Adobe

Introduction to CAI & C2PA and application to TDM and genAI

Leonard Rosenthol
Senior Principal Architect, Content Authenticity & PDF
Chair, C2PA Technical Working Group

Fact Check-Fake Trump Christmas postcard making the rounds on social media

Influencers In Norway Will Legally Have To Disclose Their Photoshopped Images

Fact Check-Photograph does not show explosion at Kabul airport; image was taken at the airport on August 16

TECHNOLOGY

Slick Tom Cruise Deepfakes Signal That Near Flawless Forgeries May Be Here

Lies on Social Media Inflame Israeli-Palestinian Conflict

GOVERNMENT

FBI alert warns of Russian, Chinese use of deepfake content

US Intelligence Agency Photoshops Cover Photo of Diversity Report

How Disinformation Corrodes Democracy

European MPs targeted by deepfake video calls imitating Russian opposition

The U.S. Surgeon General Is Calling COVID-19 Misinformation An 'Urgent Threat'

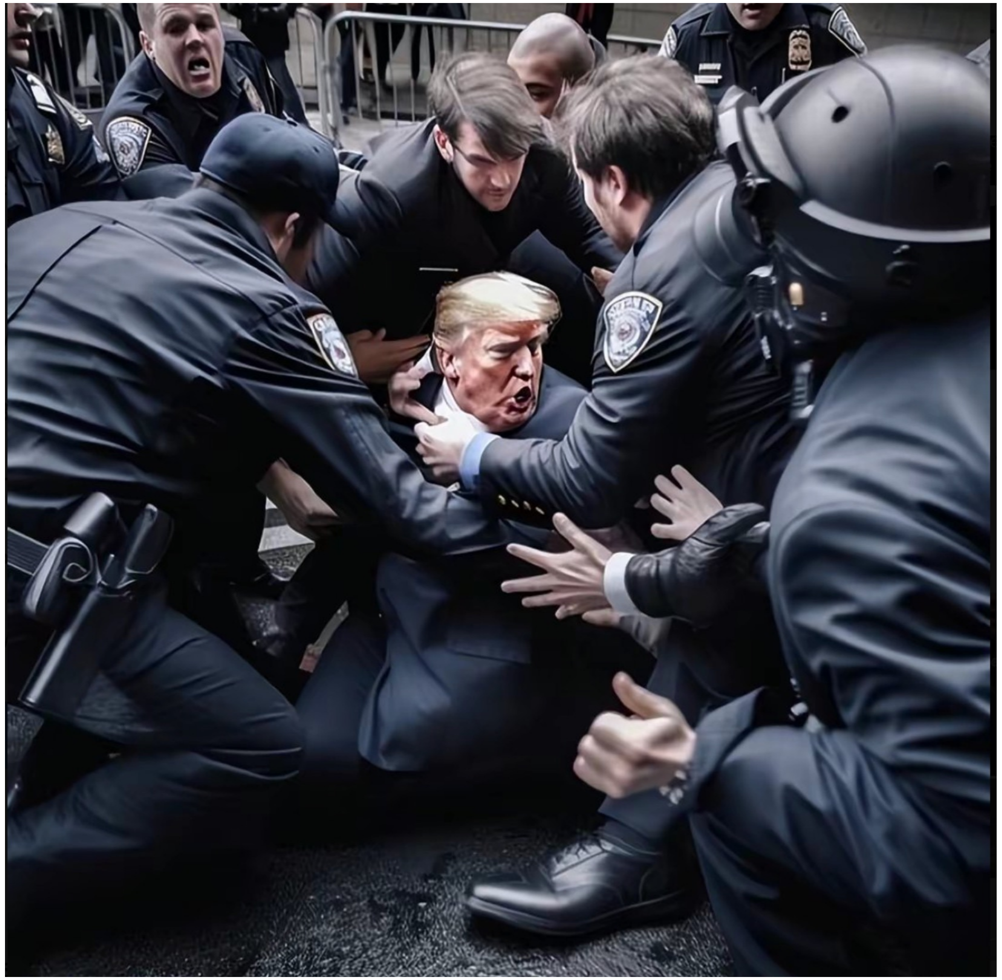
Sweden sets up Psychological Defense Agency to fight fake news, foreign interference

