Position Paper IAB Workshop on AI-CONTROL (aicontrolws)

TDM·AI – Making Unit-Based Opt-out Declarations to Providers of Generative AI

Authors

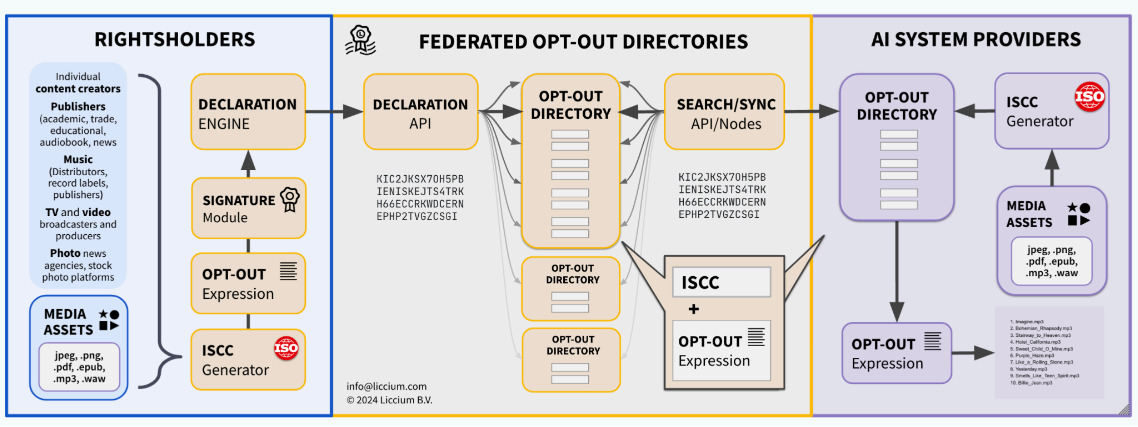
Sebastian Posth (M.A.), Founder and CEO of Liccium.com, sebastian@liccium.com

Sabine Richly, German Qualified Lawyer (LL.M., MBA), sabine.richly@medialaw.digital

Abstract

TDM·AI is a protocol for creators and rightsholders to inseparably bind their machine-readable preferences for text and data mining (TDM) to digital media assets, specifically tailored for training models and applications of generative AI. TDM·AI addresses the main problem in controlling AI crawlers, namely the problem of metadata binding, by proposing a reliable method of soft-binding restrictions or permissions to use content for training models of generative AI to content-derived identifiers.

The TDM·AI protocol utilises the International Standard Content Code (ISCC), a new ISO standard for the identification of digital media content (ISO 24138:2024[1]) and Creator Credentials[2], based on W3C recommendation for cryptographically verifiable credentials[3], to ensure that verifiable and machine-readable declarations include proper attribution of preferences and claims to the legitimate rightsholders. Although the protocol has its origins in the European DSM Directive on Copyright 2019/790, Article 4, it may in many cases also be applicable to content published by rightsholders outside the EU.


RIGHTSHOLDERS
Individual content creators
Publishers (academic, trade, educational, audiobook, news
Music (Distributors, record labels, publishers)
TV and video broadcasters and producers
Photo news agencies, stock photo platforms
MEDIA ASSETS
jpeg, .png, .pdf, .epub, .mp3, .waw
DECLARATION ENGINE
SIGNATURE Module
OPT-OUT Expression
ISCC Generator
ISO
FEDERATED OPT-OUT DIRECTORIES
DECLARATION API
KIC2JKSX70H5PB IENISKEJTS4TRK H66ECCRKWDCERN EPHP2TVGZCSGI
OPT-OUT DIRECTORY
OPT-OUT DIRECTORY
OPT-OUT DIRECTORY
SEARCH/SYNC API/Nodes
KIC2JKSX70H5PB IENISKEJTS4TRK H66ECCRKWDCERN EPHP2TVGZCSGI
ISCC
+
OPT-OUT Expression
info@liccium.com
© 2024 Liccium B.V.
AI SYSTEM PROVIDERS
OPT-OUT DIRECTORY
ISCC Generator
ISO
MEDIA ASSETS
jpeg, .png, .pdf, .epub, .mp3, .waw
OPT-OUT Expression
1. Imagine.mp3
2. Bohemian_Rhapsody.mp3
3. Stairway_to_Heaven.mp3
4. Hotel_California.mp3
5. Sweet_Child_O_Mine.mp3
6. Purple_Haze.mp3
7. Like_a_Rolling_Stone.mp3
8. Yesterday.mp3
9. Smells_Like_Teen_Spirit.mp3
10. Billie_Jean.mp3

Overview TDM·AI Protocol – Source

[1] ISO 24138 ISCC – International Standard Content Code, https://www.iso.org/standard/77899.html.
[2] Creator Credentials, https://docs.creatorcredentials.com.
[3] Verifiable Credentials Data Model v2.0, https://www.w3.org/TR/vc-data-model-2.0/.

