can not only recover the content credentials of the watermarked image but also that of the original picture, as captured and signed by its photographer. Figure 46 shows the complete provenance information about the picture. John can rest assured that the image is authentic according to its photographer.

## 5.2 C2PA and the AFP IMATAG Proof of Concept

IMATAG<sup>67</sup> is a digital watermarking company. Together with press agency AFP<sup>68</sup> (and AFP's camera provider Nikon), they developed a proof of concept (PoC) to tackle the problem of protecting professional photographic work against the proliferation of synthetic media, i.e. AI-generated images.

The PoC's solution relies on directly authenticating Nikon captured images using the C2PA specification. And then watermarking prior to distribution, in order to track the origin of the image when it gets replicated over the internet. The goal for this solution is to provide fact-checkers with a working technical solution, to distinguish camera-captured images from images from unknown sources. The PoC integrates this solution in the *Fake News Debunker* in the *Verification Plugin*<sup>69</sup> (Teyssou, 2017).

<sup>67</sup> IMATAG: <https://www.imatag.com/> (accessed on 18 July 2025)

<sup>68</sup> AFP (Agence France-Presse): <https://www.afp.com/en> (last accessed on 18 July 2025)

<sup>69</sup> The Verification Plugin was previously called the InVID-WeVerify plugin, as it was first developed within the vera.ai predecessor WeVerify project (<https://weverify.eu/>). It is now further developed within vera.ai, although not the Fake News Debunker specifically: <https://u.afp.com/plugin> (both last accessed on 18 July 2025)

### 5.2.1 Authentication Workflow

The following schema (Figure 47) describes the workflow that enables checking and proving the authenticity of an image.

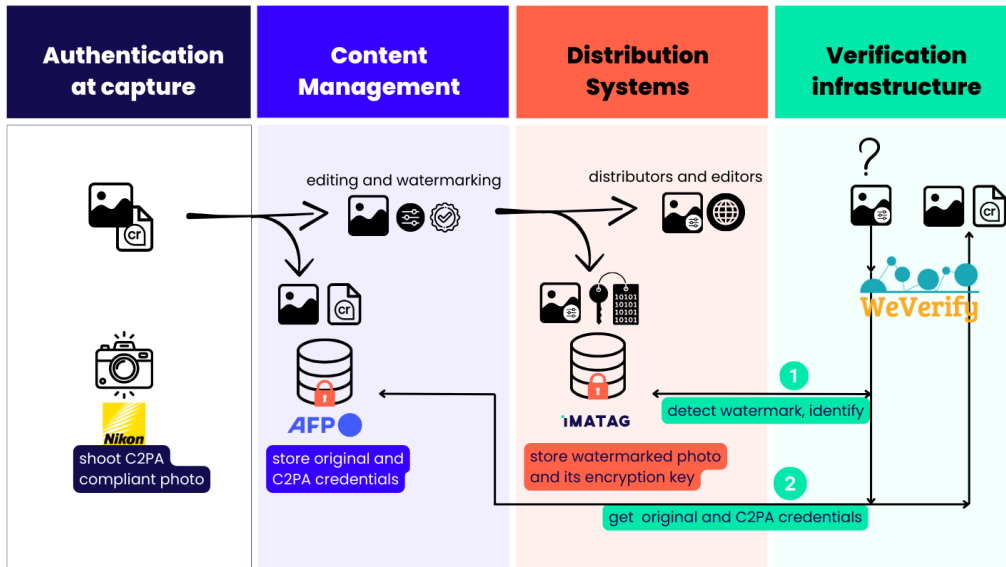


Figure 47: Authentication workflow, from capture to verification (Desoubeaux, 2025).

First, the professional photographers that are the source and origin of the image (e.g., at AFP) take a picture with a C2PA-enabled camera. Leica and Sony were the first to support C2PA capabilities for image capture. Other camera manufacturers such as Nikon, Canon and Fujifilm have also recently (2024) included support for some of their camera models<sup>70</sup>.

The camera generates a C2PA manifest for the content. The metadata information may include the camera model, the date and time of capture, its location, copyright information with contact info for the photographer, and other provenance information. However, no details were disclosed by the camera manufacturers about the metadata that goes into the C2PA manifests.

A copy of the image with its content credentials is stored in the AFP database. The other copy is passed on for editing and watermarking without its C2PA metadata. The reason why the C2PA metadata is not passed on along the rest of the workflow, could for example be that:

- It is not relevant for this PoC
- The rest of the workflow must be C2PA compatible to enable full provenance.

After editing, the edited image gets watermarked using IMATAG technology. The embedded watermark persistently links the edited image to the original manifest. It is very important that the watermark be

<sup>70</sup> Camera manufacturers and camera models supporting C2PA: <https://c2pa.camera/> (last accessed on 18 July 2025)

applied prior to distribution. IMATAG keeps a copy of the edited image and information about the watermark to be able to decode the watermark payload later.

Once watermarked, the edited image is distributed to AFP's editors and third-party publishers. From there news publishers may share the image on their platforms. Images may end up on social media or sent over forums.

Now suppose a fact-checker came across an image during an investigation. Their concerns might be:

- Is the image real (i.e., camera-captured by AFP)?
- Has it been so heavily edited that its meaning has changed?

Using the Verification Plugin, the fact-checker can verify at least whether the image has an ancestor from AFP, meaning that it has been captured by one of AFP's journalists. The verification is as follows:

1. The user uploads the photo to the plugin.
2. The plugin requests an IMATAG watermark detection:
  - a. Detected: the embedded payload is extracted which provides needed identifiers, such as database indices, for step 3.
  - b. Not detected: either
    - i. the watermark was destroyed, hence the importance of the watermarking scheme's robustness
    - ii. the image was never watermarked by IMATAG: one could assume it was out of scope for this workflow.
3. Using the extracted watermark payload, the plugin requests the original image with its C2PA metadata from the AFP database.
4. The user compares the image at hand with the original AFP image and makes their own conclusions.

AFP was able to successfully use this setup to verify the authenticity of its photos during the latest US elections (AFP, 2025).

### 5.2.2 Guarantees

---

The success of the technical solution to the problem described above, relies on the following security properties:

- **Availability:** High availability of both the source image database and the watermark detection servers.
- **Embedded device security:** A secure implementation of C2PA into the camera:

- The signing key should not be extracted from the camera, otherwise this allows attacker to sign any image as camera-captured
- Data cannot be added to the camera's connection between the lens and the signing system. This prevents anyone from tricking the camera into signing fake images without having the signing key.
- **Watermark robustness:** Watermarks can still be detected even if content is edited, altered, or transcoded.
- **Watermark Security:**
  - Watermarks cannot be copied from one content to another i.e., spoofed
  - Watermarks cannot be created and embedded on demand by an attacker, for any content such that it decodes correctly.

## 5.2 C2PA and Audio

To the human ear, AI-generated audio can sometimes be indistinguishable from real content. Since provenance information can also be added to audio content, C2PA can help fact-checkers find the provenance of pieces of audio content.

### 5.2.1 AI-Extended Audio Label

Let us consider the following user scenario. Amy, a video editor, is working on a video edit. She is using a C2PA-enabled video editor that keeps track of the C2PA ingredients she uses in the montage.