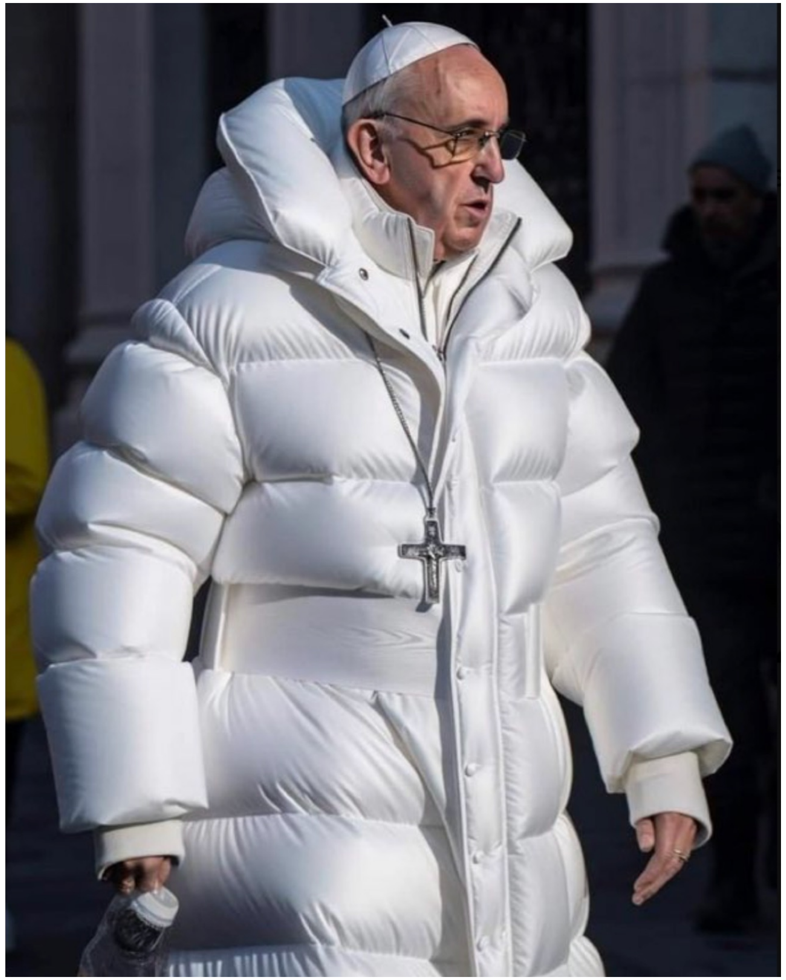
A photo shared on social media does not depict Senator Rand Paul receiving a COVID-19 jab, according to fact-checkers

