Content ARCs: Decentralized Content Rights in the Age of Generative AI

Kar Balan, Andrew Gilbert, John Collomosse

15-07-2025

Overview

- Rise of GenAI led to **creative economy tensions**
- Need to **balance creator vs AI developer interests**
- Debate over **opt-in/opt-out** of content for AI training
- UK Government consultation: 3 options, **opt-out most legally feasible**

5:27
Data for Generative AI I...

Can LinkedIn and its affiliates use your personal data and content you create on LinkedIn to train generative AI models that create content?

Use my data for training content creation AI models — Off

This setting controls the training of generative AI models used to create content. When this setting is on LinkedIn and its affiliates may use your personal data and content you create on LinkedIn for that purpose. Learn more.

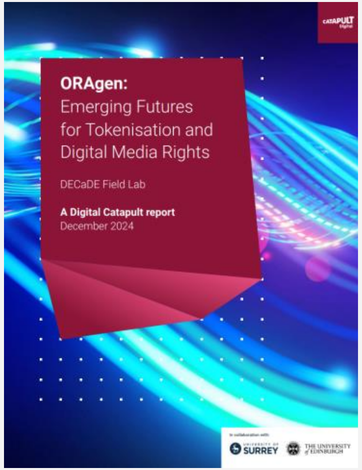
Search results for "AI training **opt-in**"

Background: opt-in/out

Opt-in/out mechanisms:

- **Site-based (location-level):** robots.txt, TDMRep for opt-out.
  - **Advantages:** efficient expression of opt-out in bulk
  - **Downsides:** signal does not persist when content is copied, no mechanism for specifying licensing arrangements for AI re-use
- **Unit-based (asset-level):** IPTC metadata, C2PA opt-in/out.
  - **Advantages:** opt-in/out may be specified separately for granular AI uses
  - **Downsides:** metadata can be stripped by non-compliant platforms

What do creatives want?



CATAPULT Digital

ORAgen: Emerging Futures for Tokenisation and Digital Media Rights

DECaDE Field Lab

A Digital Catapult report

December 2024

In collaboration with: University of Surrey, The University of Edinburgh

National LAB

REPORT: 'TIME TO ACCCT'

Providing Creative Industries and AI developers with a Copyright Framework of Access, Control, Consent, Compensation and Transparency

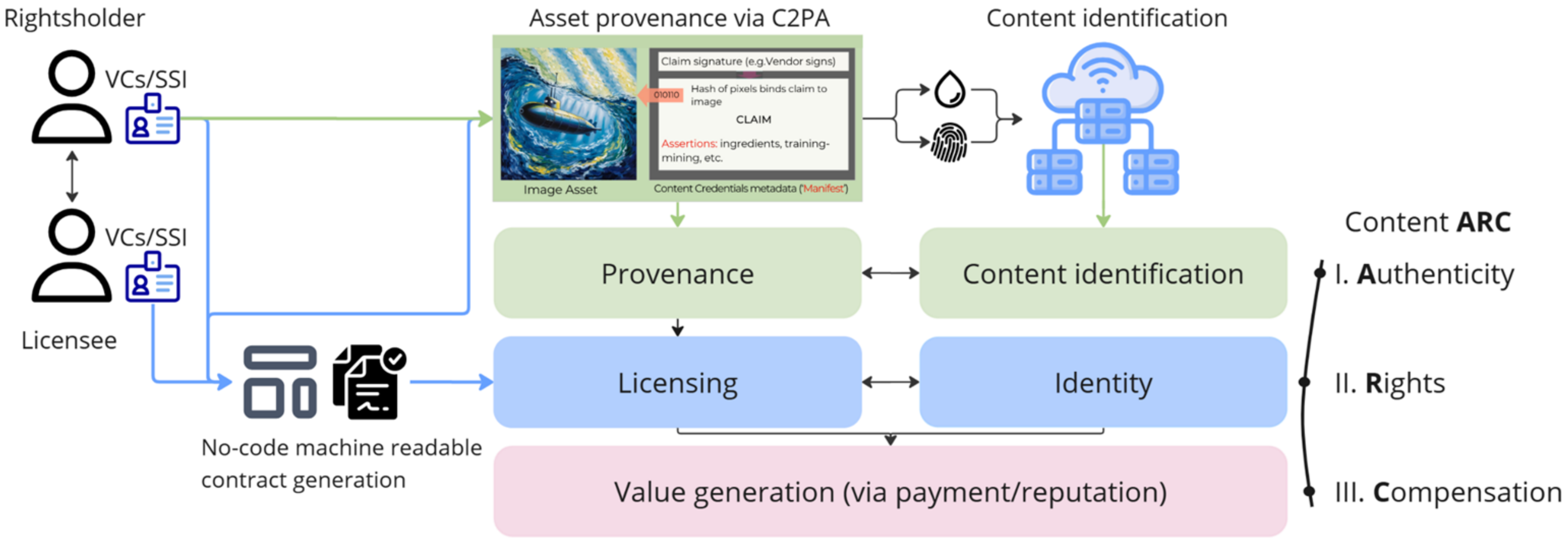
CoSTAR National LAB   DECaDE   SHERIDANS

What creators really want is more complex than opt-in/out:

- **Consent:** Granular opt-in control: who can re-use or sector-specific uses
- **Compensation:** To be paid
- **Sense of agency:** (c.f. Glaze, Nightshade etc.)

