Content Authenticity Initiative

Eric Scouten · Identity Standards Architect · Adobe

Content Authenticity 101

Content Authenticity: Who? What? Why?

C2PA data model overview

CAWG identity assertion overview

Why content provenance?

Why content provenance?

The New York Times — *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.*

Forbes — The Battle Against Fake News Enters The Age Of Deepfakes

CNN BUSINESS — Puffer coat Pope. Musk on a date with GM CEO. Fake AI 'news' images are fooling social media users

The Guardian — Elections in UK and US at risk from AI-driven disinformation, say experts

TIME — From Scams to Music, AI Voice Cloning Is on the Rise

CBC NEWS — With rise of AI-generated images, distinguishing real from fake is about to get a lot harder

Why content provenance?

We can't uninvent this.

What we *can* do is …

provide a mechanism for well-intentioned content creators to stand apart from misinformation.

Our goals

Allow **content creators** to make tamper-evident, digitally-signed statements about what they've created.

Allow **content consumers** to evaluate those statements and use them to make trust decisions.
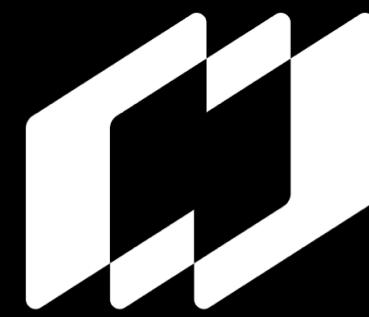
Our non-goals

Content Authenticity is **not:**

- fact-checking

- fake image detection

- politically opinionated

Demo: How it works in Photoshop
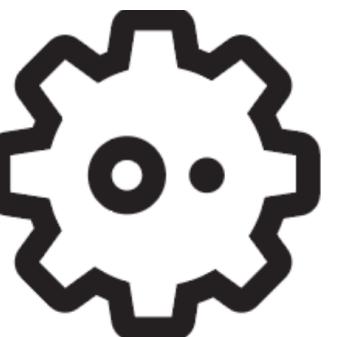
Who's who?
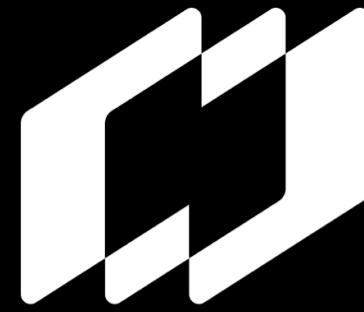
Content Authenticity Initiative

C2PA Coalition for Content Provenance and Authenticity

Creator Assertions Working Group

Who's who?

Content Authenticity Initiative
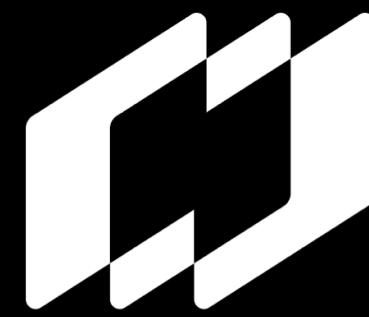
Outreach · Advocacy · Open Source
also name of Adobe's team

C2 PA — **Coalition for Content Provenance and Authenticity**

Technical Standards: **What / How**

Creator Assertions Working Group
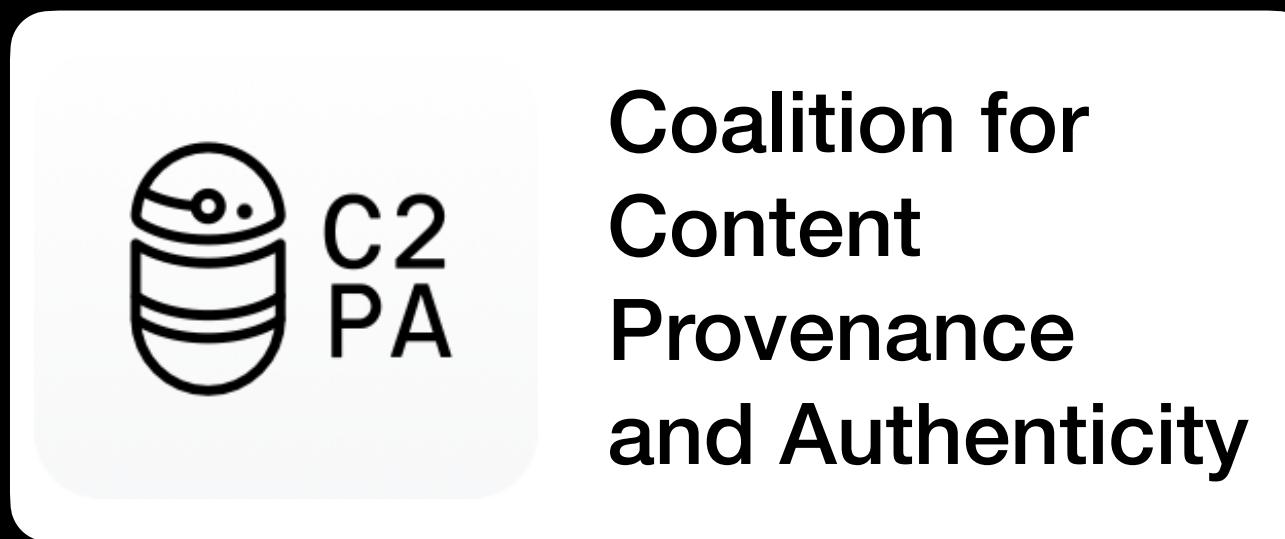
Technical Standards: **Who**

Who's who?
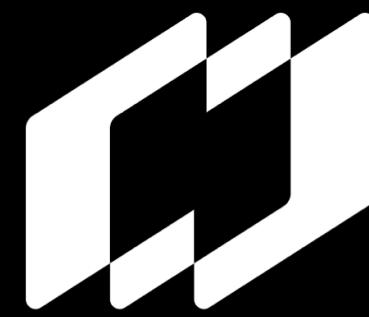
Logo	Name	URL
	Content Authenticity Initiative	contentauthenticity.org
C2 PA	Coalition for Content Provenance and Authenticity	c2pa.org
	Creator Assertions Working Group	creator-assertions.github.io

Who's in?

~3000 members, including ...
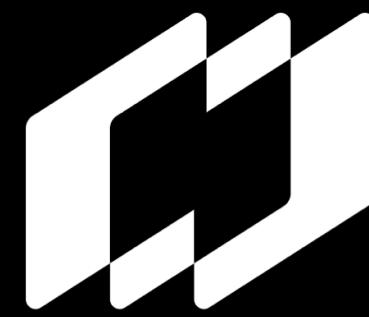
Who's in?

Major camera manufacturers

Who's in?

~100 members, including …

C2PA data model

C2PA data model
Overview

An **asset** is any piece of digital media that we wish to describe.

asset

Currently supported asset types include:

photo

video

audio

documents

fonts
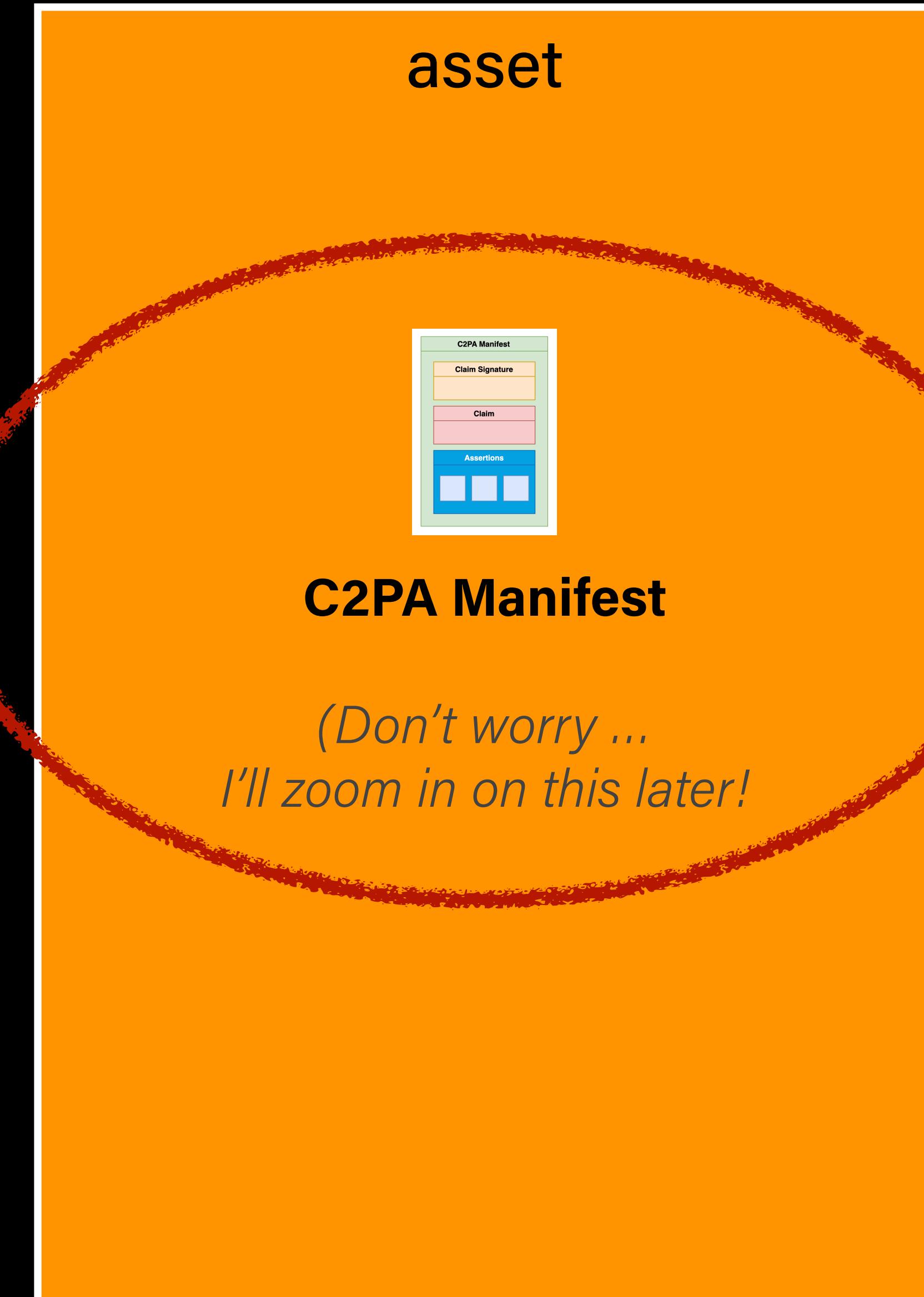
C2PA data model
Overview

An **asset** is any piece of digital media that we wish to describe.

It is described by a **C2PA Manifest**.

asset

C2PA Manifest

(Don't worry ...
I'll zoom in on this later!
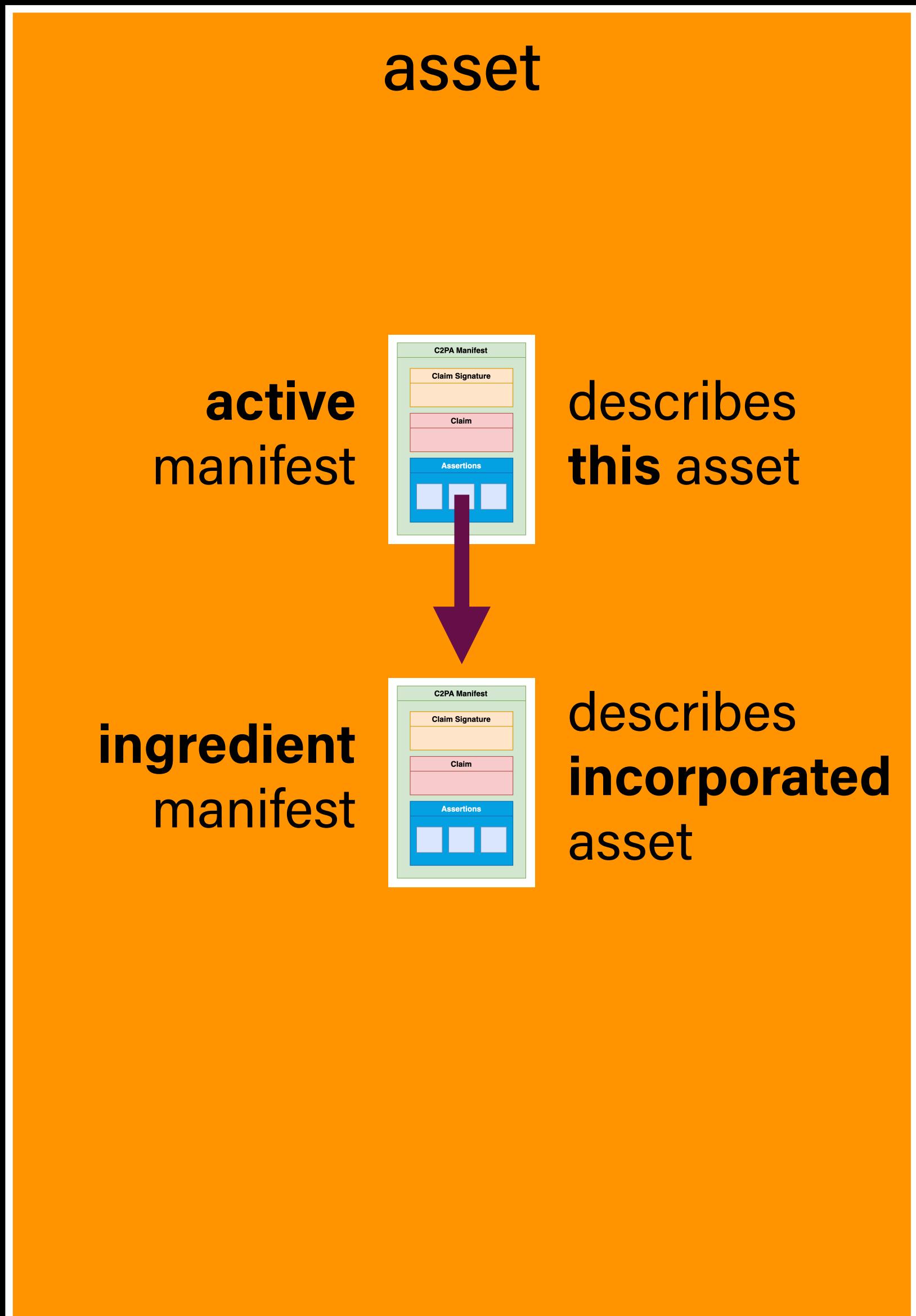
C2PA data model
Overview

An **asset** is any piece of digital media that we wish to describe.

It is described by a **C2PA Manifest**. Each asset in C2PA has an *active manifest* which describes the current asset.

That C2PA Manifest may refer to *ingredient manifests* when earlier content is incorporated.

C2PA data model
Overview
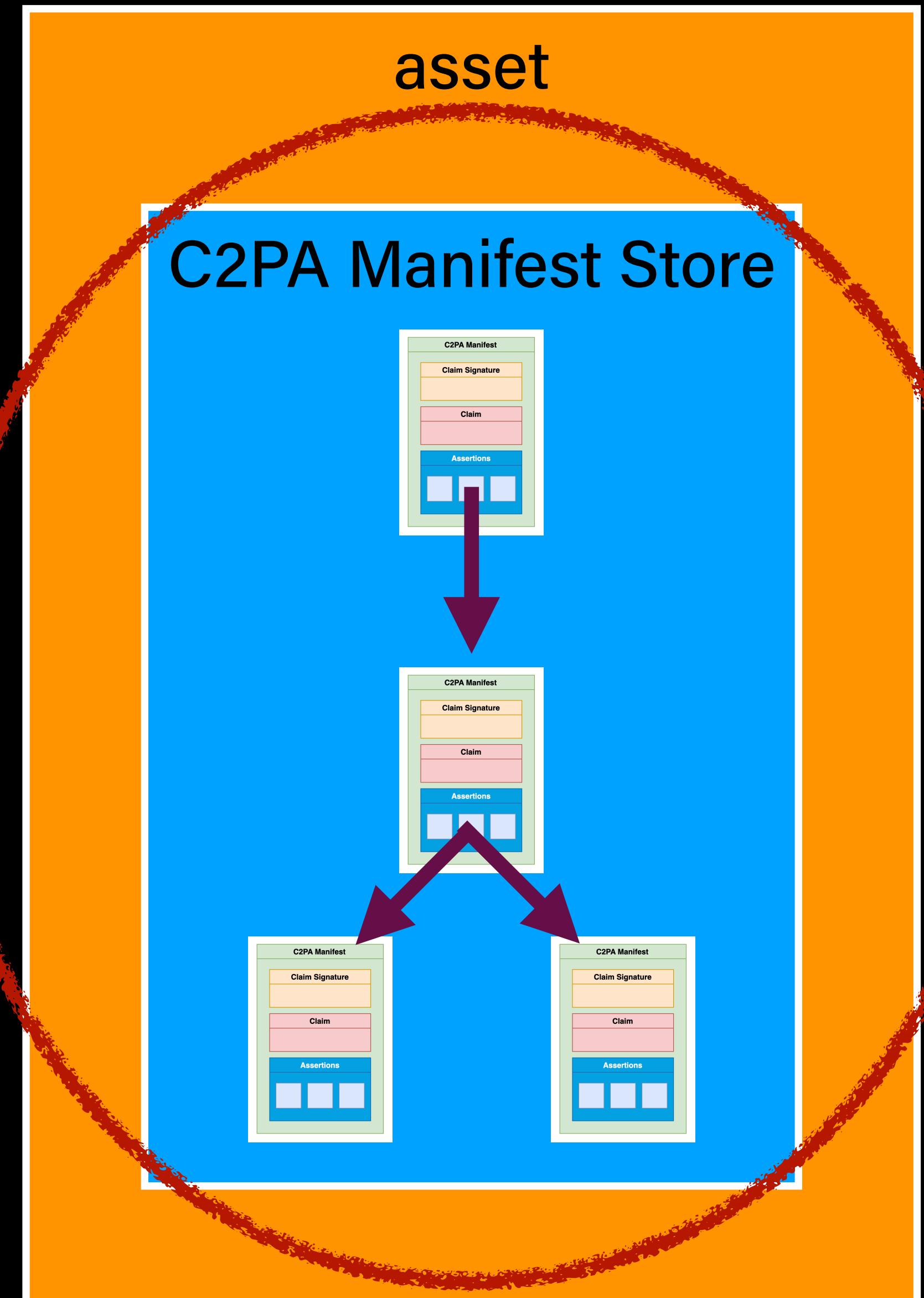
An **asset** is any piece of digital media that we wish to describe.