The video editor is *next gen* (or next generation): it features AI capabilities that can extend by several seconds a video imported in the track view (yellow box in Figure 48). All she must do is use a ‘magic tool’ provided by the editor and stretch the video. The video editor then generates synthetic frames to AI-extend the original video. This *magic* extend feature is also applicable to audio and gives astonishingly good results. The green box represents the audio component in the editor track view.

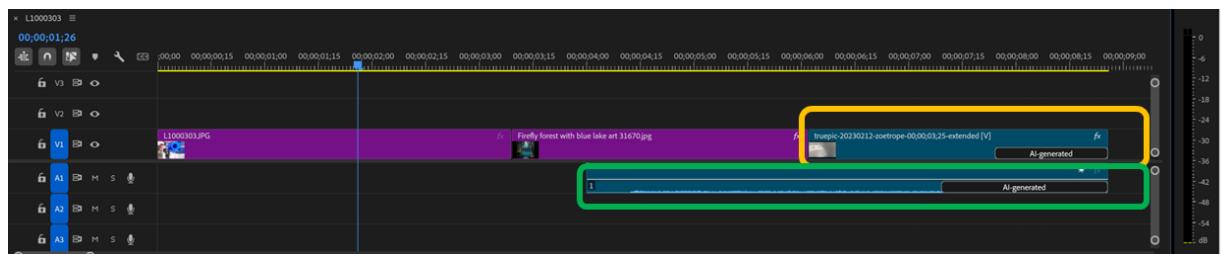


Figure 48: Video editing tool track with AI-generated audiovisual clips.

Exporting the video which has C2PA clips produces the following final video with content credentials. At the root (see Figure 49) is the active manifest of the produced video, and in the green box, the C2PA metadata of the AI-generated audio part. Inspecting the content credentials of these ingredients shows

that the audio extension piece was indeed AI-generated/extended. In this specific example as illustrated in Figures 48 and 49, the audio was generated using Adobe Firefly.

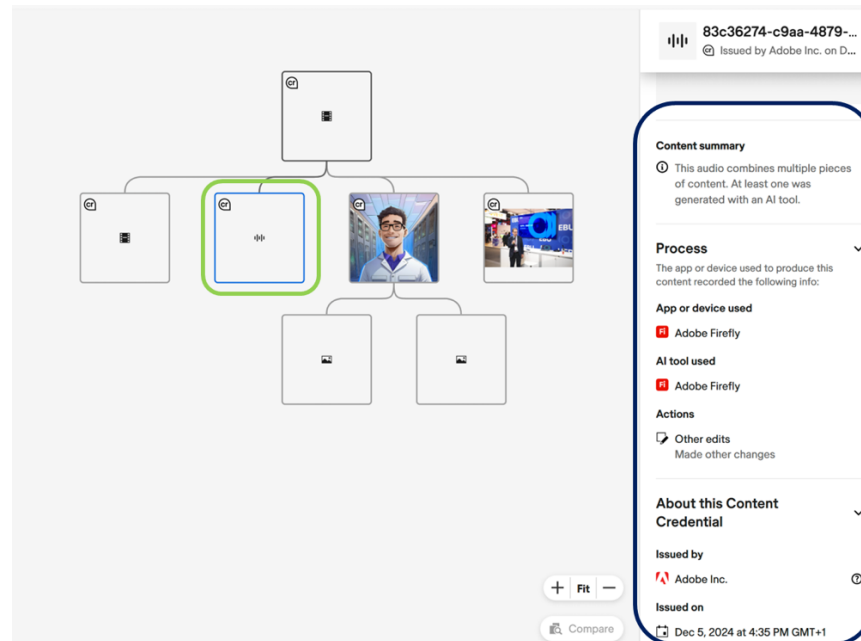


Figure 49: Inspected content credentials of the exported video.

A more technical view of the C2PA metadata of the audio file using the c2patool shows the following (Figure 50):

```
{
  "urn:uuid:aee5ad42-89d2-44d0-ba53-1a887c67d64e": {
    "claim_generator": "Adobe_Firefly_adobe_c2pa/0.12.4_c2pa-rs/0.32.7",
    "title": "Generated Audio",
    "format": "audio/wav",
    "instance_id": "xmp:iid:1d735b4b-5431-44ef-b1fd-2a994081a87e",
    "ingredients": [],
    "assertions": [
      {
        "label": "c2pa.actions",
        "data": {
          "actions": [
            {
              "action": "c2pa.edited",
              "softwareAgent": "Adobe Firefly",
              "parameters": {
                "com.adobe.firefly.version": "0.0.1",
                "com.adobe.firefly.operation": "audio_extend"
              },
              "digitalSourceType": "http://cv.iptc.org/newscodes/digitalsourcetype/compositeWithTrainedAlgorithmicMedia"
            }
          ]
        }
      }
    ],
    "signature_info": {
      "alg": "Ps256",
      "issuer": "Adobe Inc.",
      "cert_serial_number": "28651076926158642445677524766118780318",
      "time": "2024-12-05T15:35:28+00:00"
    }
  },
  "label": "urn:uuid:aee5ad42-89d2-44d0-ba53-1a887c67d64e"
}
```

Figure 50: Technical inspection of content credentials.

The action assertions show one action only, *c2pa.edited*. This is because the audio file was not generated from scratch (or a prompt) but extended from the original audio. The digital source type



*compositeWithTrainedAlgorithmicMedia*<sup>71</sup> is defined as “Augmentation, correction or enhancement using a Generative AI model, such as with inpainting or outpainting operations”

### 5.2.2 C2PA for Audio Translation and Dubbing

Another audio use case is translated or dubbed audio using AI. EuroVOX<sup>72</sup> (an open toolbox developed by the EBU for live and file-based transcription and translation) takes advantage of the rich C2PA metadata framework to transparently convey what action it has performed on a piece of input audio (or video).

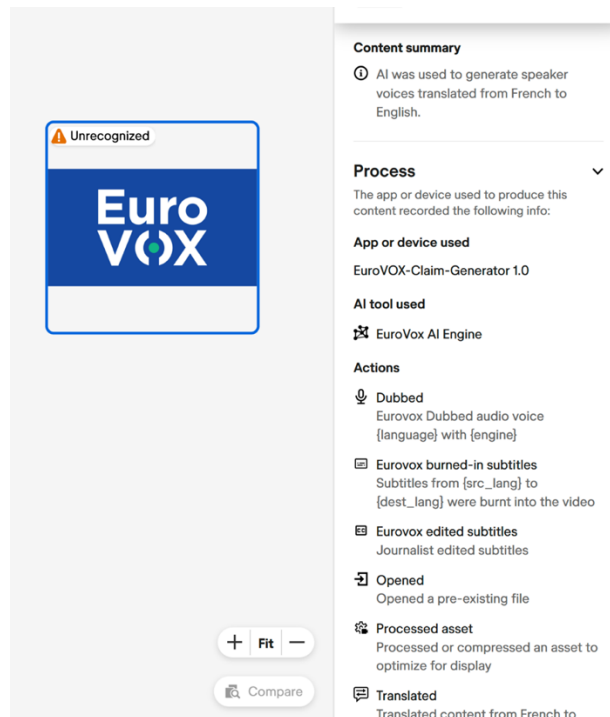


Figure 51: C2PA manifest for a video processed by EuroVox

The manifest in Figure 51, produced by EuroVOX, includes the following actions:

- **Dubbing:** which describes that the audio has been dubbed from a source language to a dubbed language, and optionally provides action parameters such as the AI engine used for the dubbing (e.g., *Amazon Transcribe*<sup>73</sup>, *Google Cloud Speech-to-Text AI*<sup>74</sup>, or *Azure AI Speech*<sup>75</sup>), the version of its AI audio synthesiser, and its voice name as given by the AI engine.