Deepfake satellite images pose serious military and political challenges

Senate Committee Advances Bill to Create Deepfake Task Force

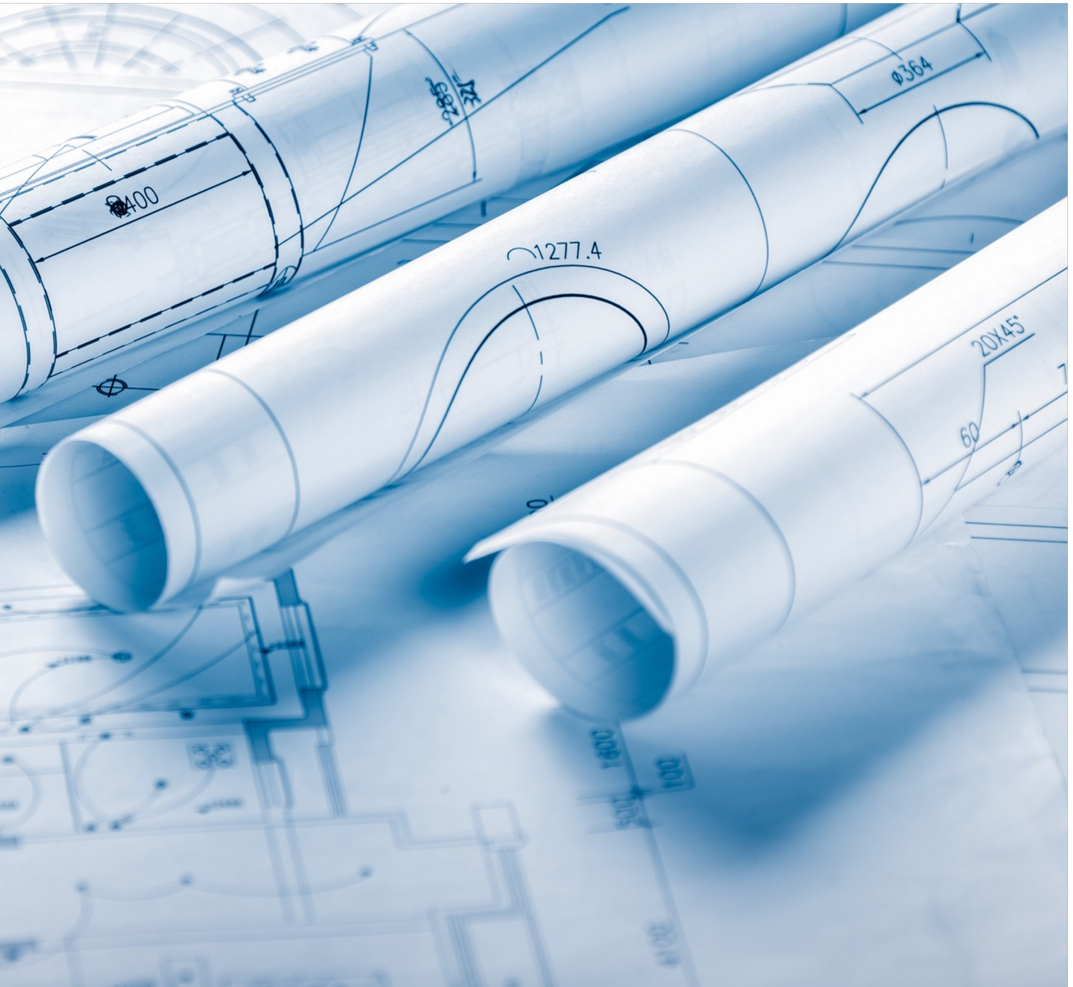
AI Generated Media





C2PA

Content Authenticity Initiative



C2PA Membership

Steering Committee Members

Adobe
BBC
intel
Microsoft
PUBLICIS GROUPE
SONY
truepic

General Members

arm
ATEME Captivate your audience
Canon
digicert
FUTUREWEI Technologies
identity
KEYFACTOR
Nikon
Numbers
RIAA FOR MUSIC
THE SOCIETY LIBRARY
WITNESS SEE IT FILM IT CHANGE IT

Contributor Members

Akamai
BANK OF AMERICA
Blockcast
CBC Radio-Canada
CLink
COMFACT
dalet
dpa
DIGIMARC
EZDRM
fastly
france•tv
ISCC Foundation
JOURNALISM.DESIGN
MEDEX FORENSICS
MELCHER SYSTEM CONSULTING
is·
METAPYXL
Panasonic
PARTNERSHIP ON AI
PIXELSTREAM
Ravnur
SafeCast
sedicii
SequenceKey.com
SERELAY TRUSTED MEDIA CAPTURE
SERUM TRUTH COUNTER FAKE NEWS
SMARTFRAME
steg.ai
THE DAILY EDIT
Tractiv
Truefy
Verisk
Web Commodore
WOPLLI TECHNOLOGIES WORK. PLAY. LEARN. LIVE
WRAPT

Liaisons

- ISO
- IPTC
- ETSI
- PDF Association

Adobe

Content
Authenticity
Initiative

Membership 1500+

AFP

AP

BBC

CBC Radio-Canada

dpa

EFE

epa images

GANNETT

gettyimages®

THE GLOBE AND MAIL

infobae

The New York Times

REUTERS

stern

EL TIEMPO

VII

THE WALL STREET JOURNAL.

