
# Open Binding of IDs to Audiovisual Essence

*Creating an open method of identifying content across multiple platforms*

SMPTE 24TB Open ID Study Group Report
30 June 2014

Society of Motion Picture and Television Engineers®
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

# Contents

# 1 Introduction

This effort had its origins within the Coalition for Innovative Media Management (CIMM). CIMM is a group of buyers and sellers of advertising-supported media formed to promote innovation and explore new, high-quality ways to measure audiences across traditional and new media in the United States. CIMM began in 2009, founded by leading television content providers, media agencies, and advertisers.

One of the keys to enable cross-platform measurement of content viewing was found to be the ability to embed registered media identifiers (e.g., EIDR for program material and Ad-ID for commercial material) into audiovisual essence in such a way that they survive processing and distribution across a variety of distribution platforms while remaining imperceptible.

CIMM and several of its media and technology members engaged SMPTE to explore how such embedding of identifiers into media might be accomplished and to develop any standard(s) required to enable such embedding to be carried out in an open way.

As a result, SMPTE's Technology Committee on Television and Broadband Media (24TB) formed a Study Group, focused on "Open Binding of IDs to Essence." The task of this Study Group was to create a report on the viability of and recommendations for an open standard to bind unique media identifiers (i.e., EIDR and Ad-ID) to content's essence. This binding must survive compression and other transformations along the entire chain, including distribution and delivery to consumers.

The problem that the Study Group effort aims to solve is stated in its project overview document and reads as follows:

> Currently, there is no open binding technology standard (e.g., watermarks, fingerprints, metadata sidecars, etc.) for embedding persistent content identifiers into audio/video essence in a way that survives compression and distribution through the media and entertainment (M&E) supply chain.
>
> Content creators and distributors need an effective way of reliably binding content identifiers to video/audio content that will robustly transit an end-to-end media ecosystem. Unique content identifiers, such as EIDR, Ad-ID – and others – have been important developments. However, within the supply chain, M&E entities "bind" the identifiers in file containers to data streams in a file container or to the structural metadata of the video/audio stream itself. These types of bindings can be destroyed in video/audio processing and delivery systems along the media production, distribution and measurement supply chain.
>
> An open standard for essence-based video/audio binding can enable a wide range of new capabilities. Most importantly it will enable increased speed, transparency and accountability in video content and advertising measurement across a wide range of delivery systems and devices. Additional benefits are improved media workflow automation within and between M&E entities, fewer barriers to deploying cross-platform

dynamic ad insertion, enablement of new anti-piracy tools and methods and broader digital locker adoption, more complete long-tail content monetization, improved accuracy in automatic content recognition and detection, better second-screen integration and improved multi-screen content discovery … the list goes on.

The Study Group proceeded to document Use Cases and, from there, technical requirements. A Request for Information (RFI) ultimately was issued, the output of which is the basis for this report.

For the purposes of this document, the term “watermark” refers to any technology that conveys hidden information within program material essence. The term “fingerprint” refers to a compact representation of program material essence characteristics, popular in the realm of automatic content recognition (ACR) technology. Please see Terms and Acronyms. This report also employs the term “metadata sidecars” to describe a broad category of data stream identification systems that are not necessarily dependent upon the processing of picture or sound essence. These systems could include those described in ANSI/SCTE 35:2012 “Digital Program Insertion Cueing Message for Cable,” SMPTE ST 12-1:2014 “Time and Control Code,” and many others.

# 2 Scope

This report investigates the feasibility of future standards development work on an open method for binding media IDs to essence.

# 3 Executive Summary

Using responses collected in an RFI issued by the Study Group, this report concludes that technologies exist today for binding media identifiers such as Ad-ID and EIDR to content. Such technologies are appropriate for documentation in an open standard and can survive most or all of the processing that content endures on its way to the consumer over an increasing variety of distribution channels.

Audio watermarking appears to be the most promising approach. It was the most frequently proposed technology solution across RFI responses and is claimed to satisfy the use cases and requirements identified by the Study Group.

The Study Group on Open ID Binding to Essence recommends that SMPTE undertake standardization of technologies for the open binding of IDs to essence, using the findings in this report as guidelines.

# 4 Technology Requirements and Features

A Request For Information (RFI) was issued to gather technical information on technologies appropriate for binding media identifiers such as Ad-ID and EIDR to content. Following is a summary of the requirements specified in the RFI. Associated with some of the requirements are additional findings that the Study Group discovered during the course of its work, and those findings are noted along with each requirement and labeled “Finding.”