Content ARCs framework [CADE 2025] influential in UK Government report

Content ARCs Framework



Rightsholder

VCs/SSI

VCs/SSI

Licensee

Asset provenance via C2PA

Claim signature (e.g.Vendor signs)

010110

Hash of pixels binds claim to image

CLAIM

Assertions: ingredients, training-mining, etc.

Image Asset

Content Credentials metadata ('Manifest')

Content identification

No-code machine readable contract generation

Provenance

Content identification

Licensing

Identity

Value generation (via payment/reputation)

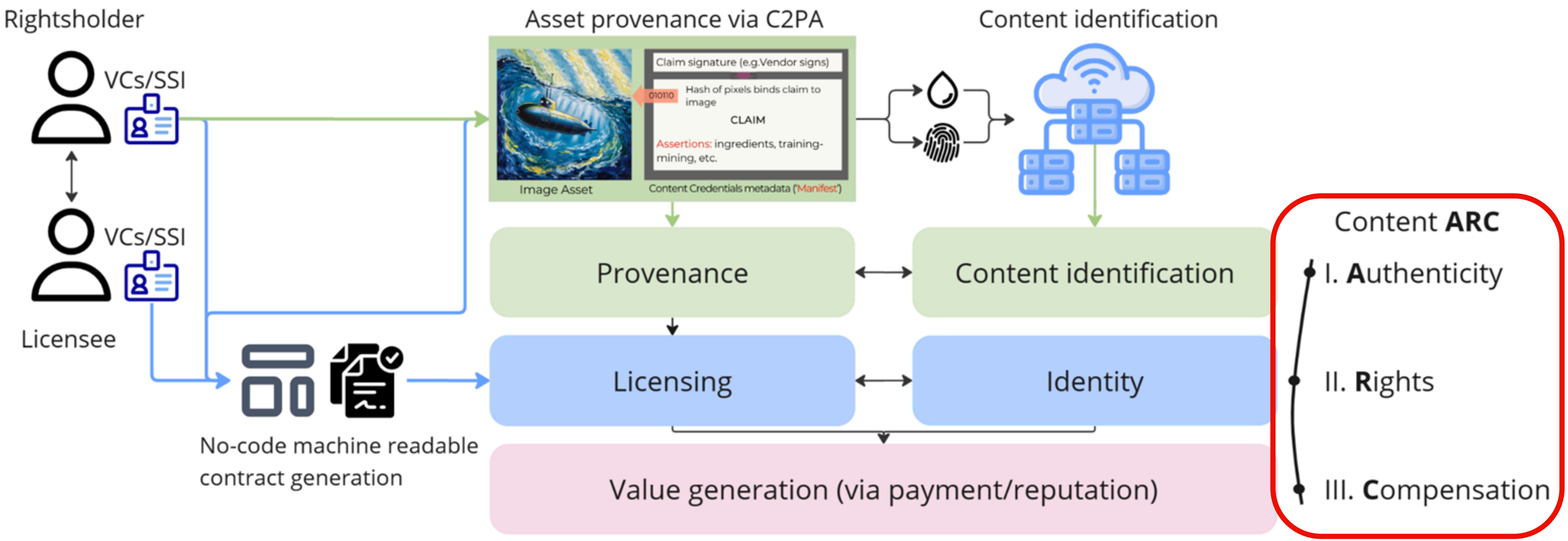
Content **ARC**

I. **A**uthenticity

II. **R**ights

III. **C**ompensation

Content ARCs Framework



Rightsholder

VCs/SSI

VCs/SSI

Licensee

Asset provenance via C2PA

Claim signature (e.g.Vendor signs)

010110

Hash of pixels binds claim to image

CLAIM

Assertions: ingredients, training-mining, etc.