<sup>71</sup> PTC digital source type definition at <https://cv.iptc.org/newscodes/digitalsourcetype/compositeWithTrainedAlgorithmicMedia> (last accessed on 18 July 2025)

<sup>72</sup> EuroVOX: <https://tech.ebu.ch/eurovox> (last accessed on 18 July 2025)

<sup>73</sup> Amazon Transcribe: <https://aws.amazon.com/pm/transcribe/> (last accessed on 18 July 2025)

<sup>74</sup> Google Cloud Speech-to-Text AI: <https://cloud.google.com/speech-to-text> (last accessed on 18 July 2025)

<sup>75</sup> Azure AI Speech: <https://azure.microsoft.com/en-us/products/ai-services/ai-speech> (last accessed on 18 July 2025)

- **Translation:** like dubbing, this action describes the translation of the transcribed subtitles with parameters like the ones described previously

## 6. Potential Integrations of C2PA with vera.ai Tools

Although C2PA and vera.ai tools are already powerful on their own, we believe that the strongest potential in terms of countering disinformation comes from uniting them together. In this section we give a number of examples of how these two approaches can be combined, as illustrated by our proof of concepts.

### 6.1 C2PA and the Database of Known Fakes

The Database of Known Fakes (DBKF)<sup>76</sup> is a searchable archive developed by Graphwise/Ontotext<sup>77</sup> within the vera.ai project, with which users can check whether a claim, image or video has already been verified or debunked by trusted fact-checking sources (International Fact-Checking Network (IFCN)<sup>78</sup> signatories). We consider that C2PA and DBKF joining forces could prove to be a powerful alliance.

Let us look into the following internet post on the X platform<sup>79</sup>, as an example (Figure 52):

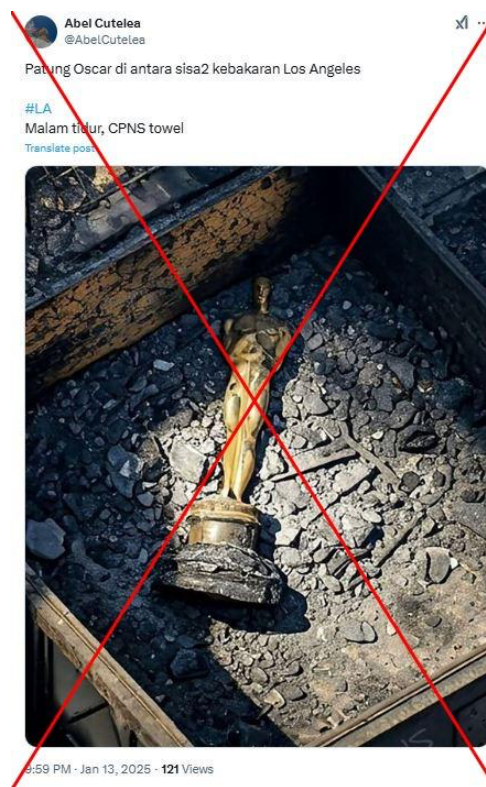


Figure 52: X post with a (fake) image of an Oscar statuette inside a burned house.

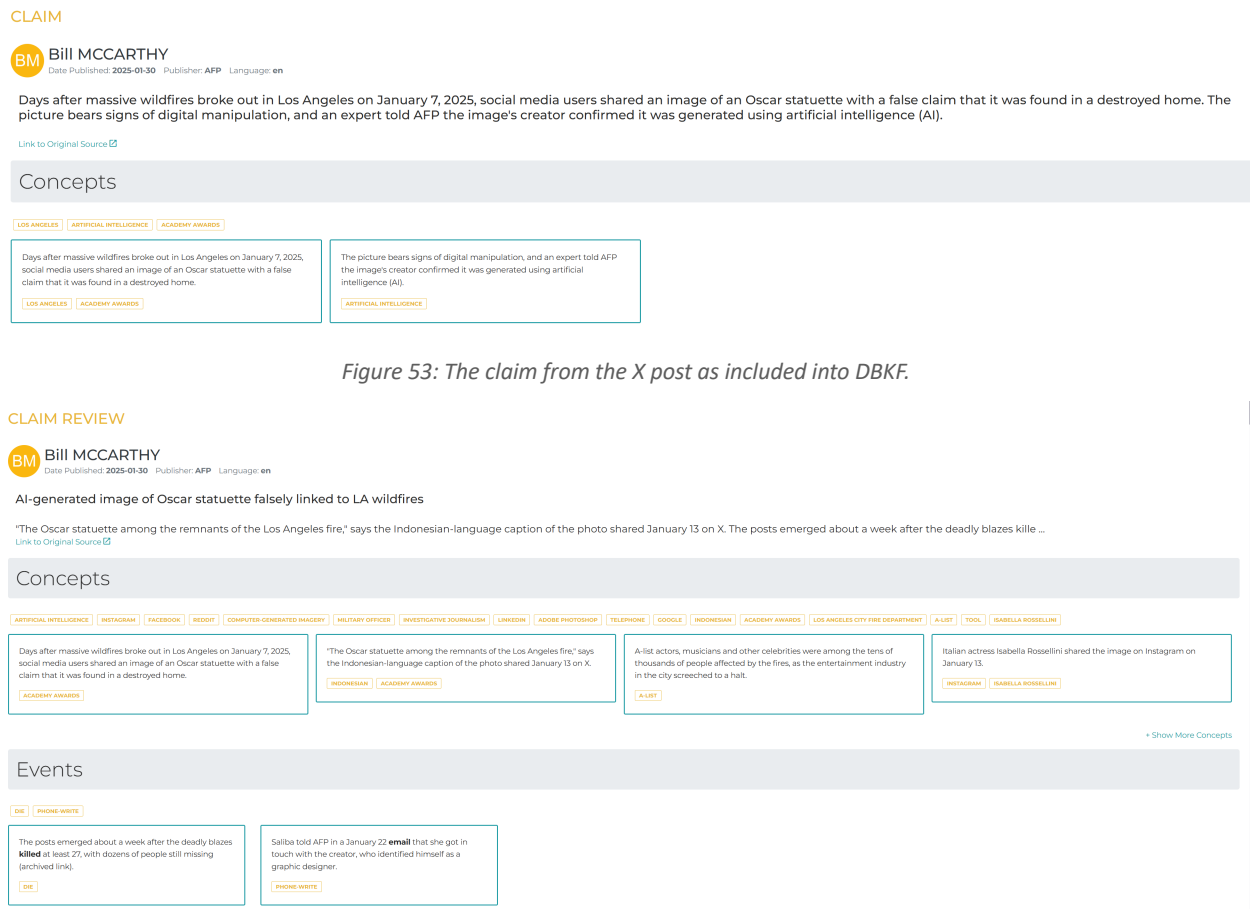
<sup>76</sup> Database of Known Fakes (DBKF): <https://dbkf.ontotext.com/> (last accessed on 18 July 2025)

<sup>77</sup> Graphwise/Ontotext: <https://graphwise.ai/>, <https://www.ontotext.com/> (last accessed on 18 July 2025)

<sup>78</sup> International Fact-Checking Network (IFCN): <https://www.poynter.org/ifcn/> (last accessed on 18 July 2025)

<sup>79</sup> Formerly known as Twitter

A fact-checking review (McCarthy, 2025) about the post was published by a fact-checker at *AFP Fact Check*<sup>80</sup> on 30 January 2025, which was then added to DBKF as both a claim<sup>81</sup> (what people claim through the post) and a claim review<sup>82</sup> (the fact-checking review of that claim), in which the fact-check's conclusion was that the image was AI-generated. See Figures 53 respectively 54.