## 4.1 Mandatory Requirements

The following requirements must be supported in an overall solution.

### 4.1.1 Payload

The binding method must accommodate a payload of a minimum of 25 bytes (200 bits) to carry simultaneously Ad-ID and EIDR (each of which is 96 bits long), along with indicator(s) (such as enumerated values) to label each ID, plus overhead.

**Finding**
Both Ad-ID and EIDR can accommodate more compact representations under certain conditions (see Section 9.2 Payload Management), allowing finer granularity (see Section 4.1.4) and/or increased survivability (see Section 4.1.2).

### 4.1.2 Survivability across Platforms

The binding technique is required to ensure that IDs survive all common distribution platforms and all associated processing.

Common distribution platforms include, but are not limited to, terrestrial broadcast, cable/satellite networks, IP (managed or open) networks, and mobile networks. Ad hoc distribution in consumer applications (e.g., AirPlay, WiDi, Chromecast, etc.) that access either file-based or streaming content also should be included.

Processing may include resolution change, scaling, transcoding, frame/bit-rate conversion, audio sampling rate change, audio up/down mix, A-to-D/D-to-A, audio spectral band replication, and the like. Proponents were asked to identify any processing and/or distribution platforms that their proposed solutions may not support.

It is not anticipated that geometric transformation and/or distortion, or partial/full overlay of video imagery will be accommodated. Proponents were encouraged, however, to state any additional types of processing that a particular proposed solution supports.

Some of the transformations that are expected to be addressed are those between video codecs such as WMV, JPEG2000, H.264, MPEG-2, VP8, VP9, and HEVC. On the audio side, transformations between codecs such as WMA, MPEG 1 Layer 2, AC-3, E-AC-3, Dolby E, AAC, HE-AAC, HE-AAC v2, and MPEG-H should be considered.

**Finding**
Resilience of the binding technique against deliberate removal or obfuscation is not considered a requirement.

### 4.1.3 Recording and Playback

The binding technique is required to ensure that IDs survive recording and replay in the professional and consumer spaces. Proponents were requested to state clearly all known limitations to their proposed solutions, including recording quality metrics such as resolution, bit rates, and/or audio sampling rates.

### 4.1.4 Granularity

This requirement was revised after analysis of the RFI responses to clarify its role in measurement, as defined in the Terms & Acronyms section of this document.

> ”The binding mechanism shall allow the transition to or from a uniquely identified or unidentified piece of content to be detected within 1 second of the transition. Transition detection does not necessarily require the recovery of the content ID.”

**Finding**
This requirement is not sufficient to satisfy proof-of-performance requirements, for which sub-second granularity may be required. Proof-of-performance was not listed as a requirement in the RFI and, as such, is out of scope of this report. Nevertheless, it was of interest to several members of the group and so is noted in this report.

### 4.1.5 Retrievability

Identifiers are required to be recoverable from any continuous three-second segment of content in the case of commercials or from any continuous five-second segment of content in the case of program material.

**Finding**
This requirement was revised to reflect (a) the payload capabilities of watermarking algorithms submitted as responses to the RFI, (b) a potentially more efficient representation of Ad-ID (see Section 9.2 Payload Management), and (c) different use cases for Ad-ID and EIDR (see Use Cases). The latter is primarily used for identification of entire programs.

### 4.1.6 Synchronization of Media IDs Across Platforms

The binding technique must enable retrieving media IDs across multiple consumer delivery platforms at the same time and matching content simultaneously (given constraints of latency across delivery platforms).

### 4.1.7 ID Replacement

IDs must be removable or capable of being overwritten in cases in which media with an ID is re-aired, used within a new work, or in similar situations.

**Finding** RFI responses indicate that the options for removing a watermark are limited. This limitation should not be confused with reversibility. For some of the audio watermarking technologies that already are in wide use, reversibility is impossible, in the sense that perfect restoration and/or recreation of the original audio content cannot be achieved once these watermarks have been applied. Because emission codecs introduce material changes to any watermark, the goal of watermark removal is problematic. Instead, an undesirable watermark inserted upstream could be overwritten or retained but superseded by the insertion of an additional watermark.

### 4.1.8 Insertion and Detection Points

The adopted method is required to allow IDs to be inserted and detected at any point between production and display to consumers. The binding technique must be suitable for documentation as an open standard that allows insertion and detection to be independently implemented.

