



# Decentralized Content Identification

## **5th JPEG NFT and Fake Media Workshop**

2022-06-30, *Titusz Pan*

# About us | Titusz Pan

- Entrepreneur with focus on Media Technologies
- Founder and CEO of Craft AG, Freiburg ([craft.de](https://craft.de))
- Open-Source Software Developer
- Inventor and developer of the ISCC



ICS

# ISO/AWI 24138

## Information and documentation — International Standard Content Code

### ABSTRACT

This document specifies the syntax and structure of the International Standard Content Code (ISCC), as an identification system for digital assets (including encodings of text, images, audio, video or other content across all media-sectors). It also describes ISCC metadata and the use of ISCC in conjunction with other schemes, such as the schemes defined by ISO/TC 46/SC 9. An ISCC applies to a specific digital asset and is a data-descriptor constructed from multiple hashes using the algorithms and rules in this document. Organisations, individuals and machines may generate ISCCs for numerous kinds of digital assets and use them for identification and management of those assets. The generation or use of an ISCC in itself does not make any statement or claim about authorship or ownership of the identified content.

# GLOBAL INTERNATIONAL STANDARD

RESOURCES							
INTERNATIONAL HARMONIZED STAGE CODES							
STAGE	SUBSTAGE			90 DECISION			
	00 REGISTRATION	20 START OF MAIN ACTION	60 COMPLETION OF MAIN ACTION	92 REPEAT AN EARLIER PHASE	93 REPEAT CURRENT PHASE	98 ABANDON	99 PROCEED
<b>00 PRELIMINARY</b>	00.00 Proposal for new project received	00.20 Proposal for new project under review	00.60 Close of review			00.98 Proposal for new project abandoned	00.99 Approval to ballot proposal for new project
<b>10 PROPOSAL</b>	10.00 Proposal for new project registered	10.20 New project ballot initiated	10.60 Close of voting	10.92 Proposal returned to submitter for further definition		10.98 New project rejected	10.99 New project approved
<b>20 PREPARATORY</b>	20.00 New project registered in TC/SC work programme	20.20 Working draft (WD) study initiated	20.60 Close of comment period			20.98 Project deleted	20.99 WD approved for registration as CD
<b>30 COMMITTEE</b>	30.00 Committee draft (CD) registered	30.20 CD study/ballot initiated	30.60 Close of voting/ comment period	30.92 CD referred back to Working Group		30.98 Project deleted	30.99 CD approved for registration as DIS
<b>40 ENQUIRY</b>	40.00 DIS registered	40.20 DIS ballot initiated: 12 weeks	40.60 Close of voting	40.92 Full report circulated: DIS referred back to TC or SC	40.93 Full report circulated: decision for new DIS ballot	40.98 Project deleted	40.99 Full report circulated: DIS approved for registration as FDIS
<b>50 APPROVAL</b>	50.00 Final text received or FDIS registered for formal approval	50.20 Proof sent to secretariat or FDIS ballot initiated: 8 weeks	50.60 Close of voting: Proof returned by secretariat	50.92 FDIS or proof referred back to TC or SC		50.98 Project deleted	50.99 FDIS or proof approved for publication
<b>60 PUBLICATION</b>	60.00 International Standard under publication		60.60 International Standard published				
<b>90 REVIEW</b>		90.20 International Standard under systematic review	90.60 Close of review	90.92 International Standard to be revised	90.93 International Standard confirmed		90.99 Withdrawal of International Standard proposed by TC or SC
<b>95 WITHDRAWAL</b>		95.20 Withdrawal ballot initiated	95.60 Close of voting	95.92 Decision not to withdraw International Standard			95.99 Withdrawal of International Standard