It is described by a **C2PA Manifest**. Each asset in C2PA has an *active manifest* which describes the current asset.

That C2PA Manifest may refer to *ingredient manifests* when earlier content is incorporated.

asset

C2PA data model
Overview

An **asset** is any piece of digital media that we wish to describe.

It is described by a **C2PA Manifest**. Each asset in C2PA has an *active manifest* which describes the current asset.

That C2PA Manifest may refer to *ingredient manifests* when earlier content is incorporated.

The collection of C2PA Manifests is referred to as a **C2PA Manifest Store.**

asset

C2PA Manifest Store
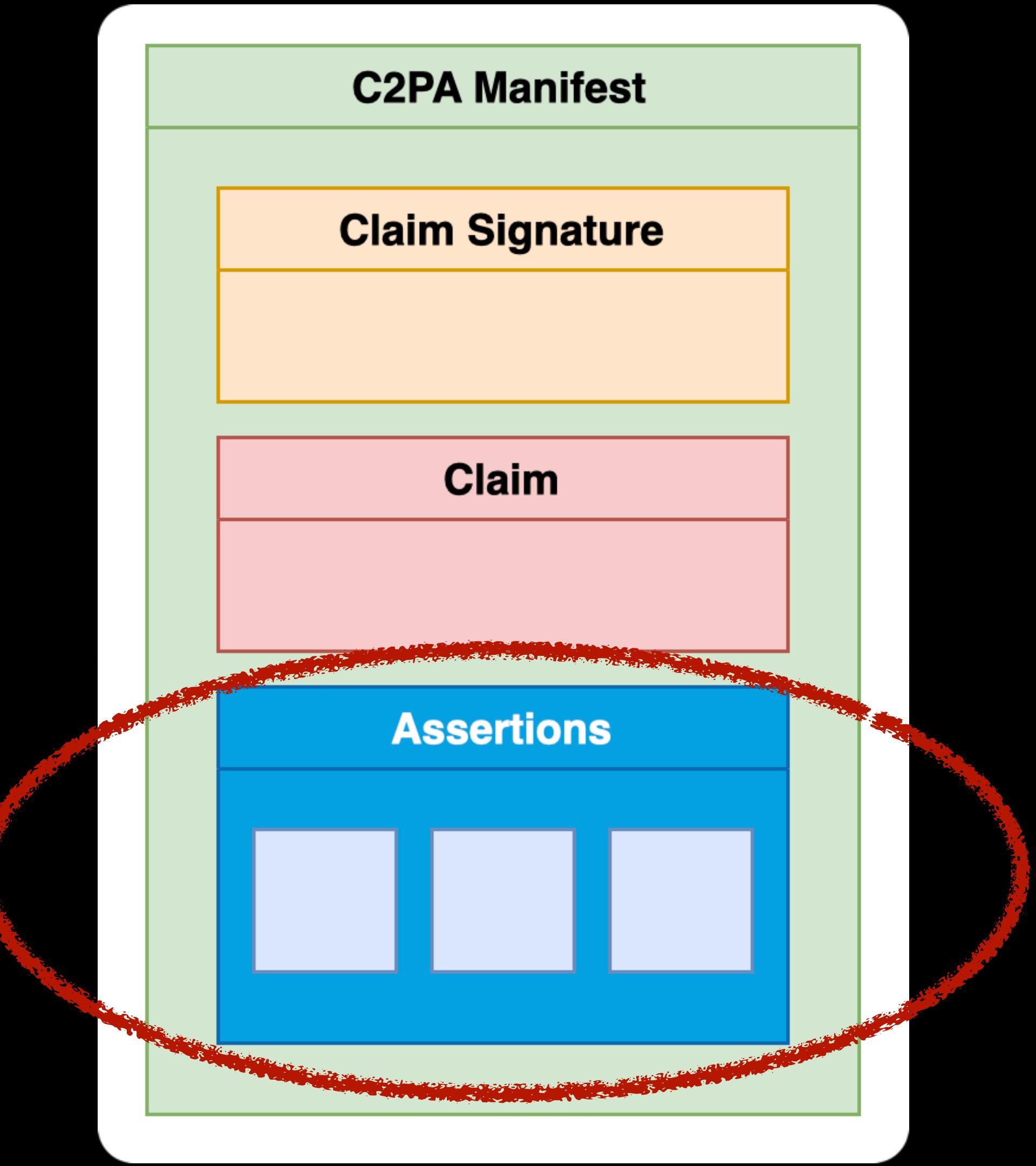
C2PA data model
Assertions

Assertions are opt-in statements that cover areas such as:

- hard binding to asset's binary content *(required – provides tamper evidence)*

- capture device details

- edit actions

- thumbnail of the content

- other content (ingredients) that were incorporated into this content
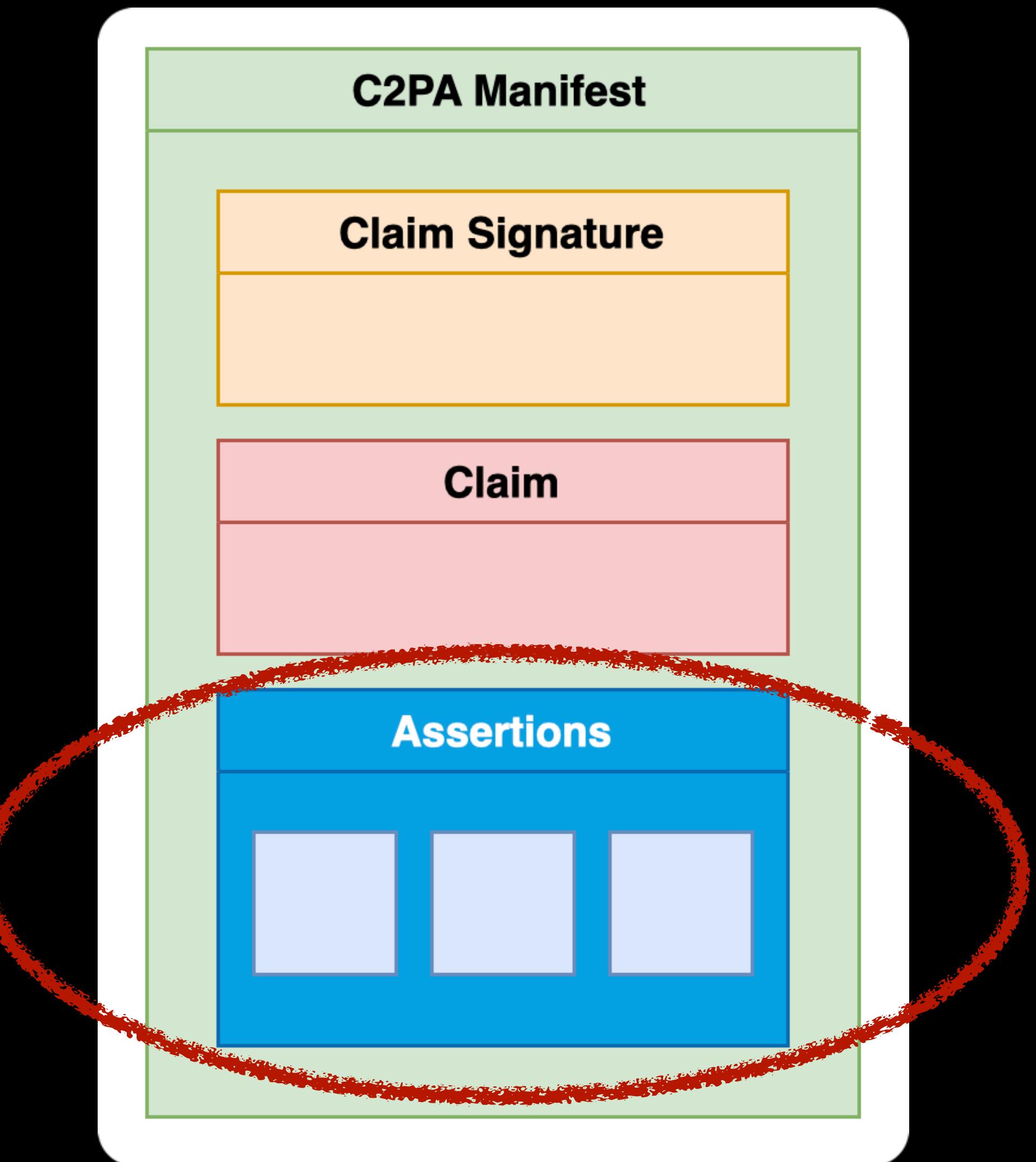
C2PA data model
Assertions, part 2