Image Asset

Content Credentials metadata ('Manifest')

Content identification

No-code machine readable contract generation

Provenance

Content identification

Licensing

Identity

Value generation (via payment/reputation)

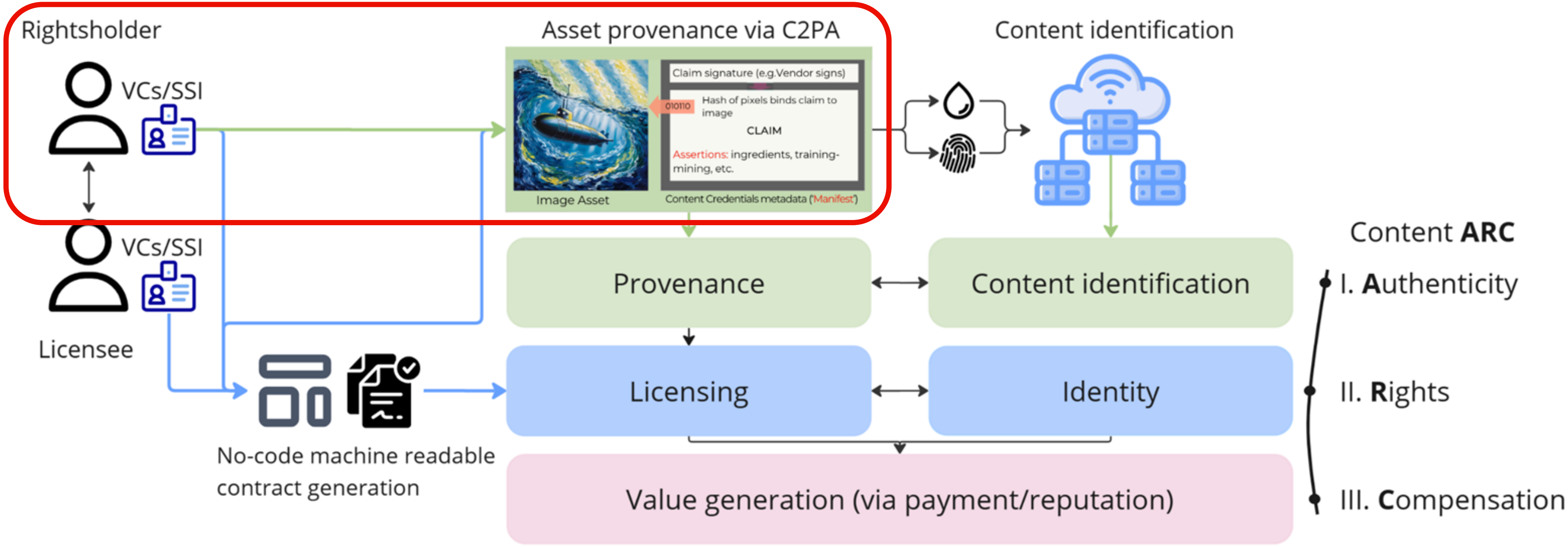
Content **ARC**

I. **A**uthenticity

II. **R**ights

III. **C**ompensation

I. Authenticity: content provenance



Rightsholder

VCs/SSI

Licensee

VCs/SSI

Asset provenance via C2PA

Image Asset

Claim signature (e.g.Vendor signs)

010110

Hash of pixels binds claim to image

CLAIM

Assertions: ingredients, training-mining, etc.

Content Credentials metadata ('Manifest')

Content identification

No-code machine readable contract generation

Provenance

Content identification

Licensing

Identity

Value generation (via payment/reputation)