Motivation

Given the current digital AI landscape in the context of an evolving international regulatory environment, there is an urgent need for a reliable way for content creators and other rightsholders to declare their consent or reservation to TDM for the purpose of training models and applications of generative AI, capable of generating text, images, and other content. The TDM·AI protocol aims to provide creators and other rightsholders with a simple and standardised way to make a machine-readable declaration as to whether or not their content may or may not be used for this specific purpose. The key differentiator of the TDM·AI protocol is that the opt-out declaration – or its revocation, which could potentially facilitate a licensing transaction – can be resolved directly from the content-derived identifier, the ISCC code, meaning that it is easily accessible to users such as AI providers using open-source identifier technology. By using ISCC, the protocol ensures a reliable method of identifying content that is robust to common problems such as the loss of embedded metadata, removal of watermarks or steganographic data, or other alteration or manipulation of content. The use of verifiable credentials adds a further layer of trust and verifiability, ensuring that the declarations are genuine and can be traced back to the original rightsholder, depending on their privacy needs and preferences.

The TDM·AI protocol is motivated by the need to:

- Provide a clear and simple way for rightsholders to declare their rightsholder preferences with regards to training models of generative AI;
- Ensure that AI providers and other stakeholders can easily read, understand and respect rightsholder preferences by machine technology.

Recommendations for Rightsholder Preferences

In this paper we would like to present the TDM·AI and discuss its underlying principles as recommendations for rightsholder preferences with regards to opt-out declarations for content being used to train models and applications of generative AI.

Rightsholder preferences should be:

- Easily discoverable, accessible and machine-readable (legal requirement);
- Inseparably bound to the content (unit/media asset based approach);
- Reliable when content is shared and distributed;
- Resilient to content manipulation or alteration;
- Resilient to the removal of embedded metadata (watermarks, and steganographic data)
- Containing verifiable attribution through digital signatures and certificates which authenticate the declaration's source and prevent false claims;
- Utilising verifiable timestamps;
- Based on reliable and transparent international standards (ISO, W3C).

Options for Metadata Binding

When it comes to managing opt-out declarations for TDM in a machine-readable way, three different approaches are discussed:

1. Location/Domain-Based Binding – Rightsholder preferences for web-published content are included in robots.txt-file or in HTML/HTTP metadata of the domain, e.g. Robots.txt[4], TDMrep[5];
2. Hard-Binding – Metadata is embedded directly into the media file, e.g. C2PA.org[6];
3. Soft-Binding – Metadata is provided in an external (sidecar) file and linked to the ISCC code.

Issues of Location/Domain-Based Approaches

Robots.txt has emerged as a practical solution to express rightsholders' preferences for web-published content, it is easy to implement, widely used and a recognised standard.

However, location/domain-based are not always effective in the following situations:

- When content is shared on websites rightsholders do not control, such as social media;
- When content is properly licensed to be used by licensors who do not use robots.txt or honour the settings specified in the original rightsholder's robots.txt file, which can happen when licensing stock photos;
- When copyrighted content is illegally republished on the Internet without authorisation.

Issues of Hard Binding

Content creators and publishers can use apps that support the C2PA method to create and embed cryptographically verifiable metadata containing information about the asset's creation and edit actions, copyright, licences, capture device details, and software used. This manifest may include TDM assertions that enable "a human actor to provide a C2PA Manifest Consumer information about whether an asset with C2PA metadata may be used as part of a data mining or AI/ML training workflow."[7] The assertions are designed to be hashed and gathered into a verifiable claim that is digitally signed, ensuring the integrity of the claim.

However, hard-binding of the embedded assertions to the content breaks in the following situations:

[4] Robots Exclusion Protocol, RFC 9309, https://datatracker.ietf.org/doc/html/rfc9309.