Assertions can be **redacted** (overwritten by zeros) by future manifest producers.

Assertions are typically serialized in CBOR.

Assertions defined by C2PA *only* describe machine-attestable content. *(New in C2PA 2.0.)*

Assertions are labeled and can be defined by other organizations.

C2PA Manifest

Claim Signature
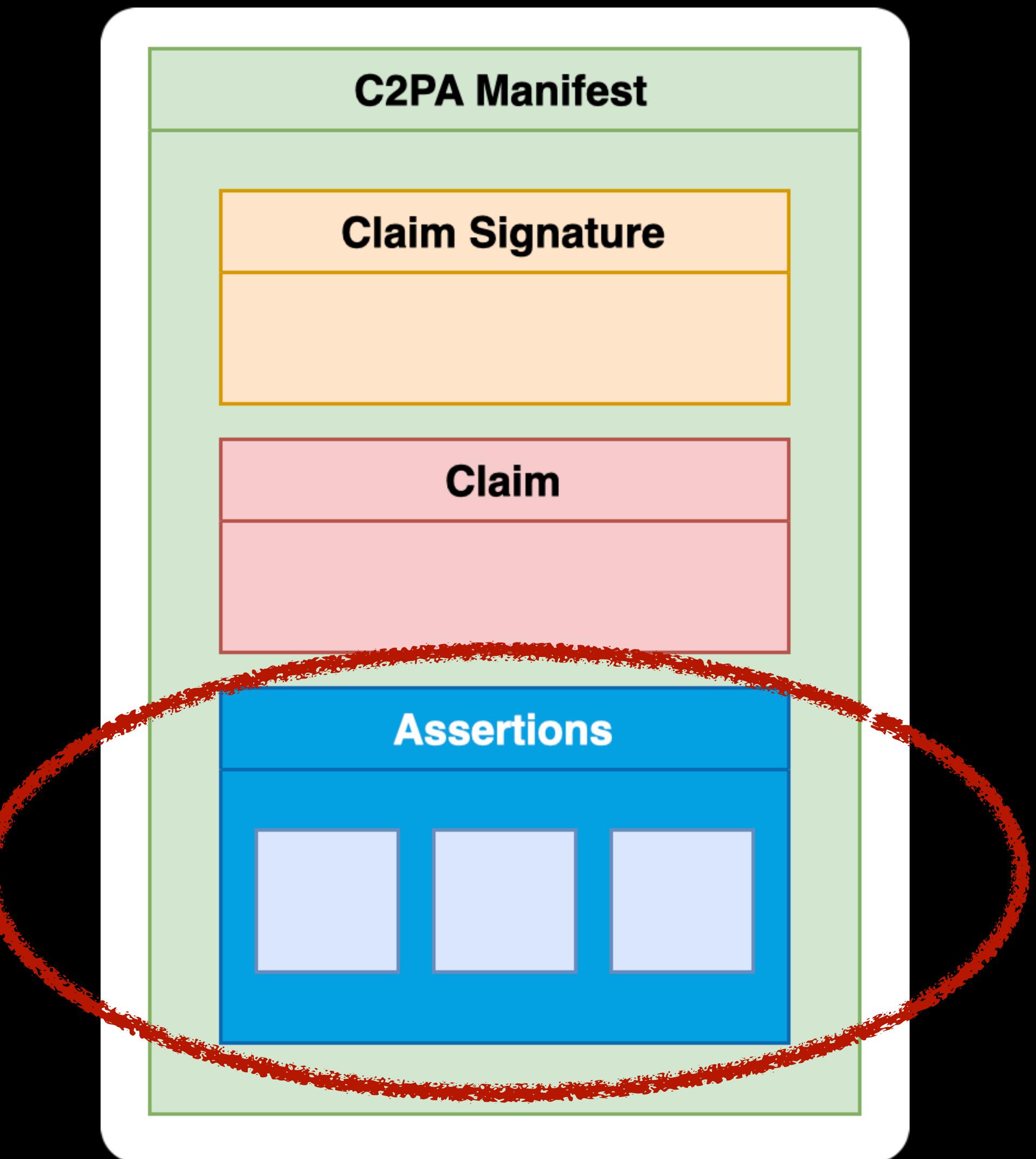
Claim

Assertions

C2PA data model
Assertions, part 3

Some examples of non-C2PA assertions:

- identity of the content creator(s) *(see later CAWG discussion)*

- training and data mining intent
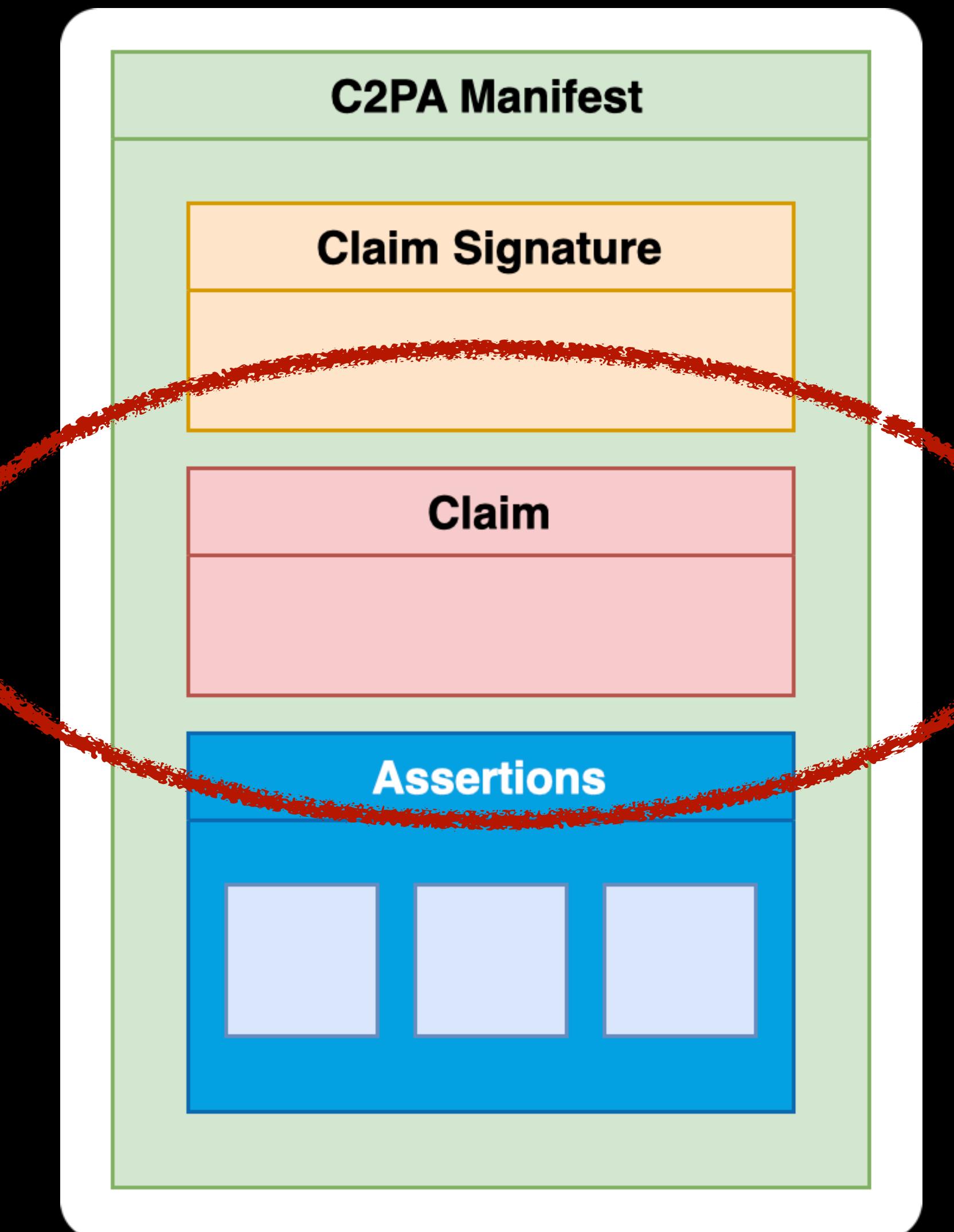
C2PA data model
Claim