Content **ARC**

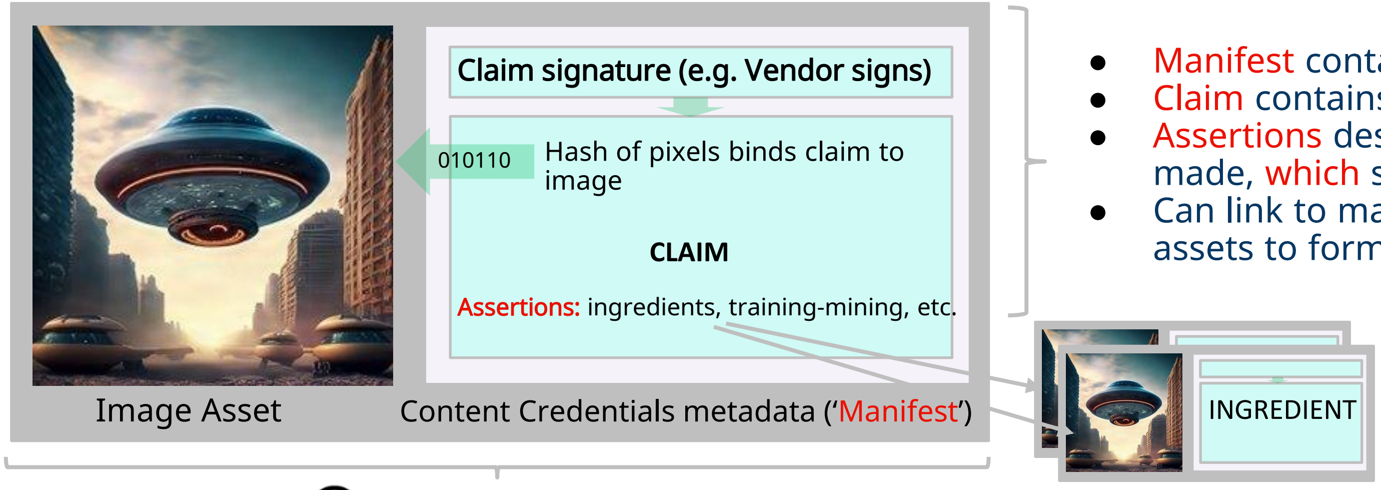
I. **A**uthenticity

II. **R**ights

III. **C**ompensation

I. Authenticity: Content Credentials (C2PA)

Content Credentials establish content provenance and authenticity at scale to give publishers, creators, and consumers the ability to trace the origin of media. **C2PA** is an open cross-industry standard for specifying provenance of media.

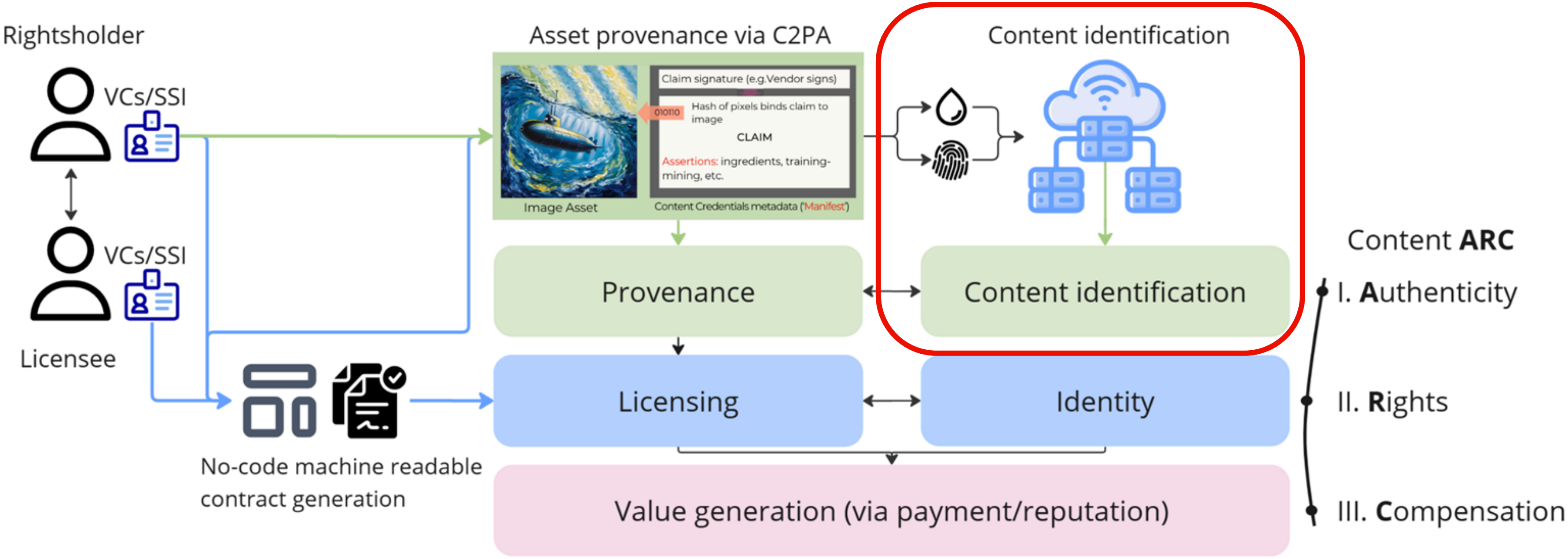

Claim signature (e.g. Vendor signs)
010110
Hash of pixels binds claim to image
CLAIM
Assertions: ingredients, training-mining, etc.
Image Asset
Content Credentials metadata ('Manifest')
INGREDIENT

- Manifest contains signed claim
- Claim contains facts called assertions
- Assertions describe how an asset was made, which sources, actions, etc.
- Can link to manifests in ingredient assets to form a provenance history

content credentials

Other metadata standards: JPEG Trust, IPTC, EXIF

I. Authenticity: content (re)identification



Rightsholder

VCs/SSI

VCs/SSI

Licensee

Asset provenance via C2PA

Claim signature (e.g.Vendor signs)

010110

Hash of pixels binds claim to image

CLAIM

Assertions: ingredients, training-mining, etc.