The results of the fact-check by AFP Fact Check include, among other details:

- An AI content detection scan using the *Synthetic Image Detection* (SID)<sup>83</sup> tool in the Verification Plugin indicates that the image from the post is 95% likely to have been generated with AI.

<sup>80</sup> AFP Fact Check: <https://factcheck.afp.com/> (last accessed on 18 July 2025)

<sup>81</sup> Claim: <https://dbkf.ontotext.com/#!/documentView?uri=http://weverify.eu/resource/Claim/a82d5e59a034ac50c7992061c03f0495> (last accessed on 18 July 2025)

<sup>82</sup> Claim Review: <https://dbkf.ontotext.com/#!/documentView?uri=http://weverify.eu/resource/ClaimReview/6dbecfa79646b3faf61af6b847e639b3> (last accessed on 18 July 2025)

<sup>83</sup> Synthetic Image Detector: Developed by ITI-CERTH (<https://mever.gr/tools/>) and University Federico II of Naples (<https://www.grip.unina.it/>), and currently only available to vera.ai beta-testers (last accessed on 18 July 2025)

- A watermark appearing in the image's bottom-right corner, further indicating that it was fabricated using AI.
- A reference to a LinkedIn<sup>84</sup> post by *GetReal Labs*<sup>85</sup>, where it was also determined that the image "was likely to be synthesized or modified", and likely generated by the AI tool *Google Gemini*<sup>86</sup> then edited in Adobe Photoshop.

Using C2PA, this list of fact-checks could be summarised in a *ClaimReview* metadata structure clearly specifying the dates, the facts in textual form (such as links, a copy of the email from the AI expert, links to other reports and exchanges), and the fact-checkers' identity provided either as an organisation or individual actor(s). The claim review would then be signed to produce a C2PA credentials using DBKF's signing key (Figure 55).

ClaimReview

A Schema.org Type

Thing > CreativeWork > Review > ClaimReview

[more...]

A fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed).

Property	Expected Type	Description
Properties from ClaimReview		
claimReviewed	Text	A short summary of the specific claims reviewed in a ClaimReview.
Properties from Review		
associatedClaimReview	Review	An associated ClaimReview, related by specific common content, topic or claim. The expectation is that this property would be most typically used in cases where a single activity is conducting both claim reviews and media reviews, in which case relatedMediaReview would commonly be used on a ClaimReview, while relatedClaimReview would be used on MediaReview.
associatedMediaReview	Review	An associated MediaReview, related by specific common content, topic or claim. The expectation is that this property would be most typically used in cases where a single activity is conducting both claim reviews and media reviews, in which case relatedMediaReview would commonly be used on a ClaimReview, while relatedClaimReview would be used on MediaReview.
associatedReview	Review	An associated Review.
itemReviewed	Thing	The item that is being reviewed/rated.
negativeNotes	ItemList or ListItem or Text or WebContent	Provides negative considerations regarding something, most typically in pro/con lists for reviews (alongside positiveNotes). For symmetry  In the case of a Review, the property describes the itemReviewed from the perspective of the review; in the case of a Product, the product itself is being described. Since product descriptions tend to emphasise positive claims, it may be relatively unusual to find negativeNotes used in this way. Nevertheless for the sake of symmetry, negativeNotes can be used on Product.

Figure 55: Schema of ClaimReview metadata structure<sup>87</sup>.

A concrete instantiation of the *ClaimReview* schema could be done as shown in Figure 56:

<sup>84</sup> Archived LinkedIn post: <https://perma.cc/F6GS-MZWY> (last accessed on 18 July 2025)

<sup>85</sup> GetReal security: <https://www.getrealsecurity.com/> (last accessed on 18 July 2025)

<sup>86</sup> Google Gemini: <https://gemini.google.com/> (last accessed on 18 July 2025)

<sup>87</sup> <https://schema.org/ClaimReview> (last accessed on 18 July 2025)

```

"stds.schema.org.ClaimReview": {
  "@type": "ClaimReview",
  "author": {
    "name": "AFP Fact Check",
    "@type": "Organization"
  },
  "@context": "http://schema.org",
  "publisher": {
    "name": "Database of known fakes",
    "@type": "Organization"
  },
  "reviewBody": "'The Oscar statuette among the remnants of the Los Angeles fire,' says the Indonesian-language caption of the photo shared January 13 on X. The posts emerged about a week after the deadly wildfires killed at least 27, with dozens of people still missing (archived link). A-list actors, musicians and other celebrities were among the tens of thousands of people affected by the fires, as the entertainment industry in the city screamed to a halt. Italian actress Isabella Rossellini shared the image on Instagram on January 13. The picture also spread in online media reports and on Facebook -- including in Greek, Spanish, Turkish, Korean, Persian, Burmese, Chinese and French. However, the image is AI-generated. The InVID-WeVerify verification tool from veraai.eu, a project in which AFP is a partner, indicated there was a 95 percent chance the picture was created with AI. This detector identifies specific traces left by AI image generation software. A reverse image search revealed the picture was previously posted January 12 on the social media platform Reddit and titled 'Symbolism' (archived link). The picture was later removed and replaced with a note saying it violated the forum's rules barring AI-generated pictures. A watermark also appears in the image's bottom-right corner, further indicating it was fabricated using AI. Emmanuelle Saliba, chief investigative officer at GetReal Labs, said in a January 15 LinkedIn post that the deepfake detection company's systems determined the image 'was likely to be synthesized or modified' (archived here). Saliba told AFP in a January 22 email that she got in touch with the creator, who identified himself as a graphic designer. He told her he originally shared it in a Facebook group dedicated to AI creations and digital art and that his intention was 'not to mislead people.' Saliba added that the creator told her he used Google's generative AI tool -- Gemini -- to generate the image on his phone before editing it in Adobe Photoshop. 'This is an example of how quickly an image can spread and be taken out of context,' she said. 'This will become more common, especially since we are moving towards a world where this type of content can easily be created on our phones.'",
  "contributor": {
    "name": "Bill MCCARTHY",
    "@type": "Person",
    "affiliation": {
      "name": "AFP US",
      "@type": "Organization"
    }
  },
  "itemReviewed": {
    "url": "https://perma.cc/Z9XQ-5JNM",
    "@type": "ImageObject",
    "caption": "The Oscar statuette among the remnants of the Los Angeles fire",
    "thumbnailUrl": ""
  },
  "reviewRating": { "@type": "... },
  "claimReviewed": "'Days after massive wildfires broke out in Los Angeles on January 7, 2025, social media users shared an image of an Oscar statuette with a false claim that it was found in a destroyed home. The picture bears signs of digital manipulation, and an expert told AFP the image's creator confirmed it was generated using artificial intelligence (AI)',
  "datePublished": "30 January 2025"
},
},
"signature": {
  "alg": "es256",
  "issuer": "The Database of Known Fakes",
  "time": "2025-07-31T19:10:38+00:00"
},
},
"validation_results": { "activeManifest": ... },
"validation_state": "Valid"

```

Figure 56: Generated ClaimReview in content credentials for the X post image.

