

# Automating and Simplifying Multiparty Workflows

Richard W. Kroon  
François Modarresse

Cite this article:

Kroon, Richard W., Modarresse, François; 2018. Automating and Simplifying Multiparty Workflows. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2018.10. Web Link: <http://dx.doi.org/10.18580/setijbe.2018.10>

# Automating and Simplifying Multiparty Workflows

Richard W. Kroon, Director of Engineering, Entertainment Identifier Registry Association (EIDR)  
François Modarresse, Business Development, EIDR

**Abstract** - Any broadcast organization that remains static runs the risk of being overtaken by newer, more agile alternatives. To remain competitive, broadcasters must constantly work to increase process velocity, accuracy, and flexibility. These goals cannot be reached without reducing time to market, manual touch-points, and associated labor costs. A major hurdle on this road to efficiency is the absence of a universal method to identify content, resulting in unnecessary manual workflows and time- and resource-consuming communications with third parties for the production, processing, and exchange of content. Root causes for these impracticalities include problems with work identification during acquisition, reconciliation, and de-duplication of assets obtained from multiple sources; placing high demands on limited resources; and causing delays or reducing content capacity. A necessary element to solve this problem is the use of globally unique and persistent works identification. As such, it will open the door to quasi-instantaneous, metadata based, machine-to-machine processes. A well-designed ID will include content and video service identifiers. Its architecture will also allow the embedding of other IDs such as pre-existing, in-house, local, application-specific, third-party, or content-rich commercial identifiers. Whether for feature films, episodic series, or even sports, significant benefits can be found by leveraging a unique ID recognized throughout the workflow and by all partners from content origination to processing, enrichment, storage, distribution, measurement, and attribution. In this paper, we explain the desired structure of identifiers plus their positive impact on workflows and speed. Finally, use case examples will demonstrate real-world applications.

## CONTENT MANAGEMENT CHALLENGES IN BROADCASTING

Broadcasting is an unforgiving, real-time environment where the show must always go on, and the VOD or EST catalogs must always remain well stocked – dead air is simply not an acceptable option. Operational considerations abound: asset management for both acquired and locally-produced content, collating and presenting guide data, audience measurement, language versioning, VOD and ancillary digital distribution channels beyond initial broadcast, content archiving and retrieval, contract compliance, etc.

Live event broadcasting raises the ante by eliminating the delay between content creation and public presentation. Sports broadcasting is more challenging still, combining all the normal broadcasting attention points with the immediacy of live event programming (variability in start times and durations, weather and other delays, groups as participants, late changes to participants, etc.) and the special nature of sports programming (local blackouts, regional presentations, multiple productions of the same competition, frequent repetition of past matchups, frequent reuse of clips, sharing clips and content among broadcast competitors, etc.).

## Why Identify?

None of the above is practical without a means to identify content as it moves along the production, storage, modification, and distribution chain.

This is just a modern example of the problem of naming things. If there is no name for something, you cannot talk about it. If you are working with someone else, the two parties must have either a shared name or a way of translating each other's names. If neither of those is present, no exchange of information is possible.

To remain financially competitive, all broadcasters must undertake continuous process improvement by increasing process velocity, accuracy, and flexibility while reducing time to market, manual touchpoints, and associated labor costs. One of the most effective ways to do this is by automating repeatable processes. Often, a necessary first step is standardizing manual processes so that they become amenable to automation. Along the way, organizations must also give up a certain amount of proprietary customization in the name of standardization, automation, and overall process efficiency.

Any asset identification scheme that is clearly defined and consistently applied can be used within a broadcast organization. In fact, organizations tend to have several different identification schemes in play at any given time, including separate IDs for accounting, production, scheduling, etc.

## The Need for a Universal Identifier

However, the limitations of these proprietary constructs, developed within broadcast organizations over time, are often not apparent until one attempts to improve workflows that include an external input or output.