The Washington Post

Adobe

arm

camera bits

Canon

Leica

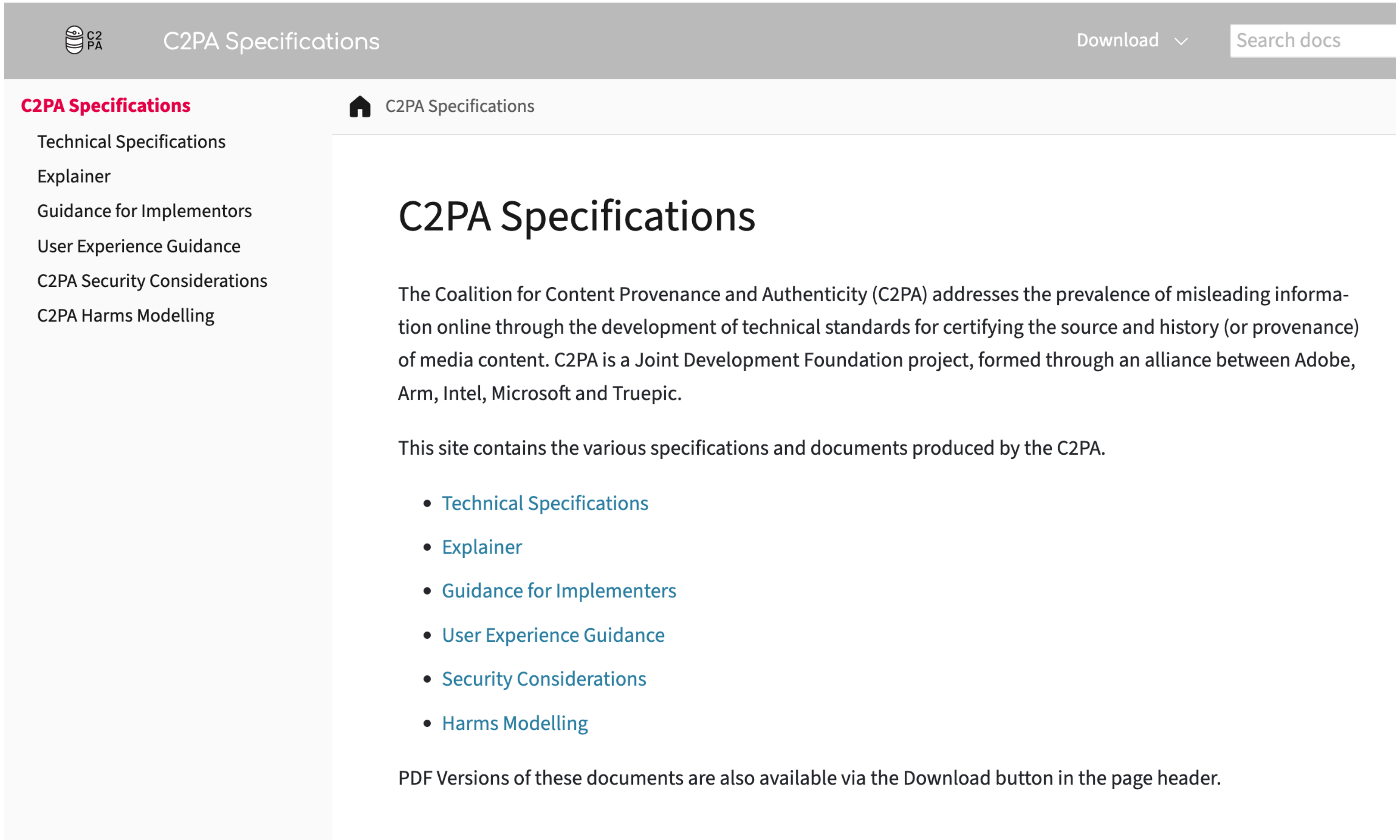
Microsoft

Nikon

Qualcomm

Truepic

Specification (v1.3) Available - https://c2pa.org/specifications/



C2PA Specifications

The Coalition for Content Provenance and Authenticity (C2PA) addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content. C2PA is a Joint Development Foundation project, formed through an alliance between Adobe, Arm, Intel, Microsoft and Truepic.

This site contains the various specifications and documents produced by the C2PA.

- Technical Specifications
- Explainer
- Guidance for Implementers
- User Experience Guidance
- Security Considerations
- Harms Modelling

PDF Versions of these documents are also available via the Download button in the page header.

CAI Open Source

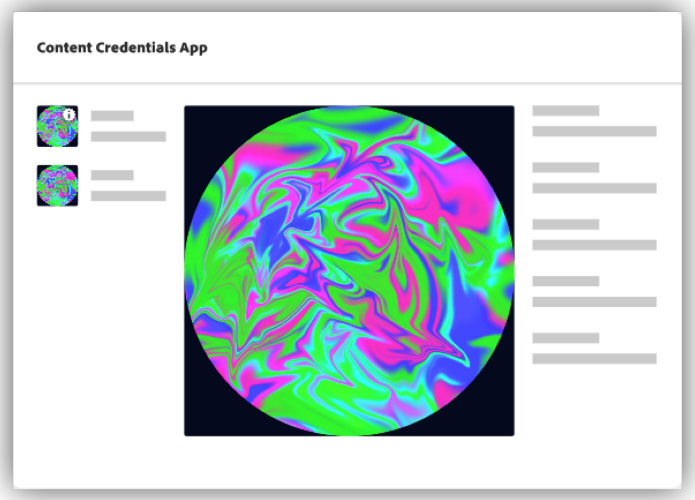
- Used by many companies
  - Adobe
  - Microsoft
  - Truepic
  - Smartframe
  - Nikon
  - Sony
  - ….

Full SDK

Develop custom applications across desktop, mobile, and services that create, verify, and display content credentials via our powerful Rust library.