Inside the red box of this PoC is the C2PA *ClaimReview* assertion that describes the claim review. It consists of the following parts:

- The *author's* field describes the organisation that performed the fact-check – in this specific case, AFP Fact Check.
- The *publisher* of the claim review is AFP Fact Check on DBKF.
- The *reviewBody* is a textual summary of the full claim review by the fact-checking organisation. Here for instance, it describes the fact-checking methodology and results as listed above.
- The lead fact-checker who debunked the claim and their review can also be referenced as *contributor*.
- The *itemReviewed* entry is of the *ImageObject* type and depicts the content under review. It contains information such as the *caption*, *URL* of the X post, and *thumbnailUrl* of the image.

- The *claimReviewed* entry contains the text of the claim that is being reviewed by the current ClaimReview. The date in the claim review that was published in DBKF is stored in the *datePublished* attribute.

Inside the orange box is the C2PA signature that could be generated by the DBKF claim generator, which has as a role to cryptographically protect the provenance of the claim information in the C2PA manifest for the image and make the tampering detectable. Note that the time of the C2PA signature is different from the time of publication. Although the C2PA creation was done at a later time for the purpose of this PoC, usually in a real-life situation, *datePublished* and time of the C2PA signature should be very close.

Moreover, to signal that this image has been fact-checked, a custom action may be added to the *c2pa.actions* assertion: here “Fact-checked. This image was fact-checked by AFP Fact Check” (Figure 57).

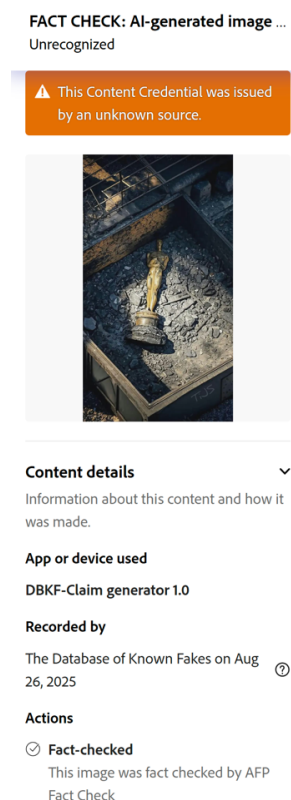


Figure 57: Fact-checked action for *itemReviewed*<sup>88</sup>.

The addition of the fact-checked C2PA metadata to the debunked claims images would be beneficial in three ways:

- The fact-checks would be hard-bound to the asset that they review; any tampering with them or the content would be detected by a C2PA verifier.

<sup>88</sup> Note that the orange warning message is due to the certificate of DBKF being unrecognized by the C2PA verifier, which is to be expected as this is only a self-generated test certificate for this PoC.



- Any other fact-checker or a user could use DBKF to media-reverse search and fetch the fact-check information related to the piece, without spending more resources debunking the claim.
- Other C2PA verifiers could use DBKF to check if images have been flagged as fake and show users the result of the fact-checks. This would be possible thanks to:
  - Fingerprinting technology to find similar images.
  - The defined *ClaimReview* metadata schema, which allows for interoperability between different combinations of C2PA claim generators and displays.

## 6.2 C2PA and the Keyframe Selection & Enhancement Service

The *Keyframe Selection and Enhancement Service* (KSE)<sup>89</sup> is a vera.ai tool developed by ITI-CERTH<sup>90</sup> for speeding video analysis. It provides relevant keyframes, organised on a timeline, and highlights faces and text present in the video for further analysis.

One of the use cases of C2PA in conjunction with KSE could be to embed the provenance of keyframes that have been extracted from the video, alongside any corresponding metadata.

Let us consider keyframe #3 from the collection of keyframes extracted with KSE from a video of Michael Hayden, former director of the US National Security Agency (NSA)<sup>91</sup>, talking about the applications and usage of metadata in information security (Figure 58).

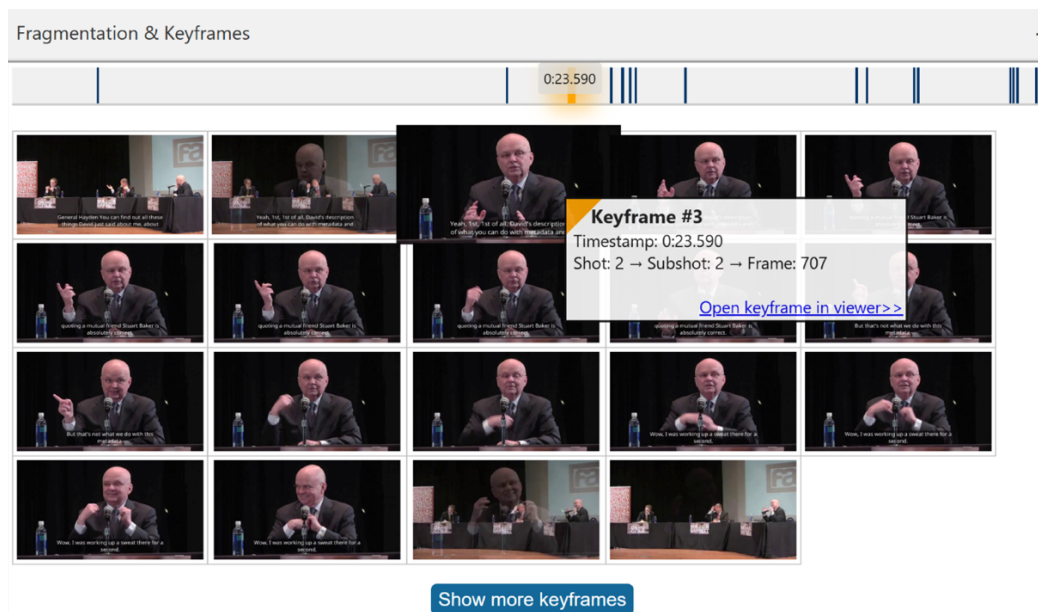


Figure 58: Extracted keyframes of a video of NSA director Michael Hayden at a conference.

<sup>89</sup> Keyframe Selection and Enhancement Service (standalone version): <https://kse.idt.iti.gr/service/start.html> (last accessed on 18 July 2025). Another version of this service exists as part of the Verification Plugin.

<sup>90</sup> Intelligent Digital Transformation Laboratory, Information Technologies Institute, CERTH (Center for Research & Technology Hellas): <https://idt.iti.gr/index.html> (last accessed on 18 July 2025)

<sup>91</sup> National Security Agency (NSA): <https://www.nsa.gov/> (last accessed on 18 July 2025)

The KSE tool could support C2PA to enable provenance of the keyframes. This would be especially useful in cases where the ingested video already has provenance metadata. In the following PoC, the example video has already some provenance metadata: it was translated and dubbed to another language before it was brought into KSE (Figure 59).

Once the video processed, the KSE tool could create a new manifest for the keyframe with all the accompanying metadata such as the *keyframe index*, the *frame number*, the *shot and subshot number*, and finally the *timestamp* from the video.