Every C2PA Manifest has exactly one **claim,** which contains:

- a list of its assertions
  (via hashed JUMBF URI)

- information about who created the
  claim (typically tool vendor)

- assertions from ingredients that were
  redacted
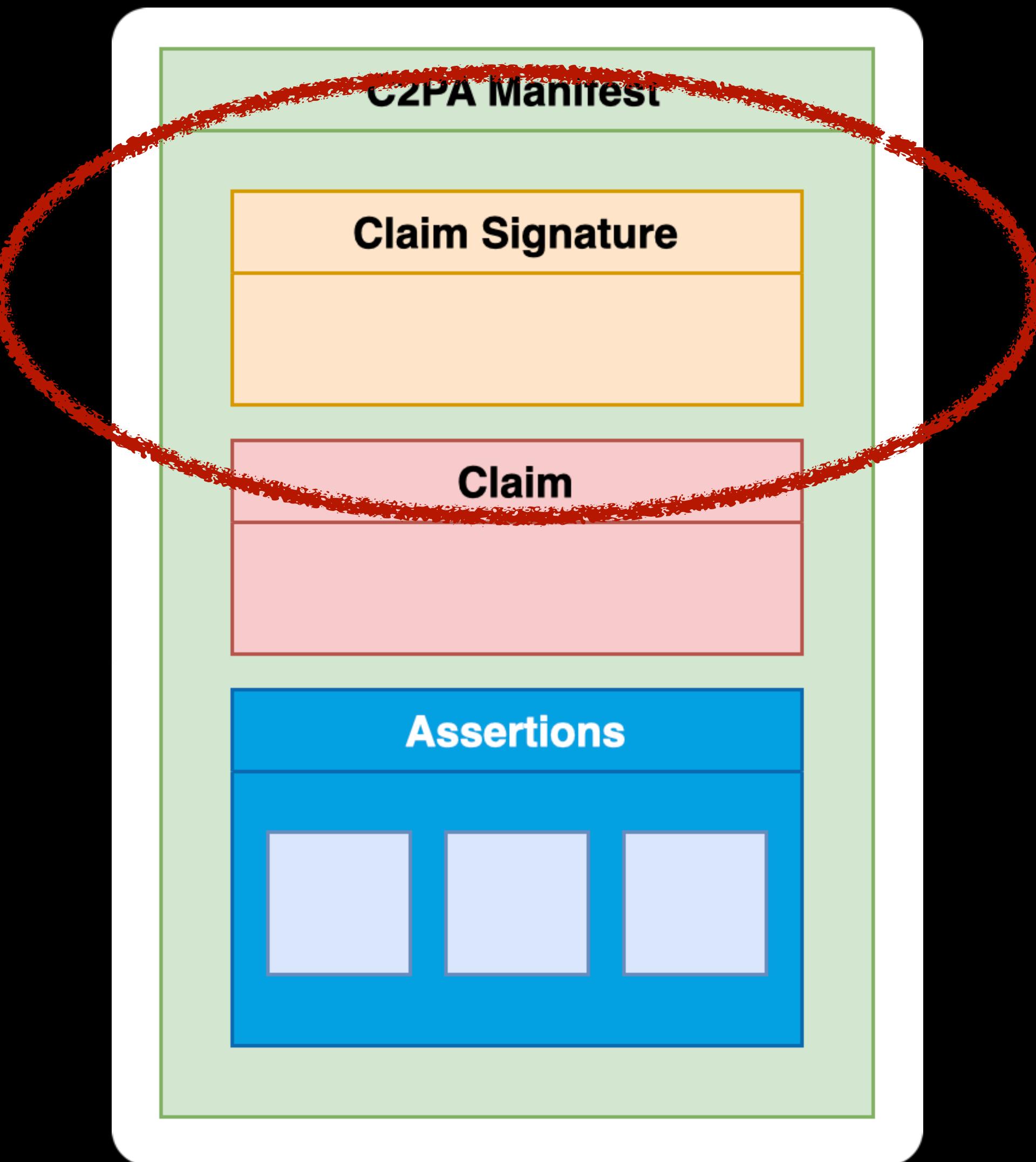
C2PA data model
Claim signature

A **claim signature** is a COSE signature that binds the claim data structure to an X.509 certificate holder.

The X.509 certificate typically identifies the *implementation* of C2PA (hardware or software), **not** the content author.
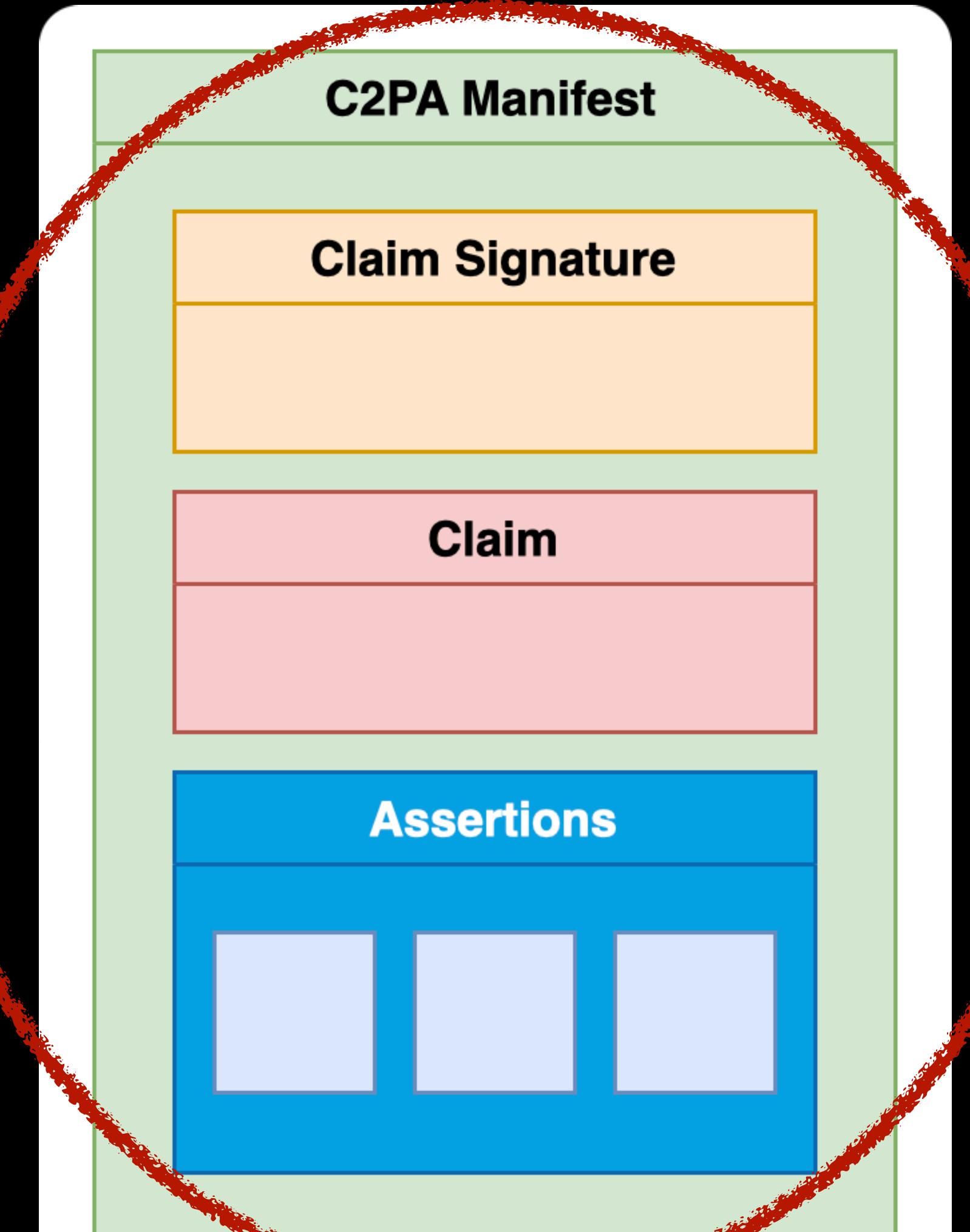
C2PA data model
C2PA Manifest

A **C2PA Manifest** is a JUMBF data structure which contains the claim signature, claim, and assertions.