Image Asset

Content Credentials metadata ('Manifest')

Content identification

Provenance

Content identification

No-code machine readable contract generation

Licensing

Identity

Value generation (via payment/reputation)

Content **ARC**

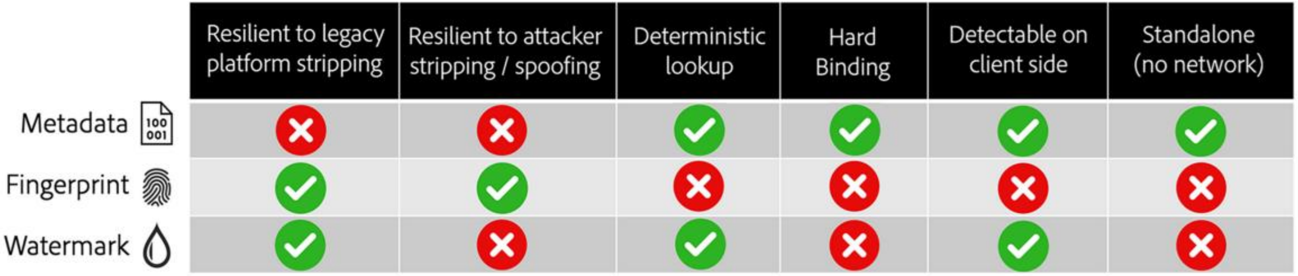
I. **A**uthenticity

II. **R**ights

III. **C**ompensation

I. Authenticity: content (re)identification

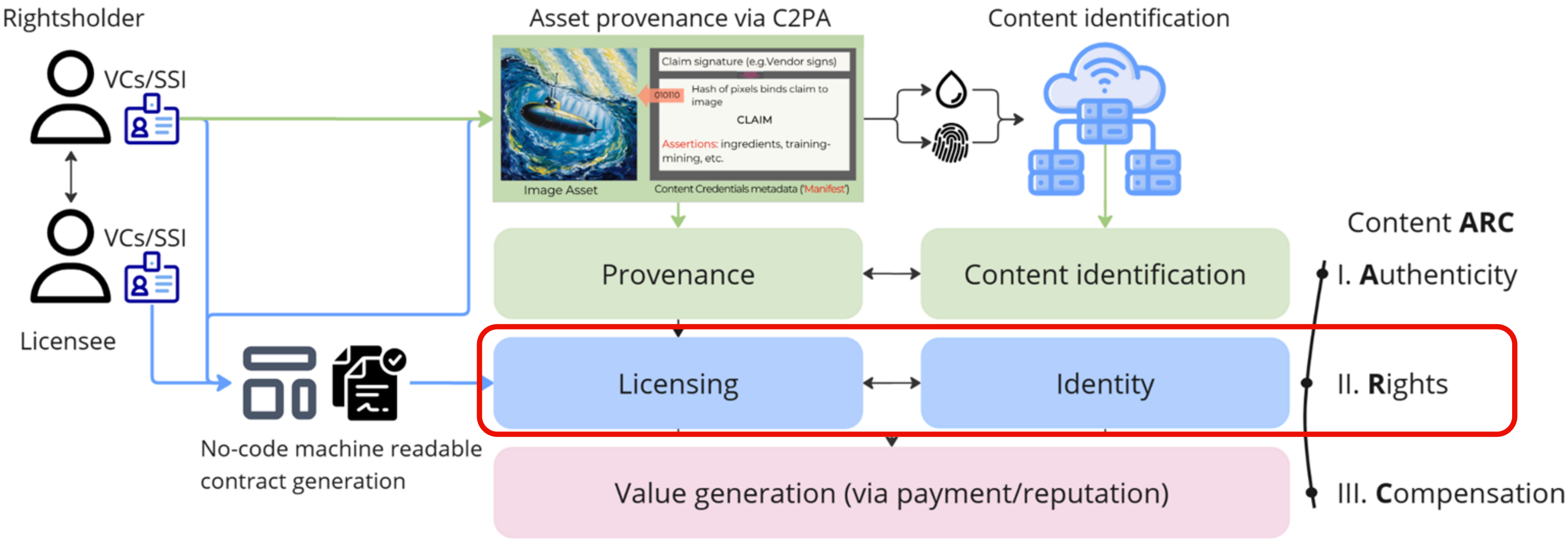
- **Metadata schemes:** fragile–often stripped on platforms
- **Content ID:** watermarking, perceptual hashes (ISCC, PDQ)
  - Durable, survives distribution
  - Risk of collisions, proprietary formats
- **Registries:** map content IDs → metadata
  - Centralised registries don’t scale globally
  - Federated or DLT-based registries enable decentralised trust



	Resilient to legacy platform stripping	Resilient to attacker stripping / spoofing	Deterministic lookup	Hard Binding	Detectable on client side	Standalone (no network)
Metadata	✗	✗	✓	✓	✓	✓
Fingerprint	✓	✓	✗	✗	✗	✗
Watermark	✓	✗	✓	✗	✓	✗