[5] TDM Reservation Protocol (TDMRep), Final Community Group Report 02 February 2024, https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/

[6] Coalition for Content Provenance and Authenticity (C2PA), https://c2pa.org/.

[7] Creator Assertions Working Group, Training and Data Mining Assertion https://creator-assertions.github.io/training-and-data-mining/1.1-draft/

- When embedded metadata (or the certificate) is removed from the media file;
- When content is altered or manipulated even to a small extent, as the method is based on cryptographic hashing;
- When content is converted into a different file format, compressed or screenshotted.

However, these are very common settings and use cases when content is shared online or on social media platforms that resize or compress media files or remove embedded metadata for security and business reasons.

Advantages of TDM·AI

The TDM·AI protocol proposes a reliable method to soft-bind rightsholder preferences to the content-derived identifier. Creators and rightsholders can generate ISCC codes directly from their content and choose the rightsholder preference for the content. ISCC codes and the selected preferences can be publicly declared in a network of open, centralised or federated, verifiable metadata directories.

These directories persistently bind the rightsholder preferences to the unique identifier of the media asset (ISCC Codes) – and 'persistently' means: the data cannot be separated or removed. Federated directories have to be publicly accessible to discover ISCC codes, resolve associated rightsholder preferences and verify the authenticity and originality of the declarations.

> "Looking at the current landscape of unit-based identifiers, an approach based on a content derived identifier such as the ISCC to identify opted-out works and record opt-outs via the proposed standardised vocabulary seems viable for at least some categories of works. Such a registry would soft-bind opt-out declarations based on the standardised vocabulary to ISCC codes. This would allow AI model trainers to use ISCC codes as a look-up key to check the registry for known opt-outs."[8]

Since anyone can make a public declaration, it is important to ensure proper authentication of the source of each declaration. To increase trustworthiness, it is suggested that declaration metadata will include publicly accessible verifiable credentials (VCs), which are based on the W3C standards for Verifiable Credentials, supported by advanced and qualified certificates that properly identify creators and rightsholders. These "Creator Credentials" serve as a means for attribution and authentication of creators and rightsholders based on social or institutional authentication, thereby increasing trust in claims and attribution. Creator Credentials provide creators and rightsholders with a sovereign, portable and interoperable way of managing their digital identities.

All declarations are digitally signed and provided with verifiable timestamps to ensure their accuracy, validity and transparency as to when exactly the opt-out declaration was published – an aspect often overlooked in the current discussion about opt-out declarations.

[8] Open Future Foundation, Considerations For Opt-out Compliance Policies by AI model Developers, https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf

Options for Opt-out Declarations

Recital 105 of the European AI Act not only refers back to the inversion of copyright and the exceptions of the DSM Directive on text and data mining (TDM), but also highlights a major concern of the creative communities: the potential loss of control over their works and commercial opportunities in the field of generative AI:

> "General-purpose AI models, in particular large generative AI models, capable of generating text, images, and other content, present unique innovation opportunities but also challenges to artists, authors, and other creators and the way their creative content is created, distributed, used and consumed. The development and training of such models require access to vast amounts of text, images, videos and other data. Text and data mining techniques may be used extensively in this context for the retrieval and analysis of such content, which may be protected by copyright and related rights. Any use of copyright protected content requires the authorisation of the rightsholder concerned unless relevant copyright exceptions and limitations apply. Directive (EU) 2019/790 introduced exceptions and limitations allowing reproductions and extractions of works or other subject matter, for the purpose of text and data mining, under certain conditions. Under these rules, rightsholders may choose to reserve their rights over their works or other subject matter to prevent text and data mining, unless this is done for the purposes of scientific research. Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works."[9]

Creators and rightsholders worldwide are currently trying to shape how and under what conditions their copyrighted works are used for generative AI training. However, there is an urgent need for stakeholders in the creative and technology industries to develop a clear vocabulary for rights expressions that defines the terms and implications of restricting TDM.

The key question of this vocabulary is whether most creators or rightsholders really want to opt-out of TDM. While this discussion is beyond our scope, it is important to recognise that TDM, especially with the inclusion of AI and large language models, supports many useful applications. Opting-out could lead to unintended or undesirable effects for creators and rightsholders, potentially limiting the utility of their works.