C2PA Manifest

Claim Signature

Claim

Assertions

C2PA data model
Sample user experience
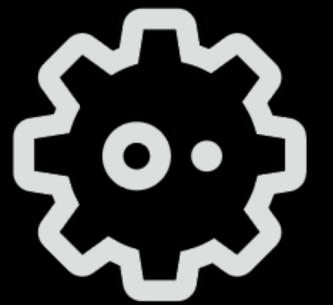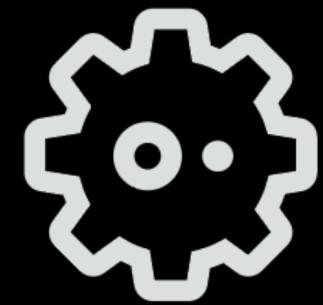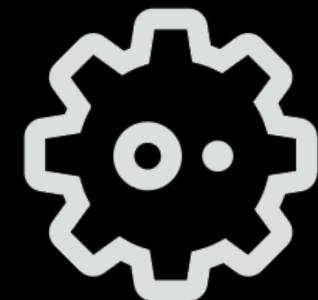
CAWG identity assertion

Identity assertion: Why?

"This specification aims to allow an actor to document their relationship to a C2PA asset produced by them or on their behalf independently from the C2PA claim generator, and to allow consumers of a C2PA asset to independently verify that the received asset was in fact produced by the named actor and has not been tampered with."

– Identity assertion specification "Design goals" section
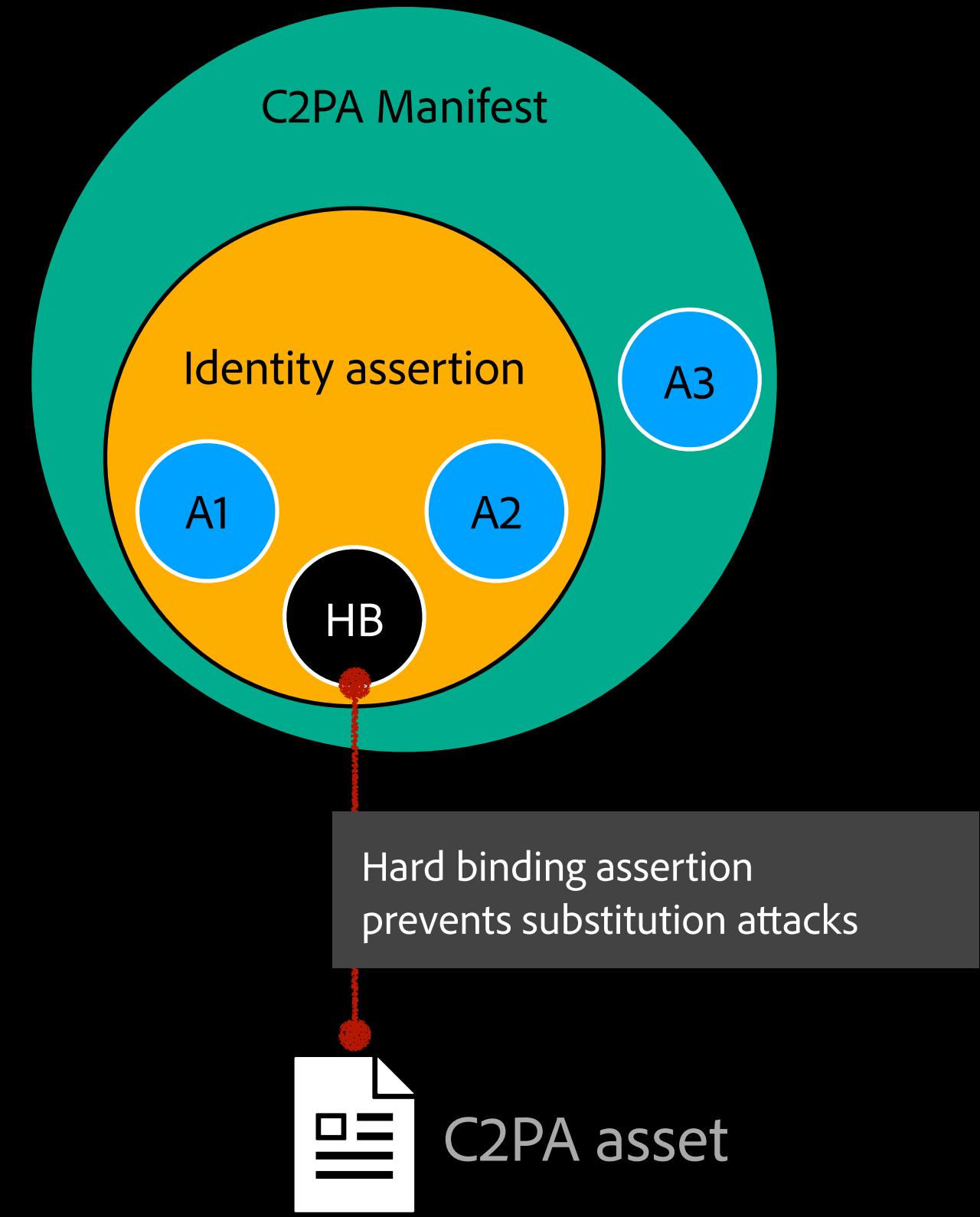
Identity assertion
## Overview
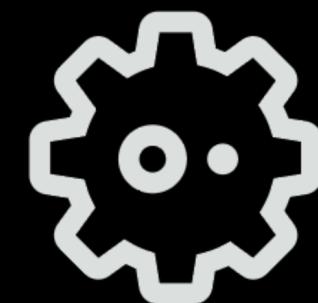
New and separate trust signal over and above the C2PA claim generator signal.

Typically meant to indicate **subject's authorization or active participation** in production of the asset.