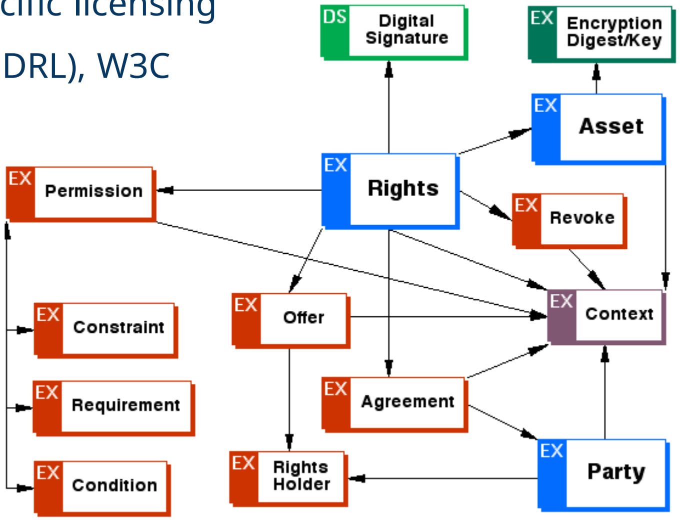
”To Authenticity and Beyond: Building Safe and Fair Generative AI on the Three Pillars of Provenance”. J.

II. Rights


Rightsholder
VCs/SSI
VCs/SSI
Licensee
Asset provenance via C2PA
Claim signature (e.g.Vendor signs)
010110
Hash of pixels binds claim to image
CLAIM
Assertions: ingredients, training-mining, etc.
Image Asset
Content Credentials metadata ('Manifest')
Content identification
Content ARC
Provenance
Content identification
I. Authenticity
Licensing
Identity
II. Rights
No-code machine readable contract generation
Value generation (via payment/reputation)
III. Compensation

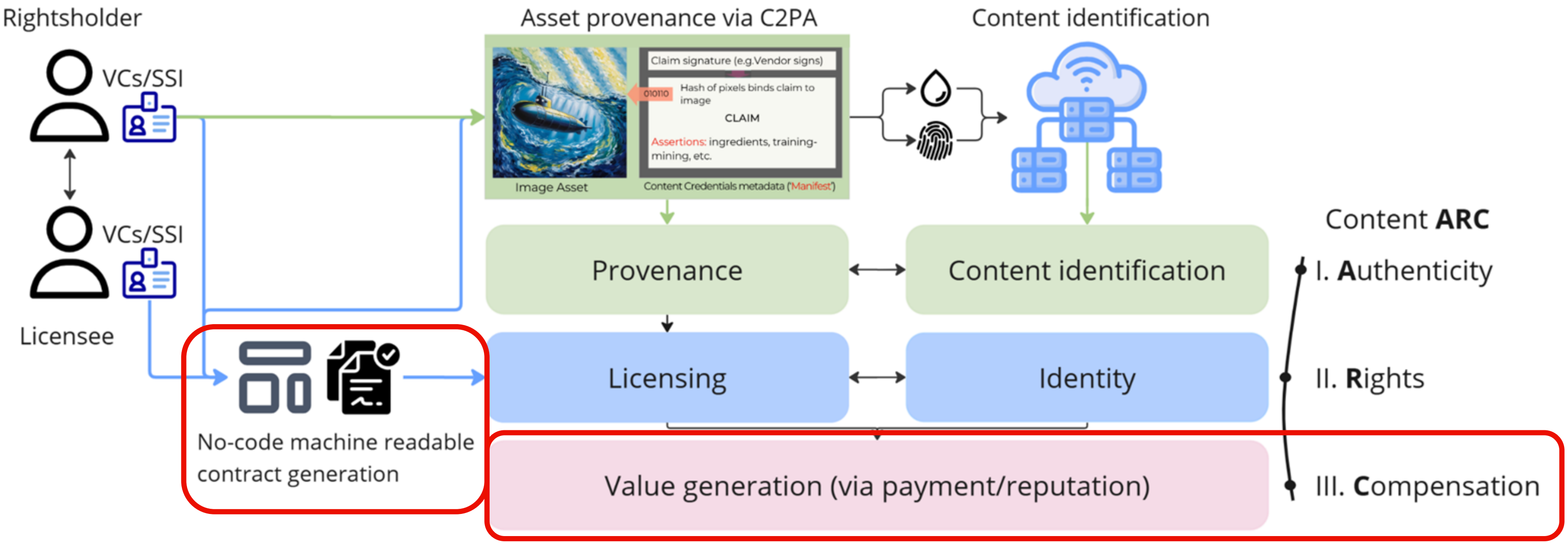
II. Rights

- **Simple metadata:** IPTC (opt-out), C2PA assertions (opt-in/out for AI training/inference) not suitable for complex rights or sector-specific licensing
- **Granular rights representation:** Open Digital Rights Language (ODRL), W3C Resource Description Framework (RDF)
    - **Dynamic licensing:** terms can change over time
    - **Machine-readable**
- Digitally signed licenses = tamper-evident + traceable
- DLT registries enable dynamic, decentralised licensing
- Identity remains a challenge
- No standard yet for rights equivalent to C2PA for provenance