Exploring the options for opting-out of TDM reveals three distinct approaches:

- Firstly, there is a comprehensive opt-out from TDM that excludes content <u>from all TDM activities</u>, with the exception of scientific research, where an opt-out is not possible. This overreaching option prevents any TDM that may or may not include AI training, thereby offering the most general form of protection for content.
- Secondly, the opt-out of TDM <u>for AI training purposes</u> is specifically aimed at excluding content from TDM activities intended for training models in scenarios involving AI technologies. While this allows general use of TDM, it restricts any

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

application where TDM is used to develop AI technologies.

- The third approach is to opt-out of TDM for use in generative AI only. This option is about preventing content from being used in TDM activities that train generative AI models and applications, or in applications that use generative AI to create synthetic audio, images, videos or text. It allows content to be used in other AI-driven applications, but prevents it from being used to generate new, synthetic content.

TDM·AI aims to create a communication protocol that can help articulate the preferences of rightsholders, with a focus on opting out of TDM for generative AI applications. This approach addresses a pressing problem, is simple to use and provides clarity in terms of scope, making it an important tool for rightsholders who want to retain control over how their content is used in a rapidly evolving digital environment.

References

TDM·AI Protocol Homepage, https://tdmai.org.

Visual presentation of the TDM·AI protocol, Google Slides (can be updated)

Overview TDM·AI Protocol, Google Drawing

1. Liccium. "Homepage." Liccium, https://liccium.com.

2. Liccium. "Creator Credentials Project Website." Creator Credentials, https://docs.creatorcredentials.com.

3. ISCC Foundation. "Homepage." ISCC Foundation, https://iscc.io.

4. ISCC Foundation. "Coded & Algorithms." ISCC Foundation, https://core.iscc.codes.

5. ISCC Foundation. "GitHub." GitHub, https://github.com/iscc.

6. ISO. "ISO 24138:2024 – International Standard Content Code." ISO, https://www.iso.org/standard/77899.html.

7. W3C. "Verifiable Credentials Data Model v2.0." W3C, https://www.w3.org/TR/vc-data-model-2.0/.

8. Open Future Foundation. "Considerations for Opt-out Compliance Policies by AI Model Developers." 24 May 2024, https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf.

9. "Robots Exclusion Protocol, RFC 9309." IETF, https://datatracker.ietf.org/doc/html/rfc9309.

10. "TDM Reservation Protocol (TDMRep), Final Community Group Report." W3C, 2 Feb. 2024, https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/.

11. Coalition for Content Provenance and Authenticity (C2PA). "Homepage." C2PA, https://c2pa.org/.

12. Creator Assertions Working Group. "Training and Data Mining Assertion." Version 1.1 Draft, https://creator-assertions.github.io/training-and-data-mining/1.1-draft/.

13. "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence." EUR-Lex, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.
14. "Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC." EUR-Lex, https://eur-lex.europa.eu/eli/dir/2019/790/oj.

About The Authors

Sabine Richly

Sabine Richly is a German Qualified Lawyer (LL.M., MBA) with a focus on media and copyright law, particularly in the context of EU regulations. She offers 15+ years of experience working with media companies, collective management organisations, and creative artists, providing guidance on all aspects of digital rights management and developing AI copyright policy strategies as a consultant for national and international organisations.

Email: sabine.richly@medialaw.digital
Homepage: https://medialaw.digital
LinkedIn: https://www.linkedin.com/in/sabine-richly-6ba49322/

Sebastian Posth

Sebastian Posth is an entrepreneur specialising in digital innovation projects in the cultural and creative industries. Sebastian is the founder and CEO of Liccium B.V. and initiator of the Creator Credentials project. He is co-initiator of ISCC (ISO 24138:2024) and has been the convener of its ISO working group (ISO TC 46/SC 9/WG 18). As a co-founder of ISCC Foundation (NL), Sebastian was a member of the Board of Directors from 2019 to 2024. During the same time, he was a research assistant at the Institute for Internet Security at the Westphalian University of Applied Sciences.

Email: sebastian@liccium.com
Homepage: https://posth.me/about-me/
LinkedIn: https://www.linkedin.com/in/posth

About Liccium

Liccium B.V. (NL) offers creators and rightsholders an easy-to-use platform to digitally sign their original creative works, providing trust in the claims, attribution and authenticity of their digital content. It supports both asset- and unit-based approaches to express rightsholders' preferences, the TDM-AI protocol and C2PA. Liccium is currently testing the first implementation of an open and verifiable opt-out directory.

Homepage: https://liccium.com