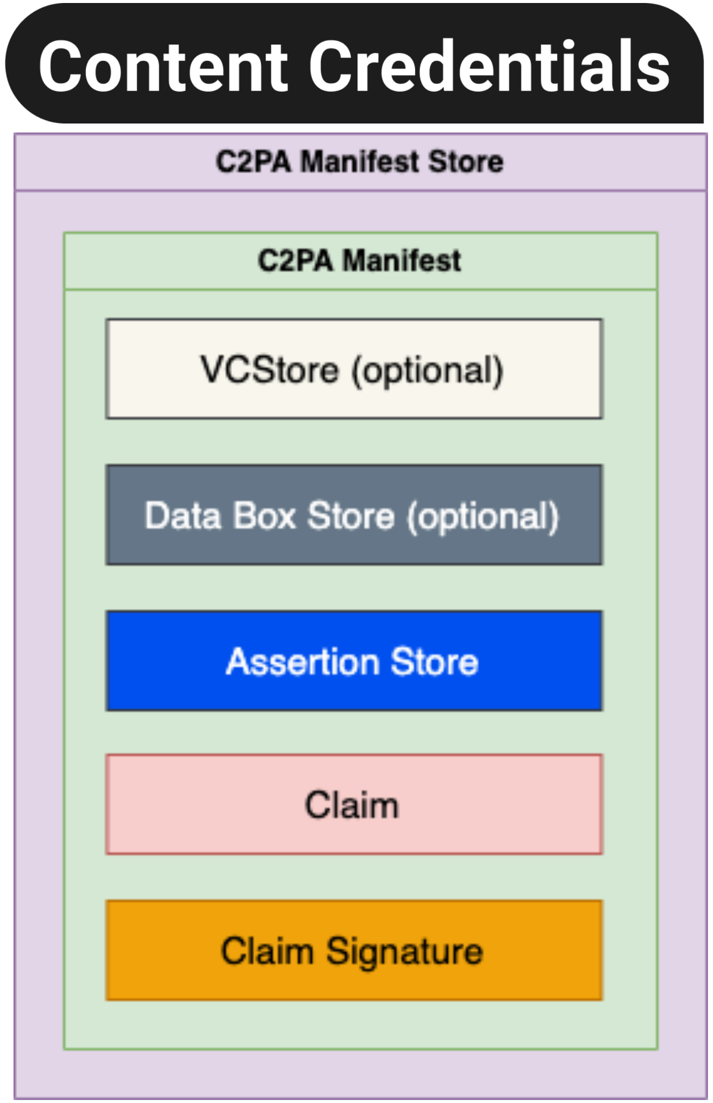
Implementors will use to:

- Display Content Credentials on your site or app
- Link Content Credentials displayed on your site or app to Verify
- Write Content Credentials data into files
- Quickly create and inspect Content Credentials data
- Customize displaying and creating Content Credentials data, with the full power of the specification
- Deploy on Web, mobile, and desktop



Core Components to C2PA

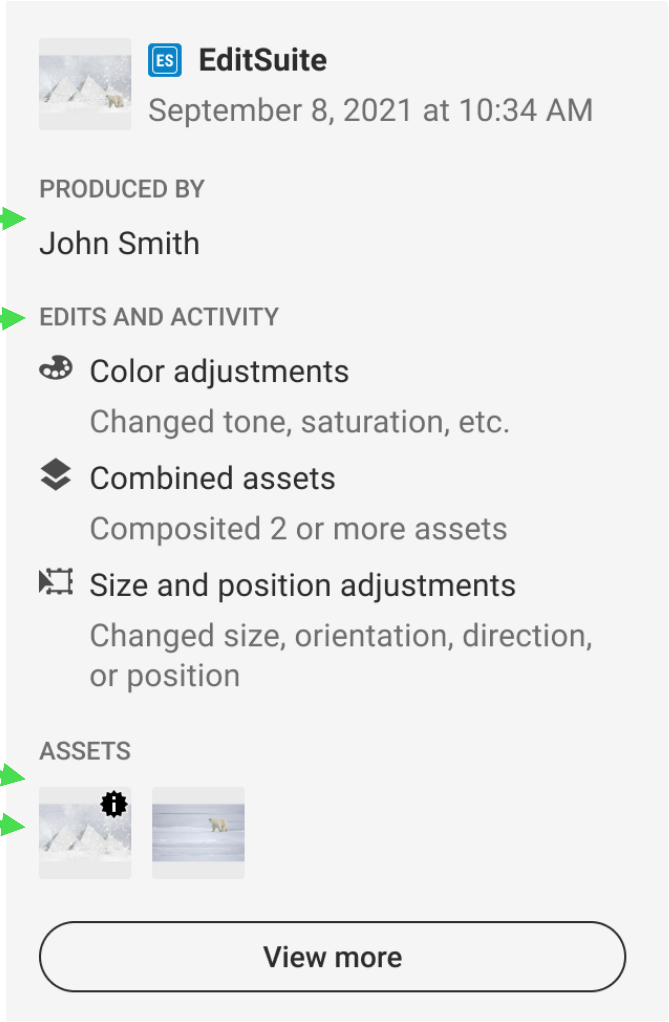
- Assertions
  - A series of statements that cover areas such as asset creation, authorship, edit actions, capture device details, bindings to content and many other subjects.
- Credentials
  - W3C Verifiable Credentials for any actor involved with an assertion.
- Data Boxes
  - Additional information about an assertion such as a GenAI prompt or thumbnail.
- Claim
  - A digitally signed entity, created by a Claim Generator, that lists the assertions being made by the Signer.
- Claim signature
  - The digital signature on the claim using the private key of an actor. This data is a part of the manifest.
- Manifest
  - A verifiable unit into which assertions, claims, credentials and signatures are all bound together. The set of manifests, as stored in the asset's Manifest Store, represent its provenance data.