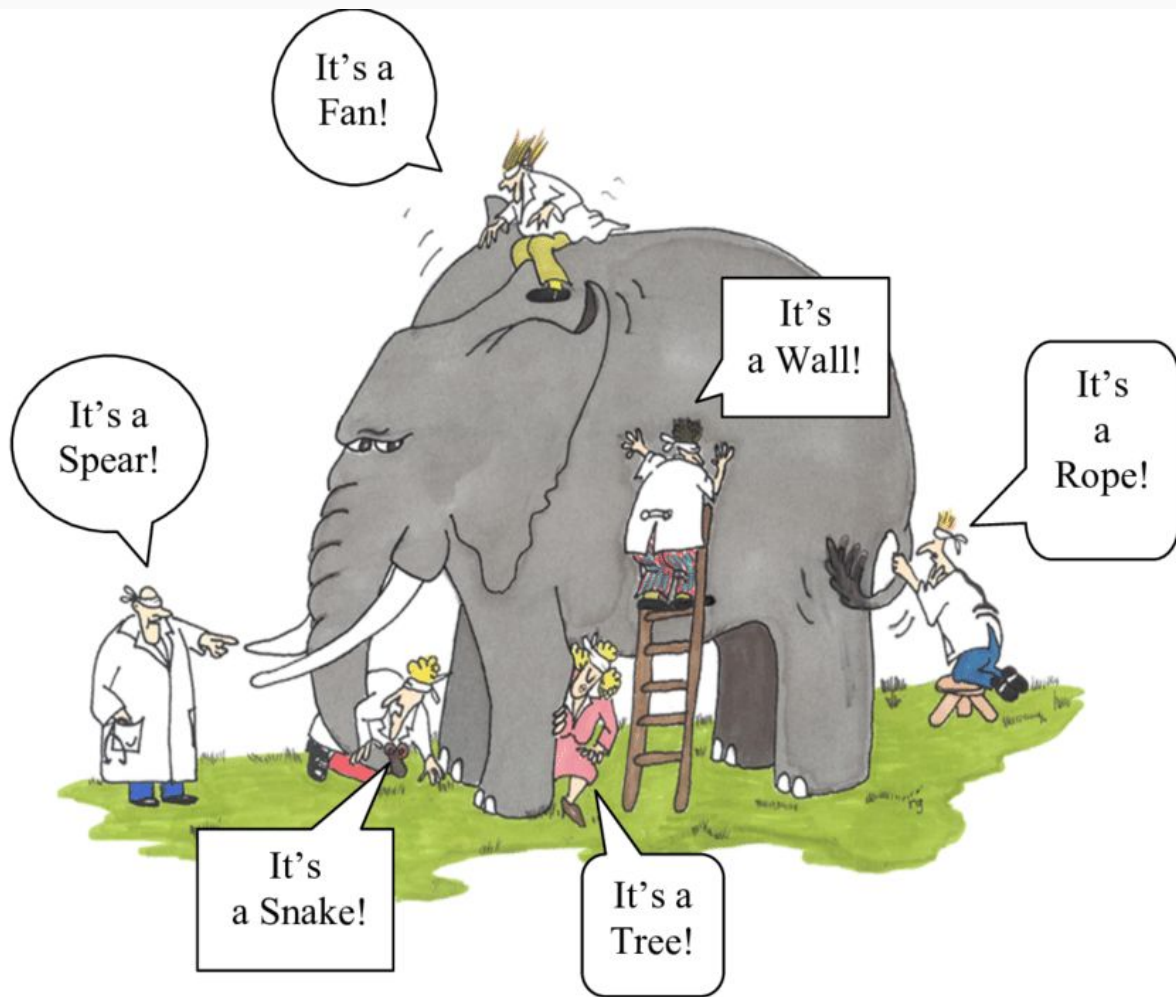
We are here

October 2022

Goal

# What is **Not** ISCC

- A persistent identifier
- A content recognition system
- A cryptographic hash
- A forensics tool
- A centralized content registry