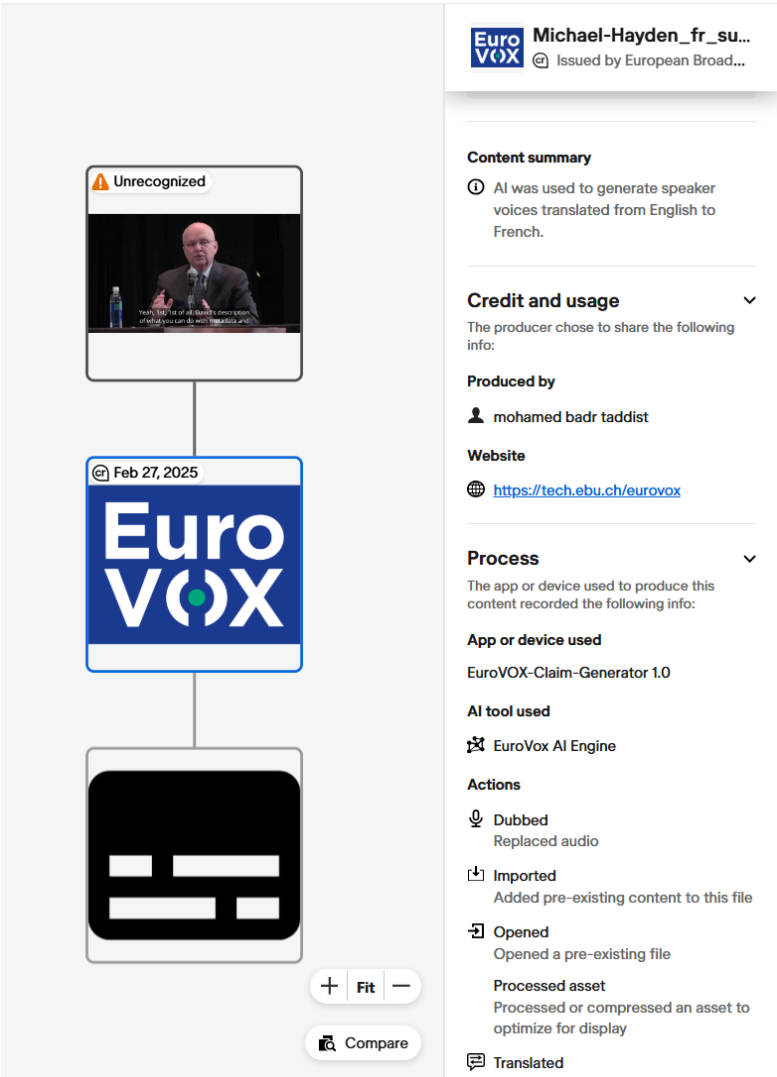


Figure 59: Inspection of the C2PA provenance data of the original video.



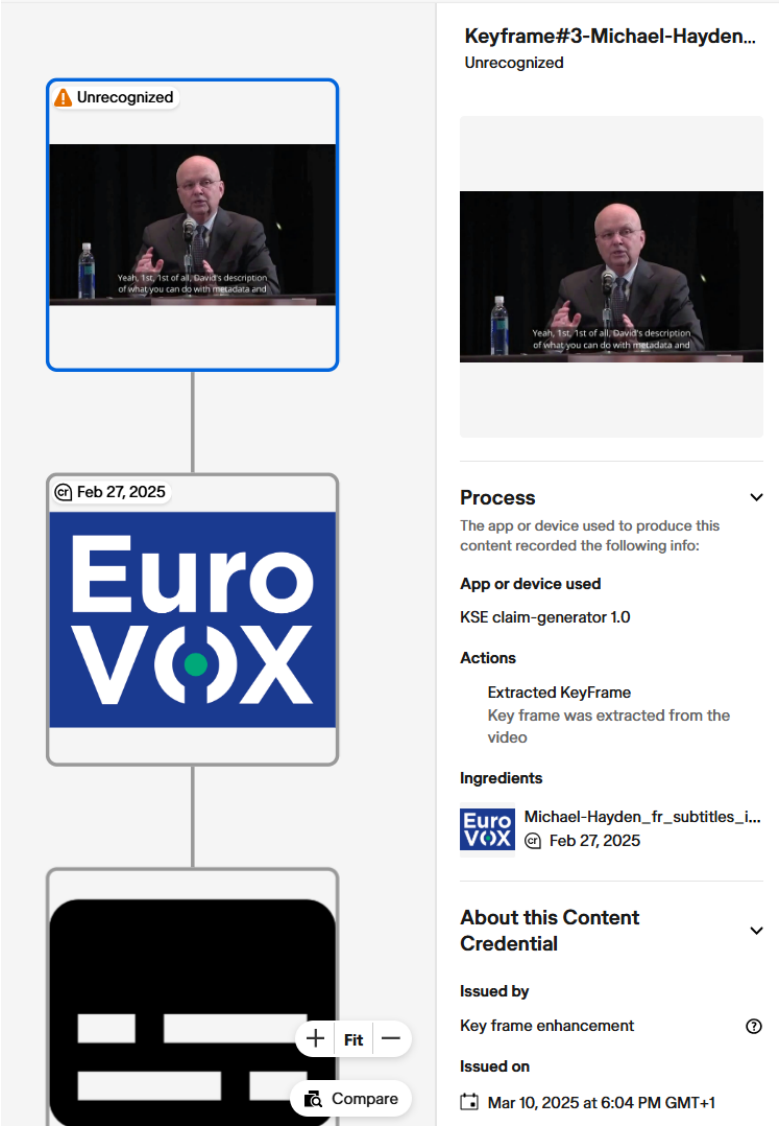


Figure 60: Inspection of the created manifest by the KSE claim generator.

Since the creation of the new C2PA manifest, along with its signature, could be costly (especially if the number of keyframes is significant), KSE could implement a lazy C2PA signing where the C2PA metadata creation is not performed until a keyframe download is requested by a user (Figure 60).

Inspecting the manifest created by KSE in more details would result in the following (Figure 61):

```

"urn:uuid:10e1c772-73e3-470e-9157-b6086290fale": {
  "claim_generator": "kse_claim-generator/1.0",
  "claim_generator_info": [
    {
      "name": "KSE claim-generator",
      "version": "1.0",
      "org.cai.c2pa.rs": "0.46.0"
    }
  ],
  "title": "Keyframe#3-Michael-Hayden_fr_subtitles_in_original",
  "format": "image/jpeg",
  "instance_id": "xmp:iid:59adb89f-0d89-4810-b09d-0331500c3d47",
  "thumbnail": {
    "format": "image/jpeg",
    "identifier": "self#jumbf=/c2pa/urn:uuid:10e1c772-73e3-470e-9157-b6086290fale/c2pa.assertions/c2pa.thumbnail.claim.jpeg"
  },
  "ingredients": [
    {
      "title": "Michael-Hayden_fr_subtitles_in_original.mp4",
      "format": "video/mp4",
      "instance_id": "xmp:iid:79e60b08-915e-4ab0-a65c-61ecd43c4182",
      "thumbnail": {
        "format": "image/png",
        "identifier": "self#jumbf=/c2pa/urn:uuid:13e02da0-6b04-4a6a-8314-cf9986caa493/c2pa.assertions/c2pa.thumbnail.claim.png"
      },
      "relationship": "componentOf",
      "active_manifest": "urn:uuid:13e02da0-6b04-4a6a-8314-cf9986caa493"
    }
  ],
  "assertions": [
    {
      "label": "c2pa.actions",
      "data": {
        "actions": [
          {
            "action": "kse.extracted",
            "parameters": {
              "kse.keyframe.shot": 2,
              "kse.keyframe.subshot": 2,
              "kse.keyframe.frame_n": 707,
              "kse.keyframe.timestamp": "00:28.280",
              "kse.keyframe.n": 3,
              "name": "Extracted KeyFrame"
            }
          }
        ],
        "metadata": {"localizations": "..."}
      }
    }
  ],
  "signature_info": {
    "alg": "Es256",
    "issuer": "Key frame enhancement",
    "cert_serial_number": "440564697469053611989247522184649911363934173816",
    "time": "2025-03-10T17:04:07+00:00"
  }
}

```

Figure 61: KSE-generated manifest for the extracted keyframe #3.