Some types of Assertions

https://verify.contentauthenticity.org/

- Content Bindings
- Creative Work
- Actions
- Ingredients
- Thumbnails
- Cloud Data
- IPTC, Exif, Schema.org



EditSuite

September 8, 2021 at 10:34 AM

PRODUCED BY

John Smith

EDITS AND ACTIVITY

Color adjustments

Changed tone, saturation, etc.

Combined assets

Composited 2 or more assets

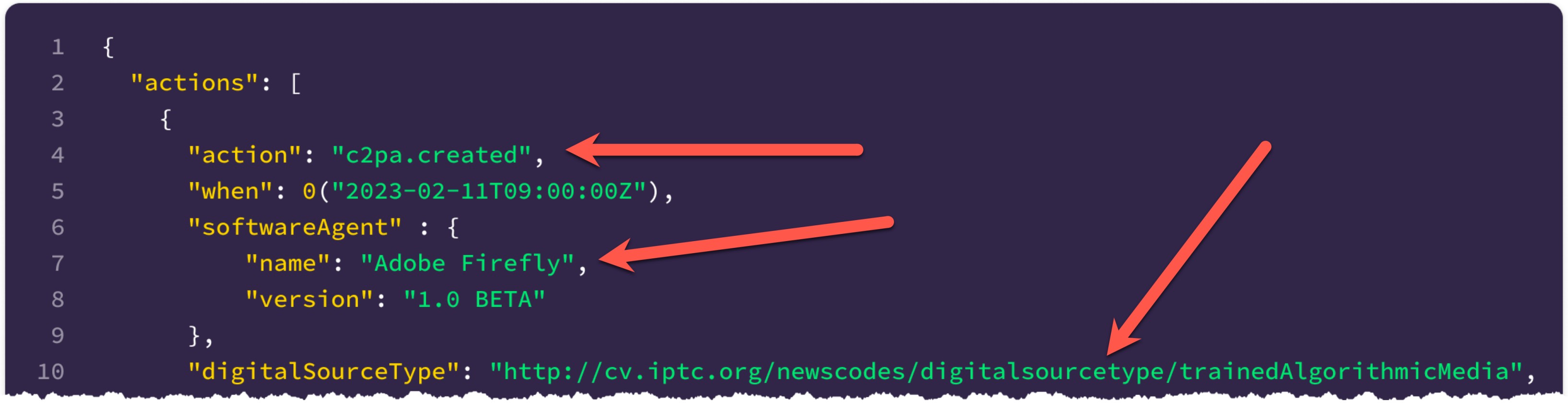
Size and position adjustments

Changed size, orientation, direction, or position

ASSETS

Identifying Assets

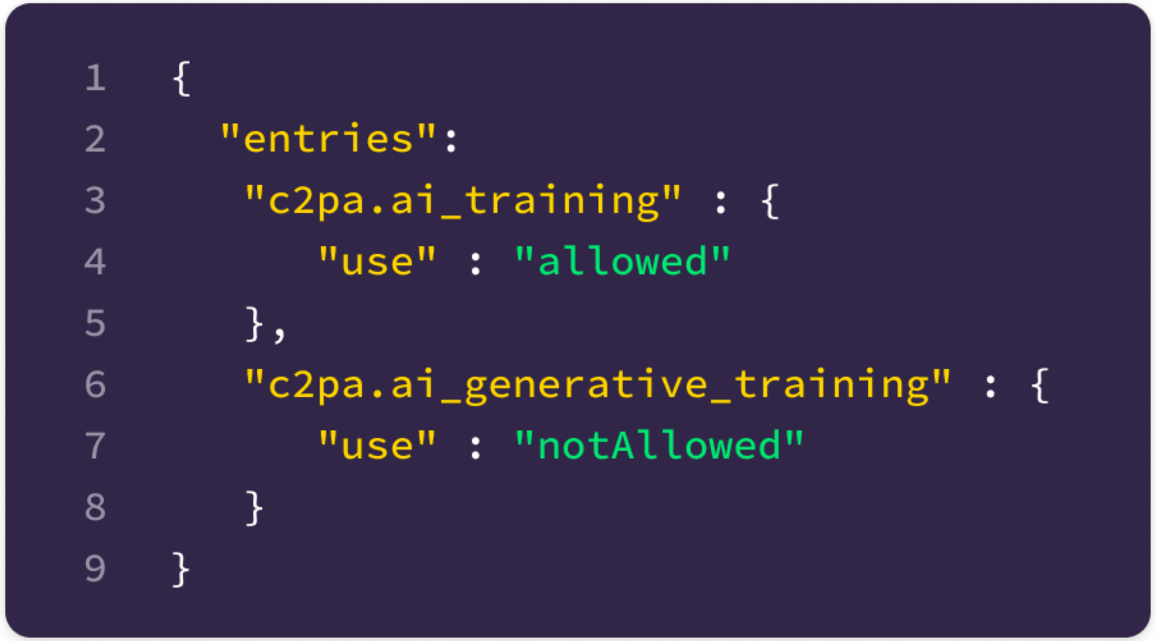
- Action Assertion
    - Usually `c2pa.created` but could be `c2pa.edited` or `c2pa.filtered` or …
    - digitalSourceType
        - Defined by IPTC
        - Also chosen by Google for identifying GenAI



```
{
  "actions": [
    {
      "action": "c2pa.created",
      "when": 0("2023-02-11T09:00:00Z"),
      "softwareAgent" : {
          "name": "Adobe Firefly",
          "version": "1.0 BETA"
      },
      "digitalSourceType": "http://cv.iptc.org/newscodes/digitalsourcetype/trainedAlgorithmicMedia",
```