# Modern and Open Content-Based Identification



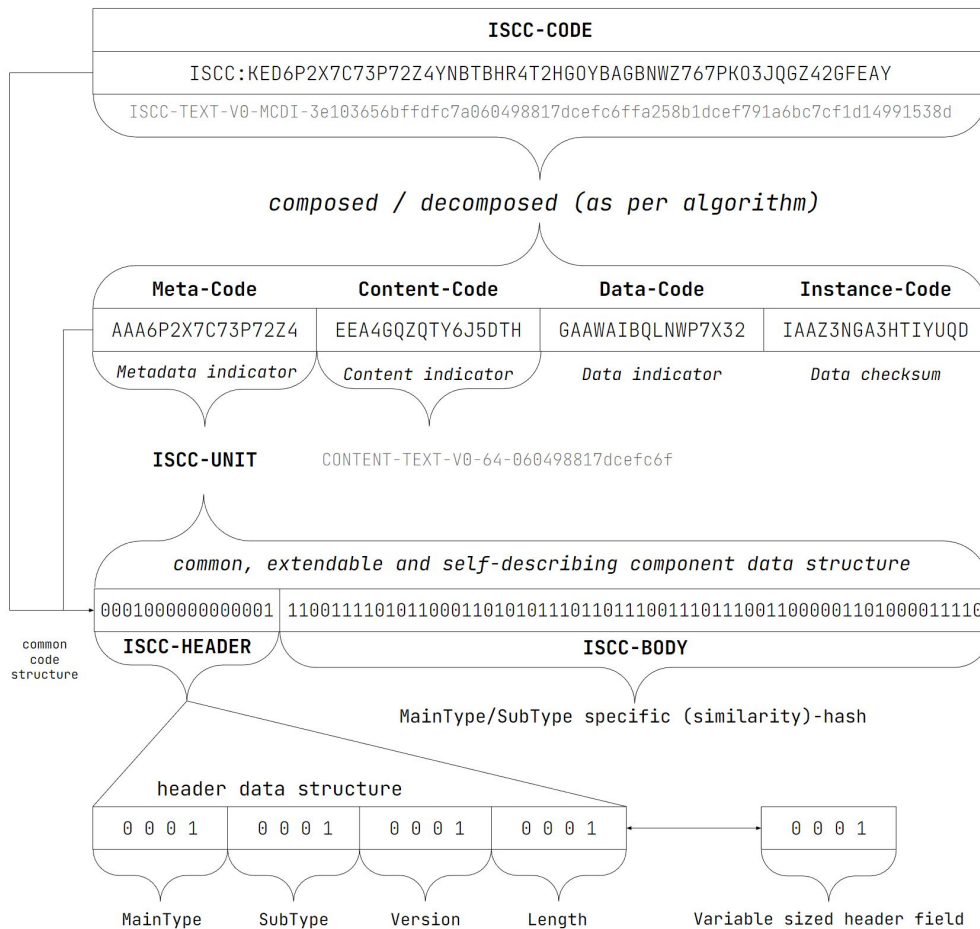
- A universal code for digital text, image, audio, video
- Lightweight, multi-faceted fingerprint designed for **digital content**
- Cross-sector applicability (photo, journalism, book & academic publishing, music, film etc.)
- Goal: establish **content** as the **subject** of **transactions** in decentralized and networked environments

# ISCC - Universal Soft-Hash Codec

A standardized way to encode and arrange fingerprints in a self-describing scheme.

- Compact
- Interoperable
- Modular
- Extensible
- Self-Describing
- Open Source

An initial set of standardized open source algorithms





# Decentralized Content Identification

## ISCC is an identifier & fingerprint

Identifier	Example	Bits	Method
<b>UUID</b>	550e8400-e29b-11d4-a716-446655440000	128	Random / Hash / Time
<b>SHA256</b>	a1bdd0de0d1f27b227cbf43ac110bb09827a40d734ea0c29585c98a34b80413d	256	Cryptographic Hash
<b>ISCC-CODE</b>	ISCC:KADV5PDFXBL7H6BXFFW64KVNP6UGTUZC2CJTDBKMFYTTZPLQQVX22FI	272	Multifaceted Fingerprint
<b>ISCC-ID</b>	ISCC:MEAJU5AXCP0IOYFL	80+	DLT / Short FP / Counter

There are 3 categories of content identifiers: Centralized Registry (Database IDs), Random IDs (UUID), Cryptographic Hashes (Content-Based). ISCC innovates with a combination cryptographic hashes, similarity preserving compact binary codes, standardized fingerprints and DLT.