Open Digital Rights Language (ODRL) Version 1.1

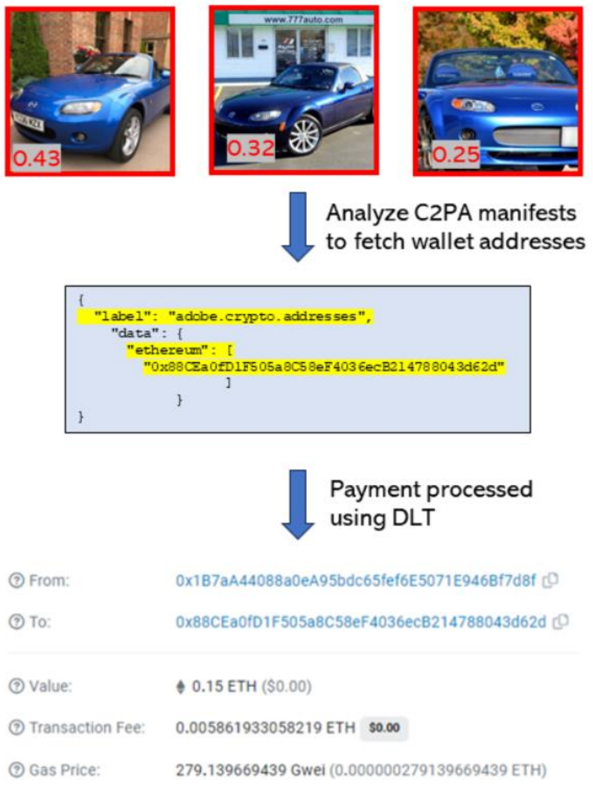
III. Compensation



Rightsholder
VCs/SSI
VCs/SSI
Licensee
Asset provenance via C2PA
Claim signature (e.g.Vendor signs)
010110
Hash of pixels binds claim to image
CLAIM
Assertions: ingredients, training-mining, etc.
Image Asset
Content Credentials metadata ('Manifest')
Content identification
Content ARC
Provenance
Content identification
I. Authenticity
No-code machine readable contract generation
Licensing
Identity
II. Rights
Value generation (via payment/reputation)
III. Compensation

III. Compensation

- **Compensation = licensing + enforcement** from Rights phase
- **Not DRM:** flexible, creator-empowering licensing
- **No code contract templates** to reduce friction
- **Compensation models:**
    - **Royalties or event-based payouts** via smart contracts (NFTs, dataset access)
    - **Attribution-based payouts** using model provenance
    - **Non-financial incentives** (tools, exposure, discounts)
- **Open challenge:** scalable attribution for billion-scale datasets



Analyze C2PA manifests to fetch wallet addresses

```
{
  "label": "adobe.crypto.addresses",
    "data": {
      "ethereum": [
        "0x88CEa0fD1F505a8C58eF4036ecB214788043d62d"
                 ]
            }
}
```