Do Not Train

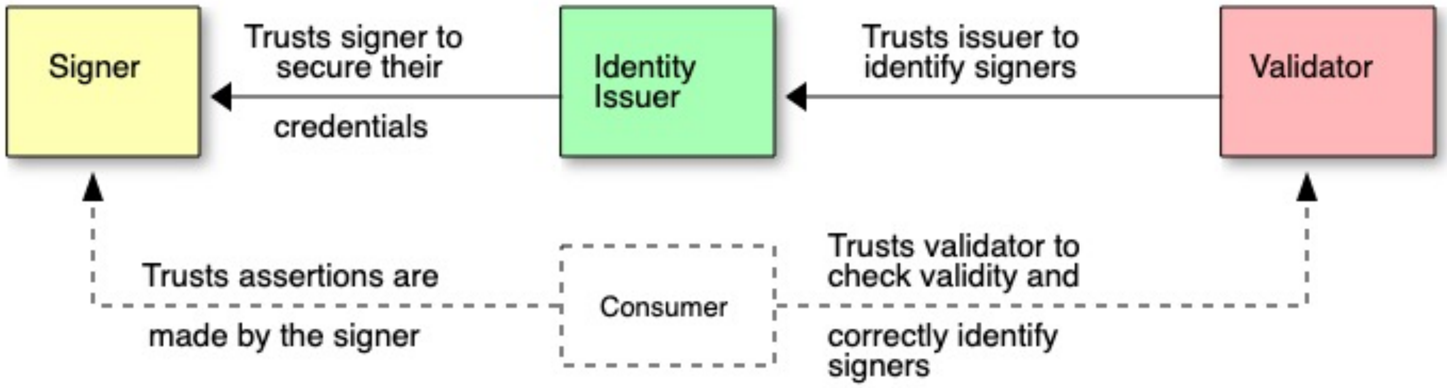
- Four categories
  - ai_training
  - ai_generative_training
  - ai_inference
  - data_mining
- Possible values
  - Allowed
  - Not allowed
  - Constrained



```
{
  "entries":
   "c2pa.ai_training" : {
      "use" : "allowed"
   },
   "c2pa.ai_generative_training" : {
      "use" : "notAllowed"
   }
}
```

Establishing a Trust Model

- Modelled on the same approach to trust as PDF and the Web
  - X.509 Certificates
  - Certificate Authorities
  - Trust Lists


Signer
Trusts signer to secure their credentials
Identity Issuer
Trusts issuer to identify signers
Validator
Trusts assertions are made by the signer
Consumer
Trusts validator to check validity and correctly identify signers

Manifests can be embedded or referenced

- C2PA Manifests can be embedded into
  - **Images** (JPEG, PNG, GIF, WebP, AVIF, HEIC/HEIF, TIFF, DNG, SVG)
  - **Videos** (MP4, MOV, AVI, BMFF)
  - **Audio** (FLAC, MP3, WAV, BWF)
  - **Documents** (PDF, EPUB*)
- They can be stored separately in file systems, the cloud, DLT/Blockchains & referenced by URL, HTTP headers, file system paths and more.