C2PA Manifest

Identity assertion

A3

A1 A2

HB

Hard binding assertion prevents substitution attacks

C2PA asset

A = any assertion other than identity or hard binding

HB = hard binding assertion
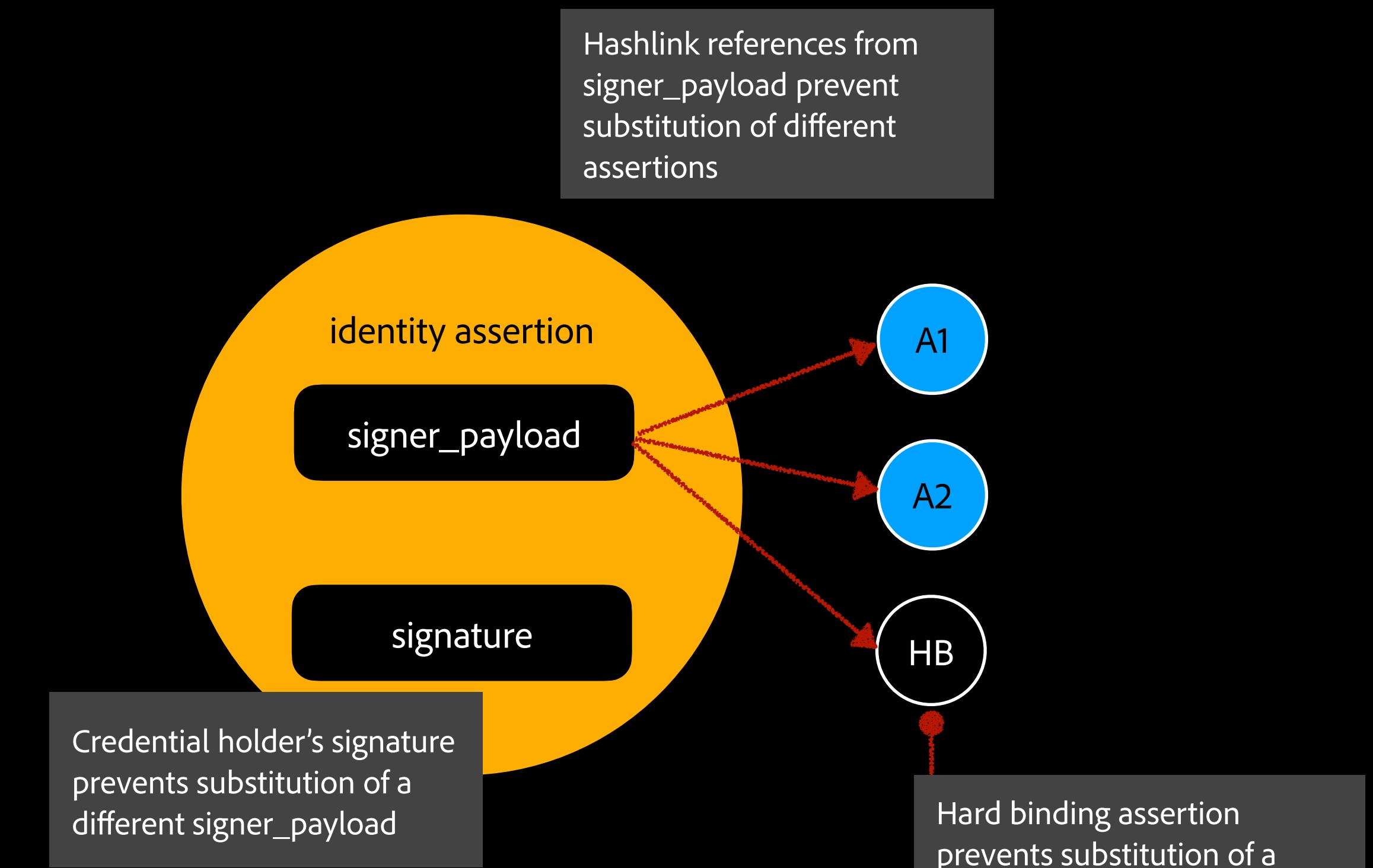
Identity assertion
Overview

Identity assertion allows a credential holder to sign a data structure we call signer_payload, which contains:

- Tamper-evident references to one or more other assertions in the same C2PA Manifest (including hard-binding assertion)

- Role of credential subject with regard to the content

Hashlink references from signer_payload prevent substitution of different assertions

identity assertion

signer_payload

signature

A1

A2

HB

Credential holder's signature prevents substitution of a different signer_payload

Hard binding assertion prevents substitution of a different C2PA asset
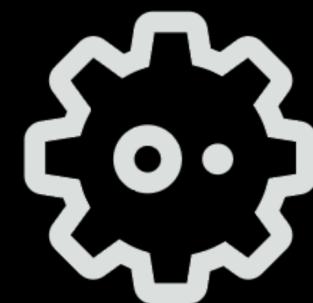
C2PA asset

A*n* = any assertion other than identity or hard binding

HB = hard binding assertion
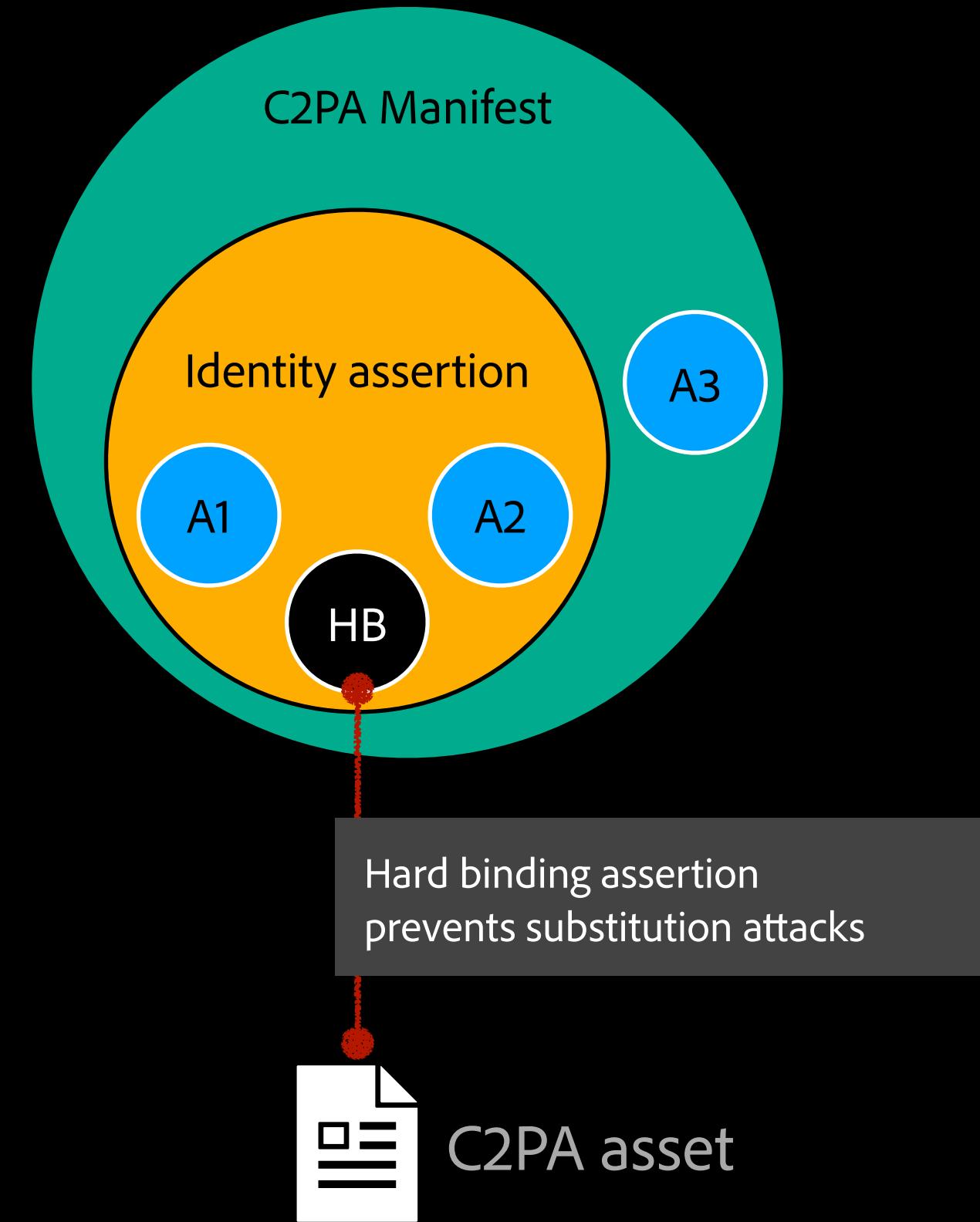
Identity assertion
Overview

Credential and signature can be:
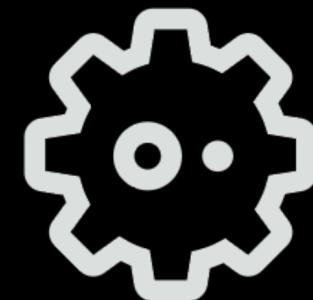
- X.509 certificate ➤ COSE signature