# Layers of “Content” Identification

There is often confusion about what exactly is being identified by a given system.

In our model for digital content identification we distinguish 6 layers that exist naturally on a scale from abstract to concrete.

Existing content identifiers usually operate on one or two of these layers.

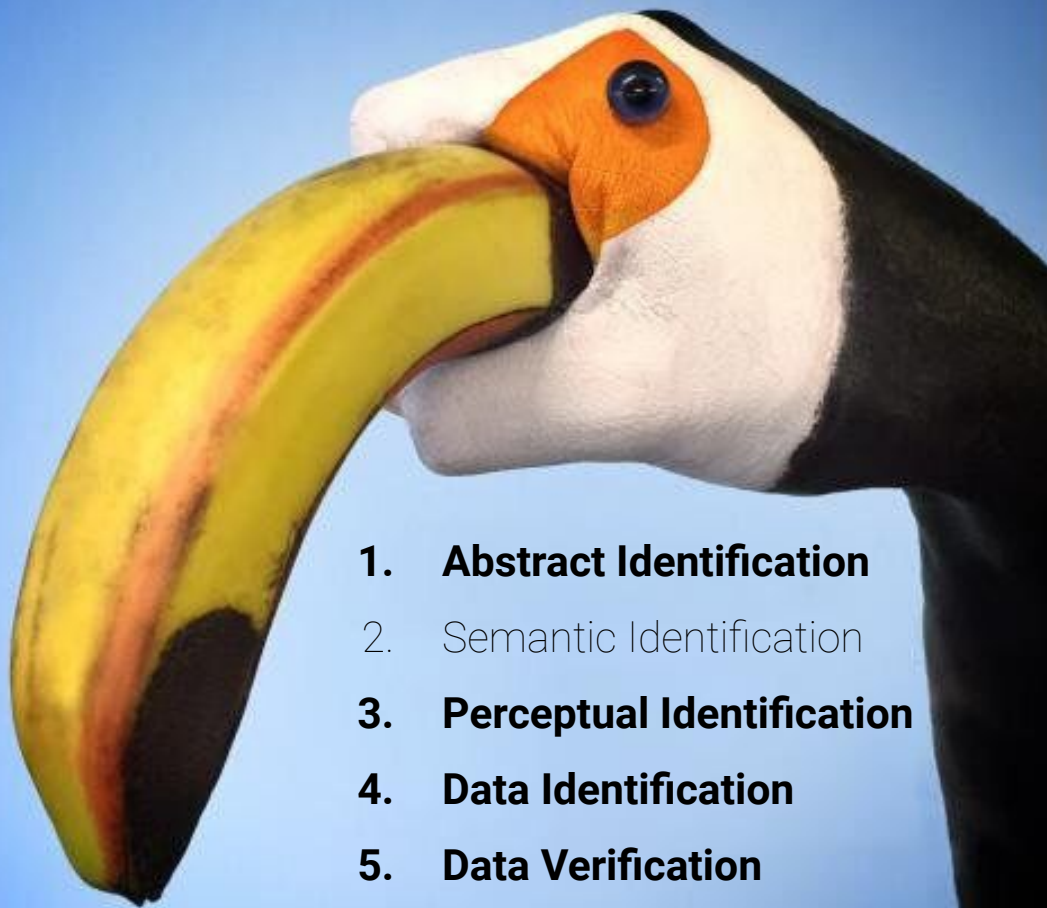


Image courtesy of Imgur

1. **Abstract Identification**
2. Semantic Identification
3. **Perceptual Identification**
4. **Data Identification**
5. **Data Verification**
6. Individual Copy