The manifest in this PoC includes a *title* of the keyframe image, a *thumbnail* for quick visualisation upon inspection in the verifier [contentcredentials.org/verify](https://contentcredentials.org/verify), and an *ingredients* array that references the C2PA manifest of the input video to the KSE tool. Most importantly, a custom action by the KSE tool is included in the actions array, signalling that a keyframe extraction took place that resulted in the keyframe. Extra metadata that describes the creation process of the keyframe is also provided as part of the *parameters* entry of the *kse.extracted* action. Finally, the C2PA signature provides its timestamp to prove that the keyframe was generated by KSE tool at *time*.

To increase the durability of keyframe manifests, KSE could implement durable content credentials by both watermarking and fingerprinting the individual keyframes. The resulting C2PA manifests of the watermarked and fingerprinted keyframes would need to be hosted by a manifest repository. The soft-binding resolution mechanism as described by version 2.2 of C2PA, would allow any factchecker to query and retrieve the C2PA information, as well as a link to the original video. The C2PA verifier tool could either be KSE itself or an auxiliary one, such as the Verification Plugin. The verifier tool reverse-searches the keyframe within the database by relying on the keyframe fingerprint or the detected watermark on the keyframe.

In cases where the original metadata would not have C2PA metadata attached, the soft-binding resolution could be used to fetch the original manifest. Again, this assumes that the video was watermarked by its

publisher and its manifest hosted within a public manifest repository; or equivalently, a fingerprinting was pre-computed and used to index the manifest of the video within a manifest repository.

## 6.3 C2PA and the Geolocalizer Tool

---

A very interesting use case of content provenance technologies is their integration with detection tools and approaches.

In fact, C2PA could be used to secure the integrity of the detection results as reported by the detection tools. Specifically, the detection report could be C2PA-signed and stored with the corresponding media in a dedicated and large database. This would prevent people from running the same set of detection tools on the same set or very similar media content, therefore skipping the detection time overhead, and thereby saving on computational cost and fact-checker time.

When running the detection tool on a previously processed or visually similar media, the tool could present the most up-to-date version of the C2PA manifest, incorporating detection results and tool version, as well as their history in previous versions. This approach introduces four implementation challenges:

1. **Storage:** A large-scale database is required for storing processed media.
2. **Strong perceptual algorithm:** a strong perceptual hashing algorithm with high precision and recall is required to compare the visual similarity of the under-inspection media with processed media already present in the database.
3. **Optimised image lookup:** because of the large size of the C2PA database, an optimised image lookup algorithm is a needed.
4. **Efficient policy strategy:** needed for handling conflicting detection results on perceptually (according to the algorithm) similar media.

Efficient solutions to challenges 1, 2 and 3 exist that could be implemented with reasonable overhead for images. However, challenge 2 is computationally demanding to meet for video, for an application requiring high precision and recall. Challenge 4 is up to the application developers.

We illustrate below how C2PA could be used to store the results obtained from vera.ai detection tools, specifically into the vera.ai the Geolocalizer tool, a service developed by ITI-CERTH<sup>92</sup> that aims to infer the geographical location of a query image, using the visual content of the image. C2PA could be integrated with this tool to enable keeping track of the detection reports across multiple versions of the tool, as an authentic provenance chain.

Given a query image, the Geolocalizer tool outputs several estimates in order of likelihood, including:

- **Latitude and longitude:** Geo-coordinates estimated by the tool
- **Confidence level** that the tool has in each geo-coordinate.

---

<sup>92</sup> The Geolocalizer tool, also called Location Extractor in the Verification Plugin, is also developed by the MeVer group at ITI-CERTH: <https://mever.gr/location/> (password-protected, last accessed on 18 July 2025)

- **Similar images with known locations**

Let us consider the following image of an area called Marieberget, in Stockholm, Sweden, as an example (Figure 65):



Figure 65: Marieberget, Stockholm, Sweden<sup>93</sup>

Running the Geolocalizer on this image returns two estimates, as shown in Figure 66. The most probable location is, with a 61% confidence, is the city centre of Stockholm, Sweden, or as indicated on the map, Marieberget on the north side of the Södermalm island. The tool's estimated location is not only correct but also very precise. The tool provides similar images to the query image, which in this case showcase Marieberget from different angles and other buildings with similar architecture on the opposite side of the water. The second most likely location (36%) is the island of Skeppsholmen, also in the Stockholm city centre, which is very close to the first estimate shown in the maps overview on Figure 66.

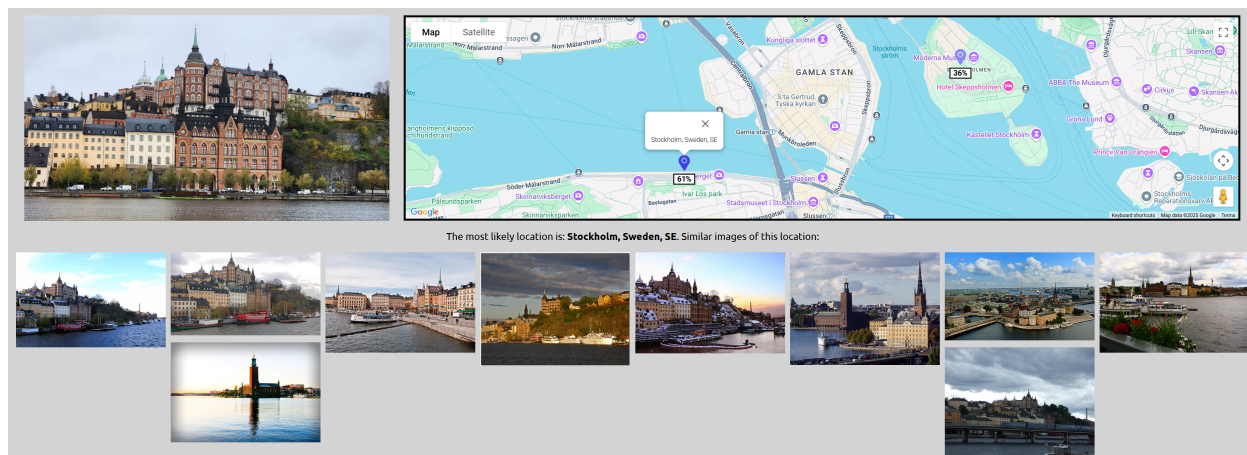


Figure 66: Extracted results from the image in Figure 65

<sup>93</sup> Image source: <https://www.kth.se/blogs/studentblog/2024/05/fags-your-essential-guide-to-starting-at-kth/> (last accessed on 18 July 2025)