### 4.1.9 Codecs and Containers

The ID binding should be able to be transported by common codec and container formats.

**Finding**
The adoption of common audiovisual ID schemes across popular media containers further facilitates the deployment of open binding of IDs to audiovisual essence. This might include, but would not be limited to, the definition of ISO MPEG-4 and MXF metadata designated for the carriage of EIDR and Ad-ID. It is required that the ID binding be able to be carried in codec and container formats such as these.

### 4.1.10 Audio and Video Content

IDs must be recoverable from material that does not necessarily have both audio and video present.

### 4.1.11 Quality

ID transport is required not to degrade the quality of audio and/or video, including lip sync, at the consumer's reception point.

**Finding**
Widespread use of overlapping watermarks due to broadcast workflows might, however, result in a tradeoff between survivability and quality.

## 4.2 Requirements Specific to Watermarks

The following requirements were recognized during consideration of the proposals submitted.

Watermarks are one technology that the Study Group has evaluated in its investigations. Insertion of a watermark modifies the essence being transmitted. Consequently, watermark technologies must meet certain requirements that do not apply to other binding methods:

- Any watermark used must be open and identifiable. A standard must provide sufficient specificity and detail for multiple parties to create implementations of both insertion and detection of watermarks that are interoperable. API-only documentation would not provide a sufficient solution.
- ID binding watermarks cannot interfere with any existing watermarks (e.g., Nielsen TV and Nielsen Audio) and vice-versa.
- There must be both linear and non-linear (file based) methods for insertion of watermarks.

## 4.3 Desirable Features

The following items are not required in a proposed solution but are desirable.

### 4.3.1 Coexistence

Coexistence is defined as the ability to not interfere with other binding techniques, such as copy control watermarks, forensic watermarks, audience measurement watermarks, and content management watermarks. It may be desirable to have a registry of the characteristics of systems in use, for purposes of enabling this type of non-interference.

### 4.3.2 Performance Impact

Processing needed for detecting IDs should not cause noticeable performance degradation on typical handheld consumer devices.

# 5 Technology Summary

## 5.1 Watermarking

Watermarking embeds a content ID into audio or video content by altering the latter. It provides a positive and definite identification of the content and allows two distinct programs with similar or identical audio or video content to be differentiated. Furthermore, extraction of a complete content ID is possible using a standalone detector without the need for an online lookup system.

This strong identification comes with the need to insert the watermark *a priori*, and the size of the payload that can be inserted is a tradeoff with robustness (the resilience of the watermark to alterations made to the essence) and transparency (the property that the watermark cannot be perceived by the audience). An additional limitation is that successive insertion of watermarks into already marked content can cause conflicts between watermarks, making watermark detection and watermark payload recovery difficult or, in extreme cases, impossible.

Deliberate overlapping of multiple audio watermarks is possible and currently practiced in some existing workflows. Such overlapping results in decreased quality (as defined in requirement 4.1.11) and potential conflicts if different technologies are used. With a few exceptions, e.g., sports highlights programs, the use cases considered in preparing this report do not require the injection of additional data at various points throughout the transmission path, and the overlapping of multiple audio watermarks is not recommended unless required.

## 5.2 Fingerprinting

Fingerprinting operates without modification of audio or video content. Prior to distribution, a fingerprint of the content is computed and added to a fingerprint database, where it is linked with the one or more IDs assigned to the content. For detection at playback, a detector computes a fingerprint from the audio and/or video content. This fingerprint then is matched against the fingerprint database. A positive match returns the content ID associated with the content. Recovering this content ID typically cannot be done by the end-user terminal alone, due to the size of the fingerprint database, and instead requires access to an external, global fingerprint database.

Since fingerprinting does not modify content, simultaneous use of multiple fingerprints does not create conflicts and does not impact quality, and a program can be added to the fingerprint database after the program has been distributed.

An inherent limitation of fingerprinting is its inability to distinguish between two programs with substantially identical audiovisual essence.

For instance, multiple programs might share the same ratings graphics (e.g. TV-MA), and multiple episodes of the same show might share the same introduction sequence. In these cases, detection time for fingerprinting will be on the order of the duration of the material in common.