International Standard Content Code

The **DNA** of your digital content  
Estimate similarity by comparing ISCC-CODEs

ISCC:KADV5PDFXBL7HGBXFFW64KVNP6UGTUZC2CJTDBKMFYTTZPLQQVX22FI

ISCC:

Meta-Code

AAAV5PDFXBL7HGBX

Content-Code

EAASS3P0FKWX7KDJ

Data-Code

GAA5GIWQSMYYKTBO

Instance-Code

IAAS0PF50CCW7LIV

Abstract & Persistent

Concrete & Volatile

Metadata  
Similarity

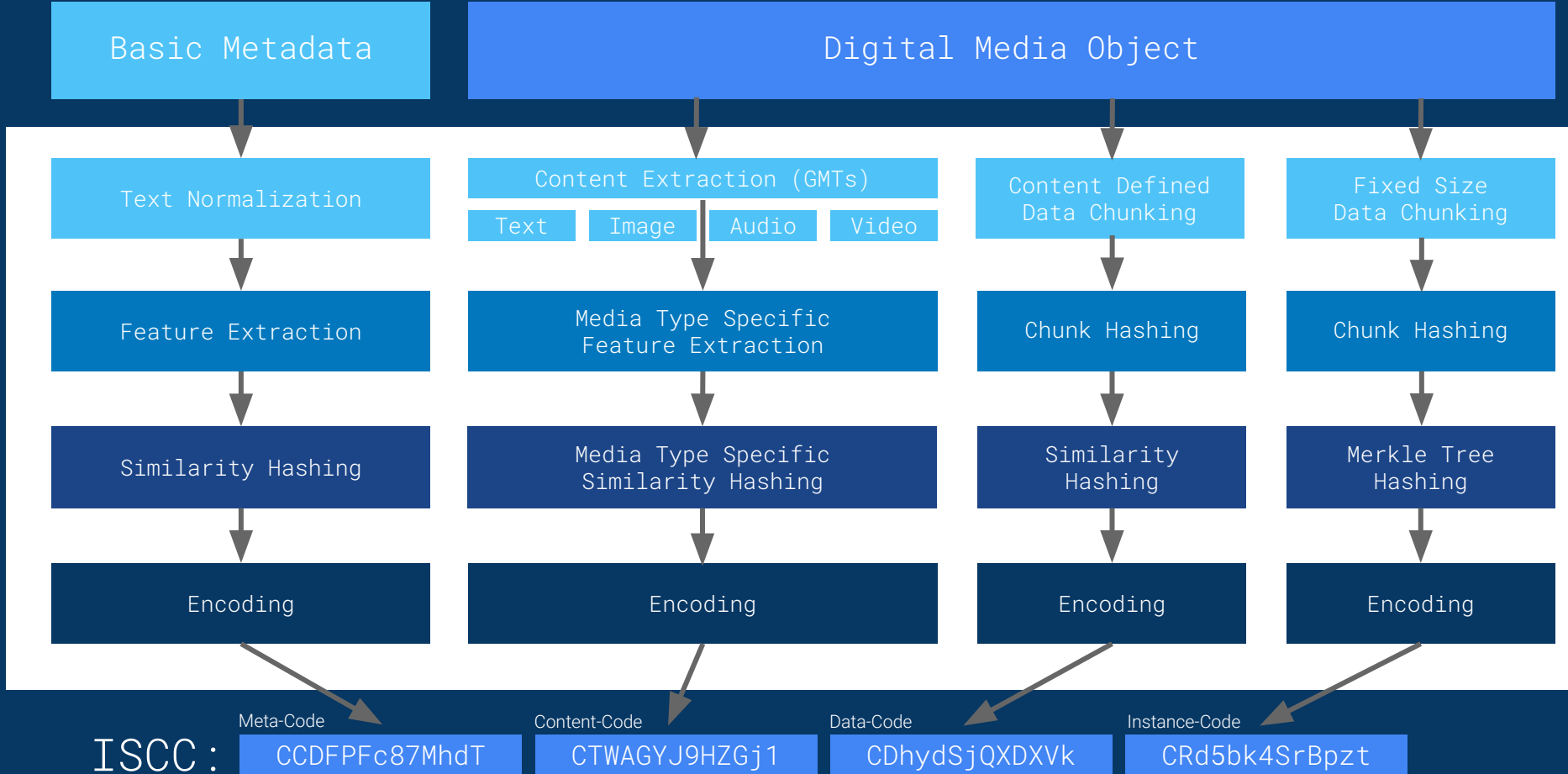