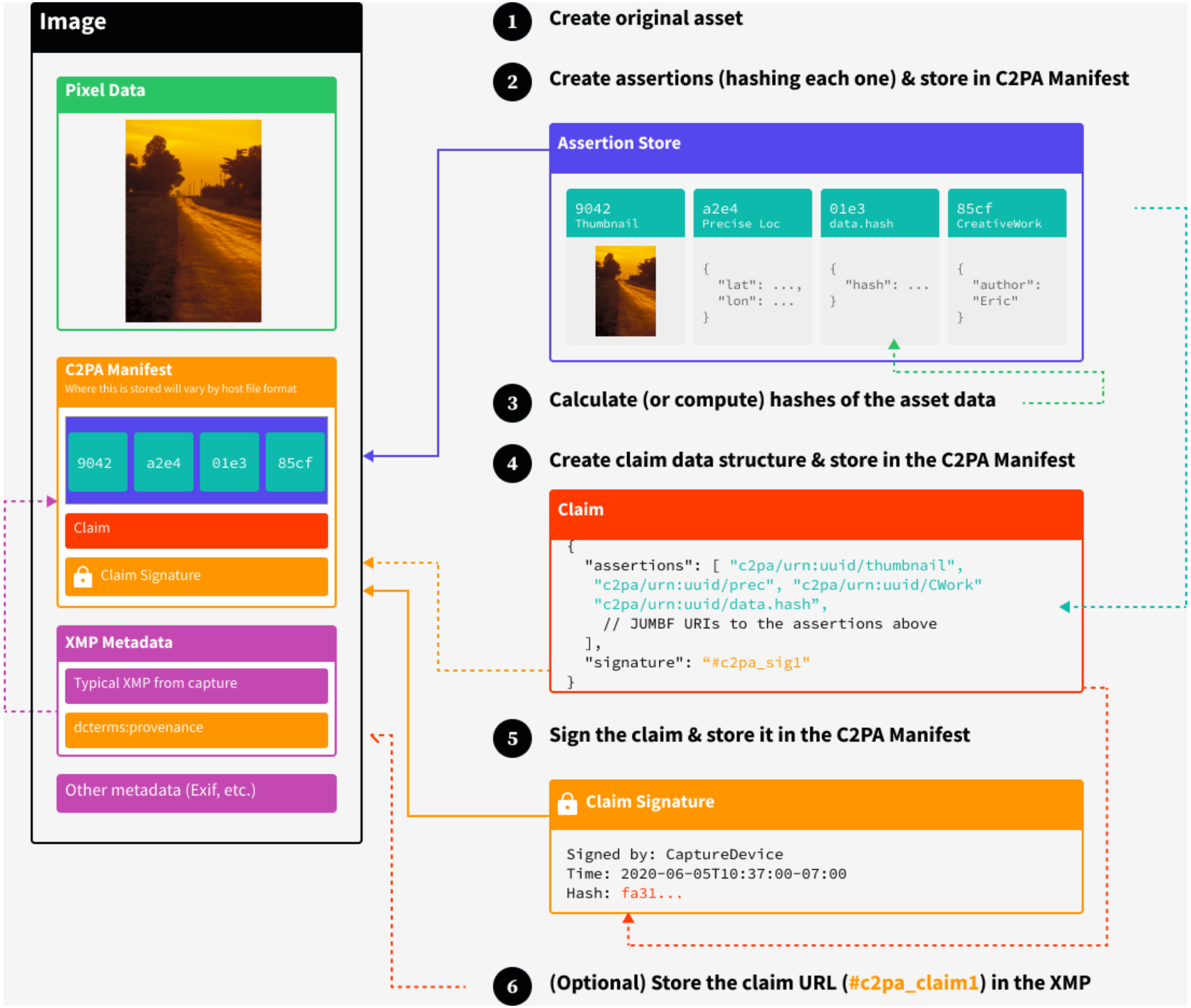
Legislative work

- A number of countries are looking to adopt legislation about
  - Identification of assets created/modified by generative AI
  - Identification of rights concerning use of an asset for AI training
  - US, UK, EU, China, AU, NZ
- C2PA addresses both needs
  - is already out in the field (Adobe, Microsoft, Shutterstock, etc)
  - Supports many file formats

Thank you!

Questions

Adobe



1 **Create original asset**

2 **Create assertions (hashing each one) & store in C2PA Manifest**

Assertion Store

```
9042
Thumbnail
```

```
a2e4
Precise Loc
{
  "lat": ...,
  "lon": ...
}
```

```
01e3
data.hash
{
  "hash": ...
}
```

```
85cf
CreativeWork
{
  "author":
  "Eric"
}
```

3 **Calculate (or compute) hashes of the asset data**

4 **Create claim data structure & store in the C2PA Manifest**

Claim

```
{
  "assertions": [ "c2pa/urn:uuid/thumbnail",
   "c2pa/urn:uuid/prec", "c2pa/urn:uuid/CWork"
   "c2pa/urn:uuid/data.hash”,
    // JUMBF URIs to the assertions above
  ],
  "signature": “#c2pa_sig1"
}
```

5 **Sign the claim & store it in the C2PA Manifest**

Claim Signature

```
Signed by: CaptureDevice
Time: 2020-06-05T10:37:00-07:00
Hash: fa31...
```

6 **(Optional) Store the claim URL (#c2pa_claim1) in the XMP**