The same limitation applies to commercials when the same "generic" commercial (i.e., the same video and/or audio) is customized, or "tagged," for local dealers or offers and, from a practical standpoint, these customized versions are assigned unique Ad-ID codes. Similarly, a commercial with an audio tag on the end of it won't be differentiable from the same commercial without that tag using only video fingerprinting technology. Even using audio fingerprinting technology, the difference won't be detectable until the tag begins to be played out (at the end of the commercial). Conversely, a commercial with graphical elements or text added to it (such as a 1-800 number) won't be differentiable from the same commercial without these elements using only audio fingerprinting and only will be differentiable using video fingerprinting at the point in the commercial where these graphics appear.

Application of fingerprinting to content synchronization is being standardized by the SMPTE 24TB Technology Committee. An evaluation of the algorithm used in the 24TB lip sync specification is available at *SMPTE Audio to Video Synchronization Measurements: Fingerprint* (ST 2064-1) once it is published, in the SMPTE Store.

## 5.3 Ad-ID and EIDR Payload Management

The shortest self-contained representation of EIDR Identifiers is the compact binary representation, which is 96 bits long. See SMPTE RP 2079 or http://eidr.org/documents/EIDR_ID_Format_v1.1.pdf).

If HTTP requests over the Internet are available, the shortDOI Service from the International DOI Foundation can be used effectively to shorten an EIDR to 40 bits (as TinyURL does for URLs), with a small risk of requiring additional bits in the future.

Ad-ID's compact identifier (CID) is a 32-bit field and the primary key for the database record, which can be used as a compact alias for the full Ad-ID for the purpose of binding an identifier to content.

# 6 Summary of RFI Responses

The following summarizes the watermarking and fingerprinting described in the RFI responses listed in Annex A. SMPTE did not evaluate the performance and features claimed by the respective RFI respondents.

## 6.1 Watermarking

Six watermarking technologies (2 video-based and 4 audio-based) were described in RFI responses. These technologies are stated to offer a granularity of one second and higher, and a payload on the order of 50 bits per second.

## 6.2 Fingerprinting

One fingerprinting technology was described in an RFI response. This technology is stated to offer a granularity on the order of seconds.

# 7 Conclusions and Recommendations

The Study Group considered requirements for standardization of the binding of media identifiers such as Ad-ID and EIDR to content, with the binding surviving all of the processing that content endures on its way to consumers over a variety of distribution channels. Based on the responses received to the RFI, it appears that the requirements considered by the Study Group may be satisfied by existing technologies. In particular, audio watermarking is potentially sufficient for binding media identifiers such as Ad-ID and EIDR to content. If it turns out that audio watermarking is insufficient, then a combination of other technologies may be considered. Any potential solution must be validated by testing.

The Study Group recommends that robustness testing be implemented to validate the viability of the Requirements in terms of reasonable implementations.

The Study Group therefore believes that SMPTE should undertake the standardization of technologies for the open binding of IDs to essence, using this report as input in the following manner:

The Study Group recommends that a request for proposals (RFP) be issued, soliciting proposed technologies for standardization

Whatever solution is adopted must be fully documented in the standard and be independently implementable by users of the standard.

# 8 References

The following documents are referenced within this document and may be helpful to the reader if additional detail is desired.


- EIDR ID Format, V 1.1, Entertainment Identifier Registry, 2013
- SMPTE RDD-17:2009 Advertising Digital Identification, Society of Motion Picture and Television Engineers, 2009
- SMPTE ST 12-1:2014 Time and Control Code, Society of Motion Picture and Television Engineers, 2014
- SMPTE RP 2079:2013 DOI Name and EIDR Identifier Representation, , Society of Motion Picture and Television Engineers, 2013
- CEA CEA-608-E:2008-04 Line 21 Data Services, Consumer Electronics Association, 2008
- CEA CEA-708-E (ANSI):2013-08 Digital Television (DTV) Closed Captioning, Consumer Electronics Association, 2013

# 9 Terms and Acronyms

**A-to-D/D-to-A (Analog-to-Digital / Digital to Analog)** - The conversion of analog essence to digital form, or vice-versa. The result is a sequence of digital values that have been converted from a continuous-time and continuous-amplitude analog signal to a discrete-time and discrete-amplitude digital signal and the reverse.

**Ad-ID** - An advertising industry standard unique identifier for commercial assets across all media platforms. Ad-ID codes are 11 or 12 alphanumeric characters. The first four characters are company identification prefixes. Ad-ID is a joint venture of the American Association of Advertising Agencies (4A's) and the Association of National Advertisers, Inc. (ANA), and is currently defined in SMPTE RDD-17-2009.