Content  
Similarity

Data  
Similarity

Data  
Integrity

Components are self-describing and can be used standalone or in combination

# Overview of ISCC Creation Process

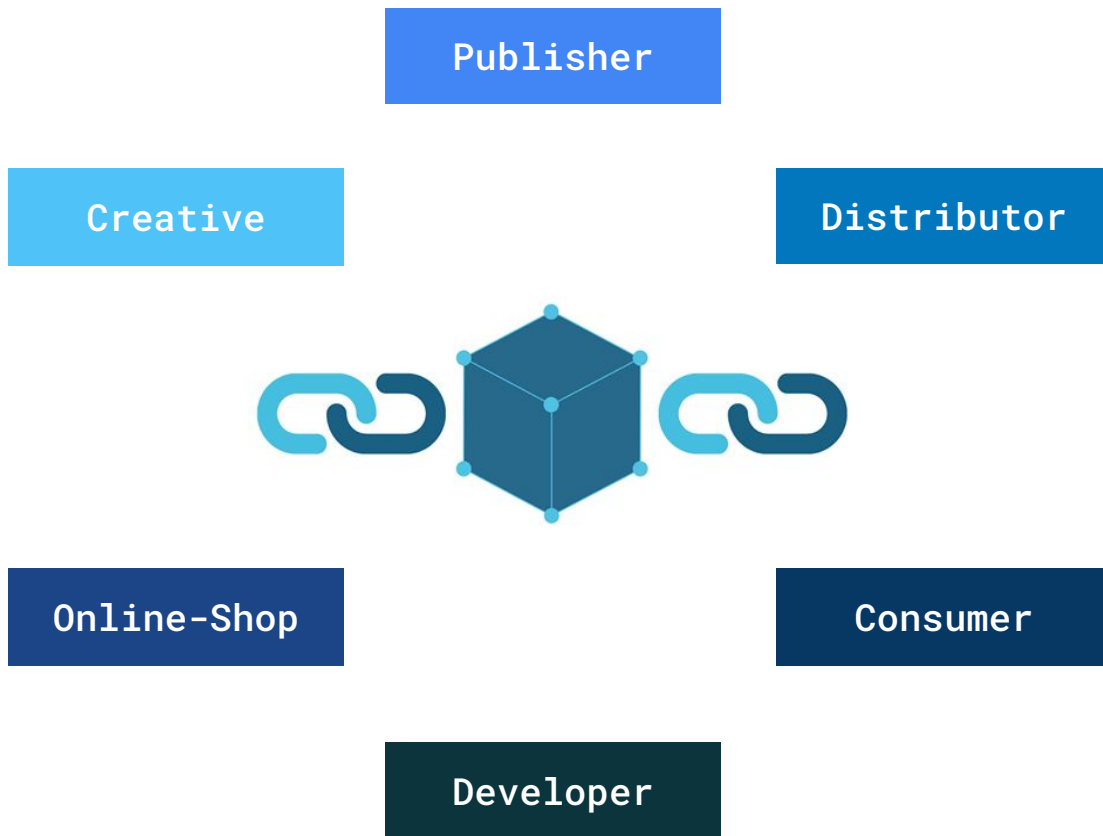


# ISCC - Decentralized Content Identification

In a multi-sided ecosystem **anybody** may have legitimate interest to create, lookup or register an **identifier** for some content.