This tool could use C2PA to protect the integrity of the results and cryptographically bind them to the processed image. Figure 67 shows the C2PA manifest that could be generated by the tool, incorporating the geo-localisation estimates. The manifest could include:

- The *version* of the tool used to extract the localisation estimates and sign the C2PA manifest.
- The location estimates which are saved as metadata under the tool custom *le.geolocalized* action.
- A secure trusted C2PA *timestamp* that proves the real time at which the estimation took place.

```
{
  "active_manifest": "urn:uuid:4227fc76-0cbb-4f50-a206-ec4b3a6c1d59",
  "manifests": {
    "urn:uuid:4227fc76-0cbb-4f50-a206-ec4b3a6c1d59": {
      "claim": {
        "claim_generator": "location_extractor/1.0",
        "claim_generator_info": [
          {
            "name": "Location Extractor",
            "version": "1.0",
            "org.cai.c2pa_rs": "0.58.0"
          }
        ],
        "signature": "self#jumbf=/c2pa/urn:uuid:4227fc76-0cbb-4f50-a206-ec4b3a6c1d59/c2pa.signature",
        "assertions": [
          {
            "dc:format": "image/jpeg",
            "instanceID": "xmp:iid:90b5a981-5e25-4bd1-aa8d-a0e298084b9c",
            "dc:title": "Geolocalizer Example",
            "alg": "sha256"
          }
        ],
        "assertion_store": {
          "c2pa.thumbnail.claim.jpeg": "<omitted> len = 152300",
          "c2pa.actions.v2": {
            "actions": [
              {
                "action": "le.geolocalized",
                "parameters": {
                  "name": "Geolocalization predictions",
                  "results": [
                    {
                      "latitude": 59.321367,
                      "longitude": 18.059806,
                      "confidence": 0.60816085,
                      "similar_images": [
                        "https://example.com/image1.jpg",
                        "https://example.com/image1.jpg"
                      ]
                    },
                    {
                      "latitude": 59.326009,
                      "longitude": 18.083539,
                      "confidence": 0.35996482,
                      "similar_images": [
                        "https://example.com/image3.jpg",
                        "https://example.com/image4.jpg"
                      ]
                    }
                  ]
                }
              }
            ]
          }
        ],
        "metadata": {"localizations": "..."},
        "c2pa.hash.data": {"exclusions": "..."}
      },
      "signature": {
        "alg": "es256",
        "issuer": "Fake News Debunker",
        "time": "2025-07-31T17:58:17+00:00"
      }
    }
  },
  "validation_results": {"activeManifest": "..."},
  "validation_state": "Valid"
}
```

Figure 67: C2PA manifest for the Geolocalizer tool.

When a new version of the tool is released, detection could be re-run and detection results recorded in an additional (updated) manifest to the content credentials of the image, with potentially more accurate detection results and a reference to the previous manifest, thereby providing the history of detections as a C2PA provenance trail. Similarly to the parent manifest, this updated C2PA manifest would include the new version number of the tool, the detection results of running the new version of the tool, an

introduction to updated localisation methods, an updated timestamp, and additionally an ingredients array referencing the parent manifest (Figure 64).

Thereby, integrating C2PA with detection tools such as the Geolocalizer tool would enable fact-checkers to authentically report its detection results and include them as verifiable proof in investigations.

## Conclusion

---

Through this report, we have dived into the C2PA standard, explored its functioning, its benefits in determining provenance and authenticity of digital content. We have also illustrated these benefits with a number of use cases in the journalistic context, and demonstrated the value of possible integrations with different vera.ai tools. As illustrated by the proof of concepts described in section 6, with additional resources C2PA could be integrated with several of the vera.ai tools and become part of their ecosystem, notably by securing the integrity and retrievability of the detection results by binding them into verified digital content. We believe strongly that C2PA and vera.ai are two complementary parts of the same solution that will allow newsrooms and citizens alike to be better equipped in their fight against disinformation.

Although the current implementation of C2PA still presents a number of challenges (see section 3.4), the combined effort of all partners involved will guarantee that these are tackled in the near future and that C2PA becomes the standard for trusted digital content.

## References

---

- EBU Technical Committee (2024). EBU TC Calls for Industry Collaboration on C2PA and Concerted Approach to Governance. <https://tech.ebu.ch/news/2024/08/ebu-tc-calls-for-industry-collaboration-on-c2pa-and-concerted-approach-to-governance> (last accessed on 18 July 2025)
- Boeyen, S., Santesson, S., Polk, T., Housley, R., Farrell, S., Cooper, D. (2008). Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. <https://datatracker.ietf.org/doc/html/rfc5280> (last accessed on 18 July 2025)
- Zuccherato, R., Cain, P., Adams, C., Pinkas, D. (2001). Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). <https://datatracker.ietf.org/doc/html/rfc3161> (last accessed on 18 July 2025)
- Monday, L. & Strappelli, L. (2024). Does provenance build trust? BBC News Lab. <https://www.bbc.com/rdnewslabs/news/does-provenance-build-trust> (last accessed on 18 July 2025)
- Halford, C. (2024). Mark the good stuff: Content provenance and the fight against disinformation. BBC R&D. <https://www.bbc.com/rd/articles/2024-03-c2pa-verification-news-journalism-credentials> (last accessed on 18 July 2025)
- Ellis, L. (2023) Increasing trust in content: Media provenance and Project Origin. BBC R&D. <https://www.bbc.co.uk/rd/blog/2023-10-media-provenance-watermarks-fingerprints-deepfake> (last accessed on 18 July 2025)



Astier, H., Avagnina, G. (2024) Haiti violence: Haiti gangs demand PM resign after mass jailbreak. <https://www.bbc.com/news/world-latin-america-68462851> (last accessed on 18 July 2025)

Thomson Reuters (2024). Claudia Sheinbaum wins landslide to become Mexico's 1st woman president <https://www.cbc.ca/news/world/mexico-election-violence-1.7222446> (last accessed on 18 July 2025)

Sellers, C. (2025). FACT CHECK: X Image Showing Netanyahu with Trump and Musk is AI-Generated. <https://checkyourfact.com/2025/02/10/fact-check-image-netanyahu-trump-musk-ai/> (last accessed on 18 July 2025)

Desoubeaux, M. (2025). AFP's Groundbreaking Photo Authentication System: A Milestone in IMATAG's Journey to Secure Digital Truth. <https://www.imatag.com/blog/afps-groundbreaking-photo-authentication-system-a-milestone-in-imatags-journey-to-secure-digital-truth> (last accessed on 18 July 2025)

Teyssou, D., Leung, J.-M., Apostolidis, E., Apostolidis, K., Papadopoulos, S., Zampoglou, M., Papadopoulou, O., Mezaris, V. (2017). The InVID Plug-in: Web Video Verification on the Browser. In: Proceedings of the International Workshop on Multimedia Verification (MuVer 2017) at ACM Multimedia. Mountain View. <https://doi.org/10.1145/3132384.3132387>

AFP (2025). AFP successfully tests a new technology to verify the authenticity of its photos during the US elections. <https://www.afp.com/en/agency/inside-afp/inside-afp/afp-successfully-tests-new-technology-verify-authenticity-its-photos> (last accessed on 18 July 2025)

McCarthy, B., Alonso, N. S., Romero, R. (2025). AI-generated image of Oscar statuette falsely linked to LA wildfires. AFP Fact Check. <http://factcheck.afp.com/doc.afp.com.36VU6WD> (last accessed on 18 July 2025)