Payment processed using DLT

From: 0x1B7aA44088a0eA95bdc65fef6E5071E946Bf7d8f

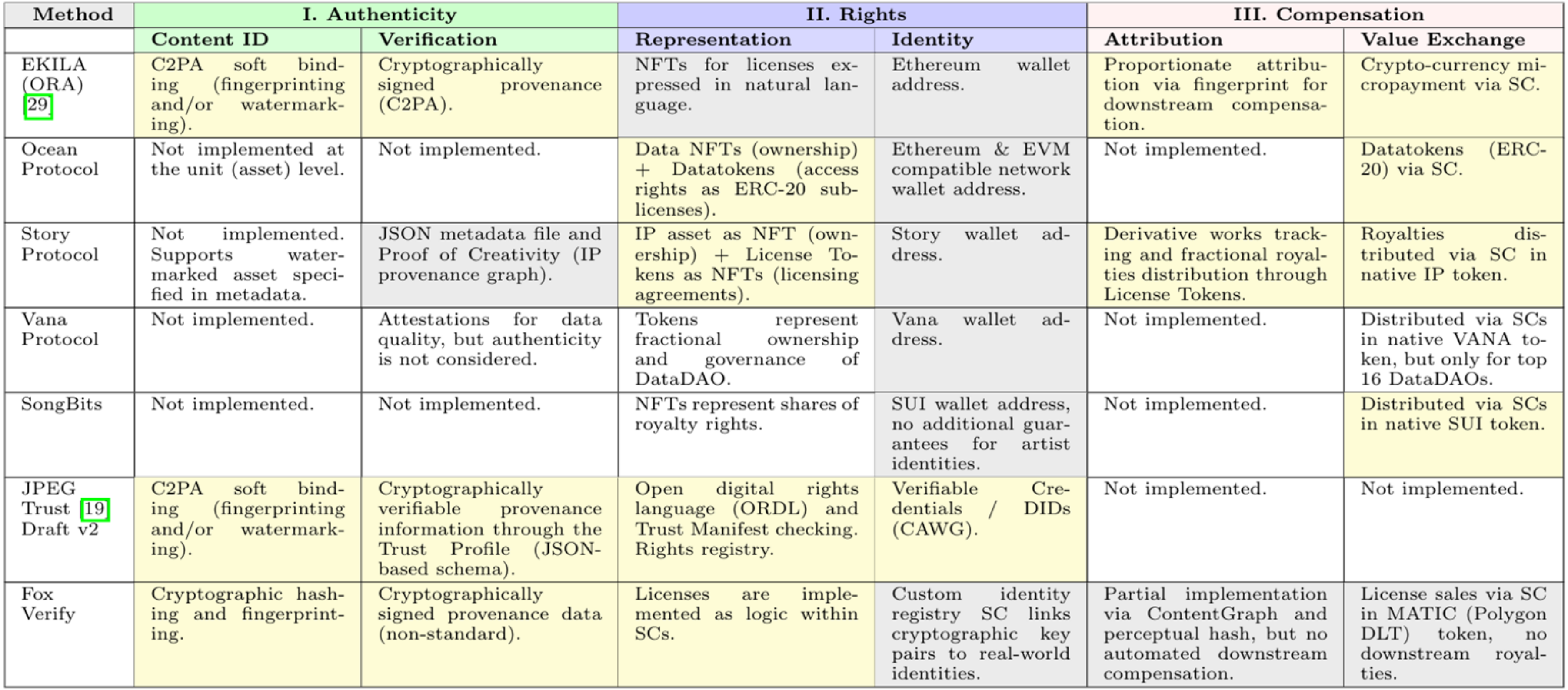
To: 0x88CEa0fD1F505a8C58eF4036ecB214788043d62d

Value: ⧫ 0.15 ETH ($0.00)

Transaction Fee: 0.005861933058219 ETH $0.00

Gas Price: 279.139669439 Gwei (0.000000279139669439 ETH)

Existing Systems



Method	I. Authenticity		II. Rights		III. Compensation	
	Content ID	Verification	Representation	Identity	Attribution	Value Exchange
EKILA (ORA) [29]	C2PA soft binding (fingerprinting and/or watermarking).	Cryptographically signed provenance (C2PA).	NFTs for licenses expressed in natural language.	Ethereum wallet address.	Proportionate attribution via fingerprint for downstream compensation.	Crypto-currency micropayment via SC.
Ocean Protocol	Not implemented at the unit (asset) level.	Not implemented.	Data NFTs (ownership) + Datatokens (access rights as ERC-20 sublicenses).	Ethereum & EVM compatible network wallet address.	Not implemented.	Datatokens (ERC-20) via SC.
Story Protocol	Not implemented. Supports watermarked asset specified in metadata.	JSON metadata file and Proof of Creativity (IP provenance graph).	IP asset as NFT (ownership) + License Tokens as NFTs (licensing agreements).	Story wallet address.	Derivative works tracking and fractional royalties distribution through License Tokens.	Royalties distributed via SC in native IP token.
Vana Protocol	Not implemented.	Attestations for data quality, but authenticity is not considered.	Tokens represent fractional ownership and governance of DataDAO.	Vana wallet address.	Not implemented.	Distributed via SCs in native VANA token, but only for top 16 DataDAOs.
SongBits	Not implemented.	Not implemented.	NFTs represent shares of royalty rights.	SUI wallet address, no additional guarantees for artist identities.	Not implemented.	Distributed via SCs in native SUI token.
JPEG Trust [19] Draft v2	C2PA soft binding (fingerprinting and/or watermarking).	Cryptographically verifiable provenance information through the Trust Profile (JSON-based schema).	Open digital rights language (ORDL) and Trust Manifest checking. Rights registry.	Verifiable Credentials / DIDs (CAWG).	Not implemented.	Not implemented.
Fox Verify	Cryptographic hashing and fingerprinting.	Cryptographically signed provenance data (non-standard).	Licenses are implemented as logic within SCs.	Custom identity registry SC links cryptographic key pairs to real-world identities.	Partial implementation via ContentGraph and perceptual hash, but no automated downstream compensation.	License sales via SC in MATIC (Polygon DLT) token, no downstream royalties.

Yellow = component is present in solution; Gray = partially present; White = absent

Conclusions

Additional resources:

- End-to-end machine-readable permissions
- Transparent licensing & automatic compensation
- **No system yet delivers fully across ARC phases**
- **Key barriers:** registries, identity, legal uncertainty, discovery
- **Real-world pilots needed** to test business models