Authorship or copyright is **not** a requirement. But **an identifier is a requirement** to communicate and agree on authorship, copyright ...

Authoritative linking of **Identifier <-> Content** is done by open standardized fingerprinting algorithms.



# ISCC-CODE + Shared Ledger Entry = ISCC-ID

ISCC-ID - 10 bytes (growable)



short, globally unique, persistent, resolvable, owned, verifiable, authenticated

ISCC:MEAJU5AXCP0IOYFL → ID-POLYGON-V0-72-9a741713dc8760ab-3



Header: ID-POLYGON-V0-64



Body: Similarity hash of ISCC-UNITs XOR-ed with hash(wallet)



Uniqueness counter: 7 bits + 1 continuation bit (uvarint)

## ISCC-ID - Global Decentralized Registration Protocol (Draft):

1. Generate **ISCC-CODE**
2. Choose any supported public shared ledger
3. Publish **ISCC-CODE** with a signed transaction
4. The rest is “history” ...
5. Any indexer can calculate your unique short **ISCC-ID**

**ISCC-ID** binds digital content to actor identities. The **ISCC-ID** emerges naturally by executing a protocol over DLT transactions that are: **Public, Authenticated, Ordered, Immutable**. Anybody who follows the ISCC indexing protocol can build the state of the full **ISCC-ID** Index. The same **ISCC-CODE** may be registered by multiple actors on different chains. They all get different but matchable **ISCC-IDs**.