For workflows related to content management, these external interfaces almost always include manual touchpoints for title matching. Traditionally, when two parties in the broadcast ecosystem have communicated about a work of common interest, they have identified the work by exchanging descriptive metadata along with any contract, query, asset, or transaction and then relied on manual labor to match everything together. Unfortunately, communicating parties do not always agree on what these descriptive data are or how they are structured – even titles may differ, especially when abbreviated to fit within arbitrary space limitations or when dealing with international releases or foreign works. Thus, record matching has been a long-standing challenge and remains an ongoing expense.

In this light, internal identifiers are of limited use when communicating with third parties. Organizations can agree on an identification scheme with each of their supply chain

partners, but this leads to an explosion of point-to-point identifiers. In theory, there could be 10 different identifiers per title in a 5-party ecosystem, though the actual number is lower thanks to the use of commonly exchanged IDs, usually those of a dominant partner. Regardless of the number, the receiving parties must still manually match each received ID. This could lead to five matching efforts per title in this example.

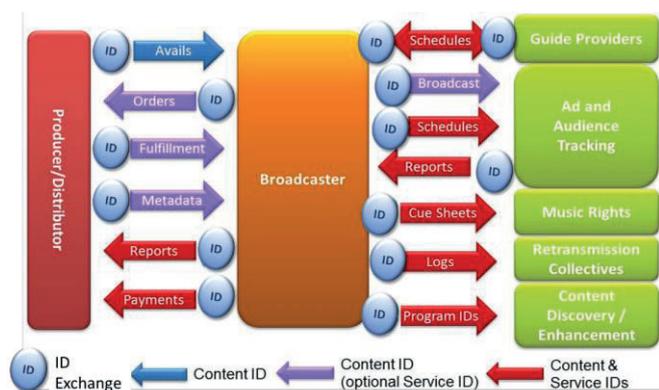


FIG 1: ID Uses in Workflows with External Touchpoints

These manual touch points must be removed if the benefits of automation are to be realized.

Using a common, global identifier for these third-party interactions reduces the number of IDs in play to one, regardless of how many partners are involved in the Exchange<sup>1</sup>, and in turn facilitates a greater automation of processes while accelerating time to market.

An independent industry services company reviewed one such use case, where several major broadcasters (ITV in the UK, ABC Television Group and NBC in the US plus Sony Entertainment Networks) and studios (Universal and Warner Bros.) reviewed the opportunities and challenges of using a universal ID in broadcast workflows. With full implementation, they anticipated an annual savings of US\$26M (i.e. 79%) in content identification and management costs<sup>2</sup>.

### Desirable Attributes for IDs

Useful identifiers for multi-party media workflows possess several key characteristics:

- **They are globally unique** – a particular identifier resolves to a particular object. That object could be a single asset or a collection of assets, but whatever it is, it the association between the identifier and the identified never changes.
- **They are permanent** – once assigned, the identifier never goes away. It is always available for use in reference to the identified asset.
- **They are not proprietary** – they may be issued by a controlling organization, but once issued, there is no restriction on their use. Anyone can use them at any time in any workflow.
- **They are large** – the ID space must have sufficient capacity to identify all the different assets that might conceivably appear over time.

- **They are resolvable** – there is a mechanism where the ID can be converted into a description of what it identifies (and ideally the reverse, where you can find the ID via its descriptive metadata). You cannot have a shared identification system without an open and accessible means of sharing the relationship between the identifier and the identified.

If an identifier fails in any one of these dimensions, then any workflow that depends upon it will eventually fail itself. Implementing a shared identification system is not without cost, so the parties in a supply chain should take great care in the selection of their identification scheme to make sure that it will continue to serve their needs far into the future.