**Audio sampling rate -** The number of samples of audio carried per second, measured in Hz or kHz (one kHz being 1,000 Hz). For example, 44,100 samples per second can be expressed as either 44,100 Hz, or 44.1 kHz.

**Bit rate** - The number of bits that are conveyed or processed per unit of time.

**Codec / video codec -** A device or software that performs compression or decompression (co/dec) of digital audio or video. Codec programs are necessary for a media player (software or device) to play music and movies. The purpose of a video codec is to represent moving images in an efficient manner. The most common digital video codecs utilized as of this writing for broadband video transmission and playback on the various Internet-connected devices include MPEG-2, H.263 / MPEG-4 part 2, H.264 / MPEG-4 AVC, and VP6.

**Coexistence -** Coexistence is defined as the ability to not interfere with other binding techniques, such as copy control watermarks, forensic watermarks, audience measurement watermarks, and content management watermarks. It may be desirable to have a registry of the characteristics of systems in use, for purposes of enabling this type of non-interference.

**Compression -** The process by which files of data, audio, or video content are compressed / reduced in size to facilitate fast transmission and to require less storage space.

**Container** – Used to package, transport, and present a variety of codecs. Some examples of containers include MP4 (MPEG-4 Part 14), QuickTime, MPEG-2 Transport Stream, Flash, etc.).

**Digital Object Identifier** - The Digital Object Identifier (DOI) is a character string used to uniquely identify an object such as an electronic document. ISO 26324:2012 specifies the syntax, description, and resolution functional components of the DOI system and the general principles for the creation, registration, and administration of DOI names.

**Downcut** - A technical error in which playback does not end at the proper “out” point of a clip, but instead playback ends at a point before the proper “out” point. This results in the last part of the desired content not being transmitted.

**Dynamic ad insertion (DAI)** - The process by which a targeted ad is inserted in real time into content.

**EIDR** - The Entertainment Identifier Registry, a global registry for universally unique identifiers for movie and television assets. EIDR is designed to allow for automated machine-to-machine communications and has a flexible data hierarchy down to the product and SKU level, including edits, composites, encodings, and relationships. EIDR is based on the ISO Digital Object Identifier (DOI) Standard, and its representation is specified in SMPTE RP 2079. EIDR's registry services are provided by the Entertainment Identifier Registry Association.

**Essence** - The raw video, audio, and data (such as closed captioning) streams that are intended to be perceived by end users.

**Fingerprint** - A compact representation of essence characteristics that can be used to identify that essence. Typically fingerprints contain far less data than is in the original essence. Some types of fingerprints can continue to identify essence even when that essence has mild impairments (such as compression, changes in sampling, or changes in resolution) and transcoding to other codecs.

**Granularity** – The time required to be able to detect the transition between two pieces of content.

**Latency** - The time delay experienced in a system.

**Measurement –** The tracking of viewership of a piece of content.

**Metadata** - Data that describes and/or provides meaning to other data.

**M&E** - Media and Entertainment.

**Professional media systems** - Systems operated by professionals to acquire, manipulate, edit, process, and distribute media.

**Professional media network** - A network infrastructure to support some or all of the activities of a professional media system.

**Proof-of-performance –** The reporting of the airing of content by a broadcaster or network, typically back to the owner/creator of the content (i.e., ad agency or program distributor).

**Retrievability** – The ability to extract an ID from a piece of content.

**Sidecar** - A broad category of data stream identification and other metadata that are transmitted in such a way that they can be processed independently of the video or audio essence.

**TinyURL -** A URL shortening service; a web service that provides short aliases for redirection of long URLs.

**Upcut** - A technical error in which playback does not begin from the proper “in” point of a clip, but instead playback begins from a point after the proper “in” point. This results in the first part of the desired content not being transmitted.

**Watermark** - A covert signal added to essence that does not cause noticeable artifacts in the essence and is imperceptible to viewers. Some types of watermarks allow watermark signals to survive mild impairments of the essence (such as compression, changes in sampling, or changes in resolution).

Society of Motion Picture and Television Engineers®
3 Barker Avenue
White Plains, NY 10601 USA
www.smpte.org

# 10 Annex A – RFI Responses

In response to the RFI generated by the Study Group, a wide array of responses was received. Those responses and supporting documentation can be found in the 24TB Open ID Study Group's Kavi document repository. For access to this material, contact the SMPTE Director of Standards & Engineering.