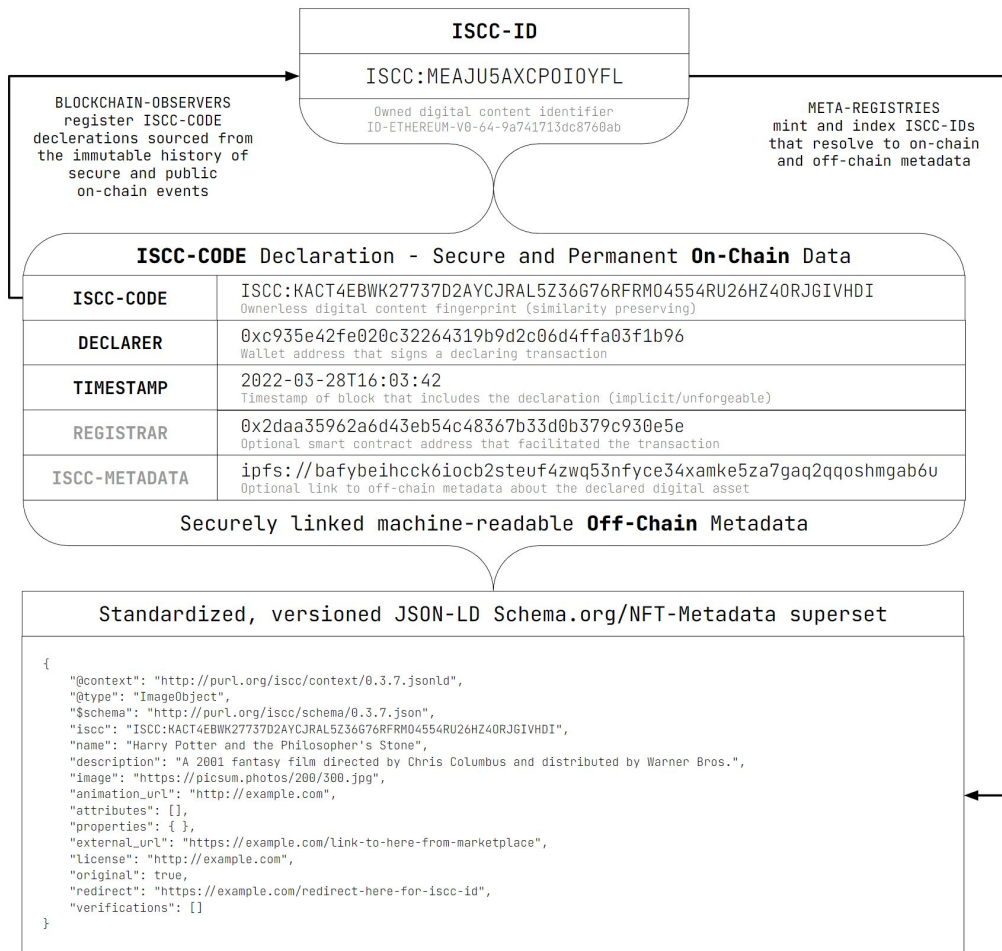
Demo: <https://testnet.iscc.id>

# ISCC - Decentralized Content Identification

In a multi-sided ecosystem **anybody** may have legitimate interest to create, lookup or register an **identifier** for some content.

Authorship or copyright is **not** a requirement. But **an identifier is a requirement** to communicate and agree on authorship, copyright ...

Authoritative linking of **Identifier <-> Content** is done by open standardized fingerprinting algorithms.



# Comparing two **ISCC** Codes yields various insights




	Meta-Code	Content-Code	Data-Code	Instance-Code	Explanation
1	=	=	=	=*	<b>Totally identical file</b> (same metadata, content structure, file encoding and file)
2	= =	=	=	=*	Different metadata, same content, file encoding and identical file > e.g. a special edition or inconsistent metadata
3	= =	= =	= =	= =	Totally different file (different metadata, content structure, file encoding and file)
4	= or ~	= =	= =	= =	Same/similar* metadata, but different content and file encoding and file, e.g. manual clustering
5	= or ~	= or ~	= =	= =	Same/similar metadata, same/similar content but in a different file encoding, e.g. related product
6	= or ~	= or ~	= or ~	= =	Same/similar metadata, same/similar content in same/similar file encoding
7	= =	= or ~	= =	= =	Different metadata, same/similar content but in a different file encoding
8	= =	= or ~	= or ~	= =	Different metadata, but same/similar content and file encoding, e.g. a special edition

=\* compare top-hash of both files to be sure there is no accidental Instance-Code collision.

first 3 components are compact binary codes (bit vectors) that can be compared to measure estimated similarity by hamming distance

# ISCC - Status

ISO

Standards About us News **Taking part** Store   

RESOURCES

INTERNATIONAL HARMONIZED STAGE CODES

STAGE	SUBSTAGE			DESCRIPTION						
00 PRELIMINARY	00	REGISTRATION	20	START OF MAIN ACTION	60	COMPLETION OF MAIN ACTION	92	REFLECT AN EARLIER PHASE	99	PRECED
	00.00	00.00	20.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
10 PROPOSAL	10.00	10.00	10.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	10.00	10.00	10.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
20 PREPARATORY	20.00	20.00	20.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	20.00	20.00	20.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
30 COMMITTEE	30.00	30.00	30.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	30.00	30.00	30.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
40 ENQUIRY	40.00	40.00	40.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	40.00	40.00	40.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
50 APPROVAL	50.00	50.00	50.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	50.00	50.00	50.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
60 PUBLICATION	60.00	60.00	60.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	60.00	60.00	60.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
90 REVIEW	90.00	90.00	90.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	90.00	90.00	90.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
95 RETIRABLE	95.00	95.00	95.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00
	95.00	95.00	95.00	Reserved for new project/under review	60.00	60.00	92.00	92.00	99.00	99.00

# ISO/AWI 24138

ISO/TC 46/SC 9/WG 18

# Roadmap:

- Standardize ISCC codec & algorithms
- Demo applications (cli/web/desktop)
- Use case testing
- Efficient indexing and NNS (matching)
- Semantic-Code
- Cross-chain meta-registry
- see: <https://iscc.codes>