- W3C verifiable credential ➤ another VC

- *Extensible ... others may be added by CAWG in future or by others*

C2PA Manifest

Identity assertion

A3

A1   A2

HB

Hard binding assertion
prevents substitution attacks

C2PA asset

A = any assertion other than identity or hard binding

HB = hard binding assertion
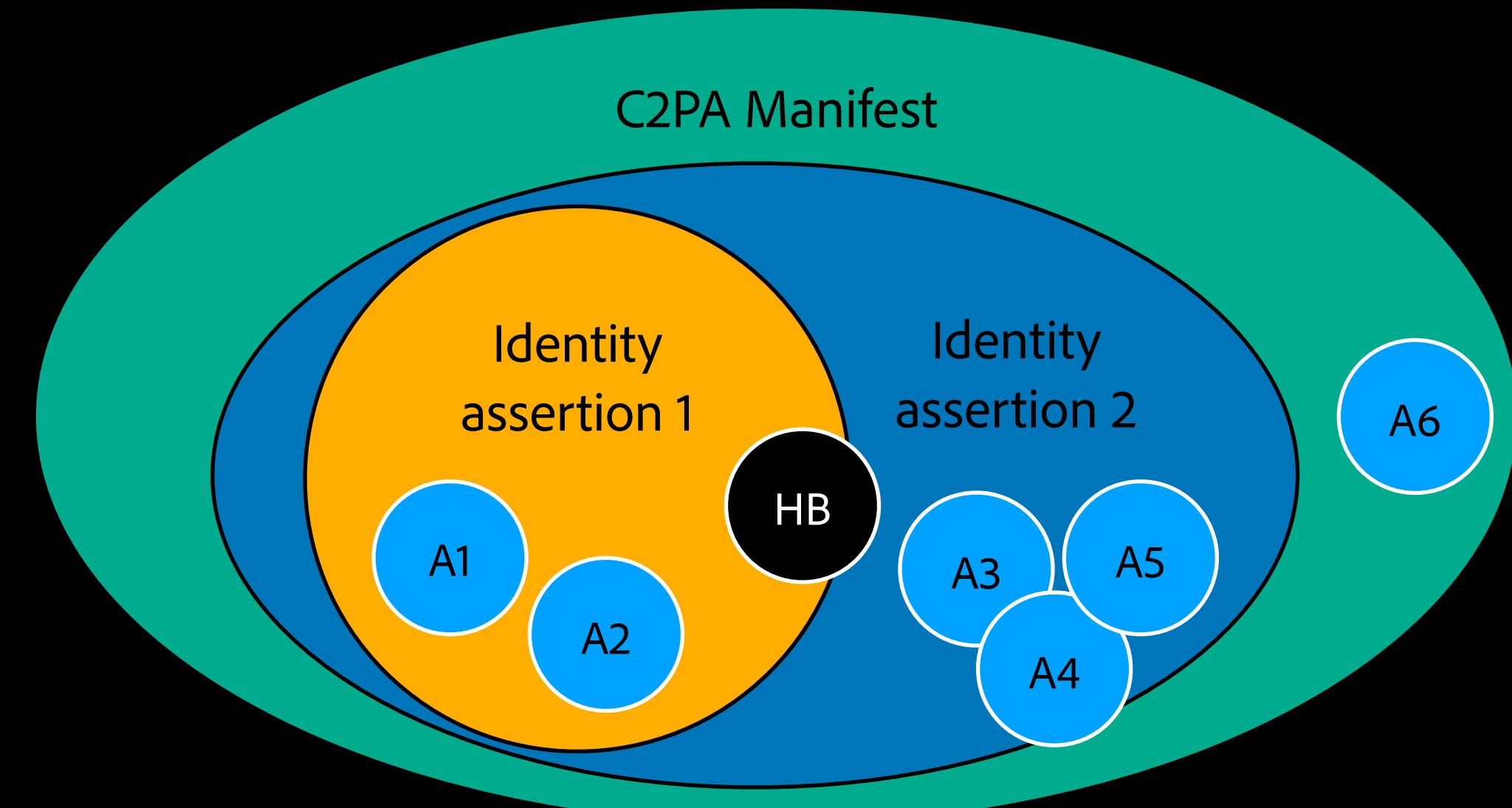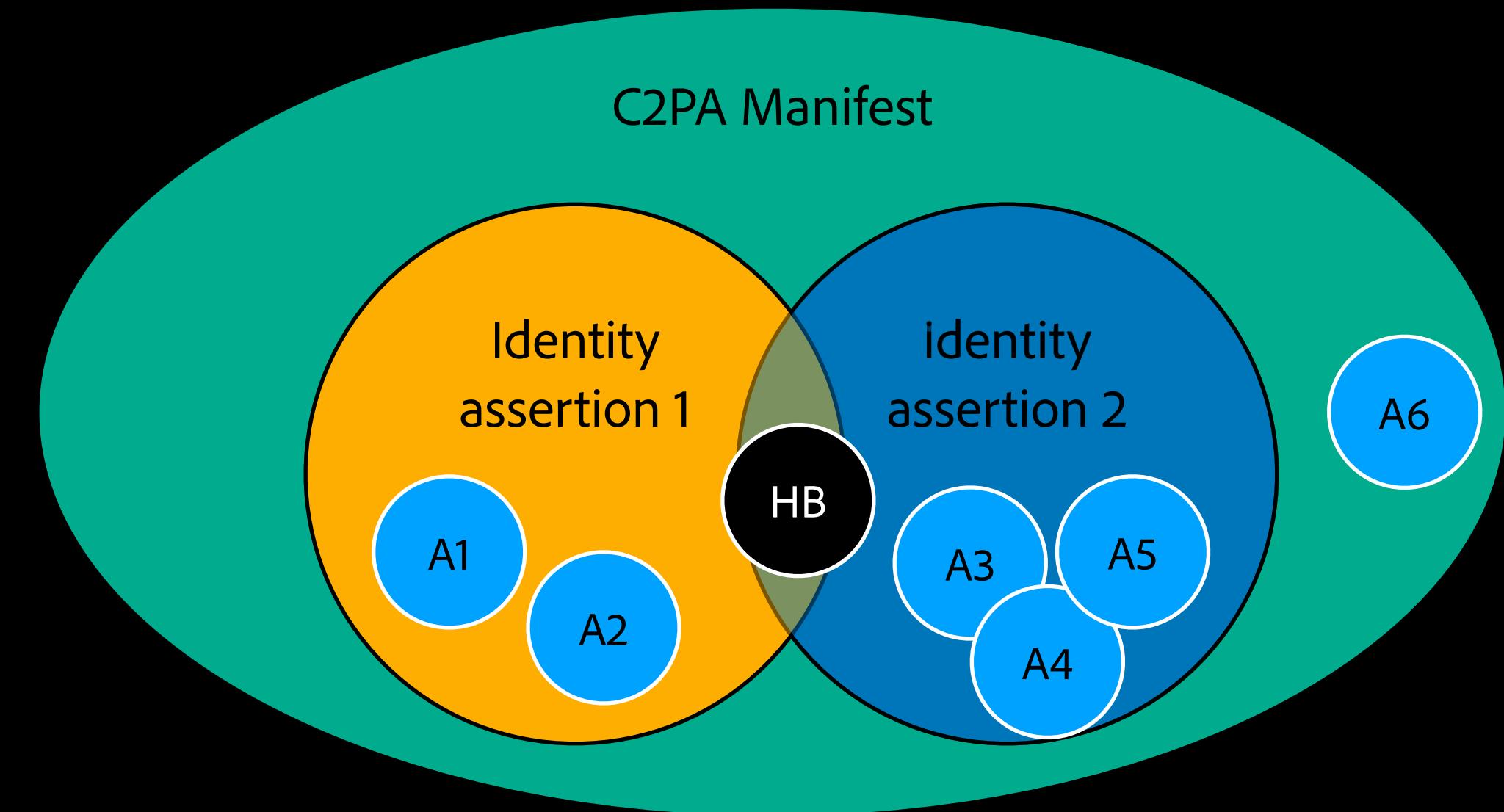
Identity assertion
Overview
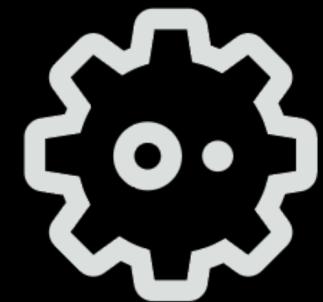
Identity assertion is **optional.**

Identity assertion may be **repeated any number of times** for multiple content producers (for example, interactive multi-author app).
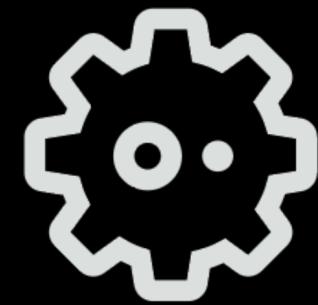
Identity assertion can be **redacted** if needed for privacy/safety reasons.

Identity assertion: Where do credentials come from?
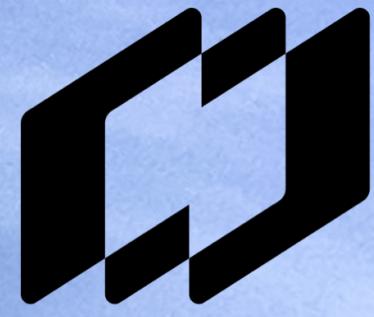
- **Trust anchors in X.509 world are well-established**

- **Trust anchor ecosystems in VC world are evolving**

  - VC world just starting to think about interoperability

- **Investigation:**

  - What do new EU identity credentials look like?

  - Can they support individual content creators in this ecosystem?

  - Can they support *organizational* content creators as well?

Identity assertion
Help us build it!

- **https://creator-assertions.github.io**

- Weekly meetings:

  - Typically on Mondays
    0830 US Pacific / 1130 US Eastern / 1530 UTC

  - Contact me for invitation

Content
Authenticity
Initiative

Thank you!

contentauthenticity.org

Stansbury Island, Utah, 2022
Eric Scouten
(yes, really!)