### Classes of ID

There is a wide variety of identification schemes at play in the media industry, each with its advantages and adherents. The most common options for shared identification include.<sup>3,4</sup>

- **Shelf Numbers (for physical assets) and Directory Paths (for digital assets)**  
 These are most useful for locating an asset once it has been identified but are not useful as identifiers themselves. They are only meaningful within the context of a particular storage system and are not durably linked to the asset being described – if the asset is relocated, then its shelf number/path is changed.
- **Classification Systems**  
 These are most commonly found in libraries<sup>5</sup>. The most widely used of these is the Dewey Decimal Classification system<sup>6</sup>. They were first applied to print works and have since been extended to include audiovisual assets. They are “intelligent” numbers in that meaning may be derived by parsing the classification identifier according to a predefined formula. By their very nature, they include subjectivity in their assignment, so the same work may receive different classifications in different collections even when using the same classification system.
- **Proprietary Inventory Indexing**  
 These systems consist of a simple identifier assigned by whoever holds an asset. They are easy to create, are unique within their domain, and are permanently associated with an asset. These are popular with media archives, such as the British Film Institute<sup>7</sup>, and are the sort of identifier most commonly associated with a digital asset management system (DAM). Since they are only unique within their domain, the same ID could be assigned by different parties to reference different things and so are not globally unique. They can be used as point-to-point identifiers between specific parties, but this does not scale well.
- **Statistically Unique Identifiers**  
 This includes things such as UUIDs (Universally Unique ID)<sup>8</sup>, UMIDs (Unique Material ID)<sup>9</sup>, and file hashes (C4, MD5, SHA-2, etc.).<sup>10,11,12</sup> They can be generated by anyone at any time in such a way that they are globally unique for all practical purposes<sup>13</sup>. The most common failing of such non-hash-based systems is that

the same asset will receive a different ID every time an ID is generated. File hashes will always be the same for a given asset no matter who creates them, but they can only be applied to digital files and a single bit's difference in a file results in an entirely different ID. Finally, there is no way to determine what is identified by a particular statistically unique identifier (they are not resolvable outside the organization that created them).

• **Shared, Curated, Globally-Unique Identifiers**

This is the ultimate media asset identification and the only type of identifier that meets all the criteria for use in multi-party media workflows. Such identifiers are issued by a controlling organization that ensures uniqueness and provides ID resolution services. The preeminent identifier of this type is the Digital Object Identifier (ISO Standard 26324) <sup>14,15,16</sup>. The Entertainment Identifier Registry (EIDR) is a DOI registration authority providing identification services for audiovisual works<sup>17</sup>.

**Desirable ID Structure**

For broadcast applications, a typical DOI-compliant content ID is organized in three fields.

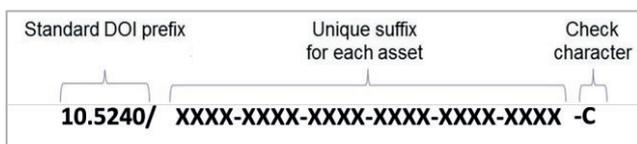


FIG 2a: DOI-Based Content ID

Such IDs are part of the Digital Object Identifier (DOI) namespace. Each DOI begins with a prefix that identifies the particular type of DOI and the organization that administers it. For audiovisual content identifiers, this is 10.5240, indicating a Content ID administered by EIDR. That is followed by a 20-character hexadecimal number, generally presented in 5 hyphen-separated groups of 4 digits for human convenience, and an alphanumeric check character. Such a number space has sufficient capacity to identify 1,208,925,819,614,630,000,000,000 unique audiovisual works, edits, and encodings, so it will remain viable for decades – if not centuries – to come:

With 1 Septillion, 208 Sextillion, 925 Quintillion, 819 Quadrillion, 614 Trillion, 630 Billion different values, that works out to roughly 160,000 IDs for every grain of sand in the world, if this were a sand identification system.

As with all DOIs, this is a “dumb,” or opaque, number. Unlike a cataloging system, no intelligence can be derived from the ID itself other than the fact that it is an ID of a given family as determined by the prefix. The ID associated with a work in the abstract (a “title” ID) looks just like the ID for a specific encoding of a particular edit of that work. The only way to know for sure what the ID references is to resolve the ID. EIDR IDs are always “read-for-free,” and public resolution services are offered by both DOI and EIDR<sup>18</sup>. In addition, users of such IDs, such as EIDR members, offer their own, internal resolution services integrated into their proprietary workflows <sup>19</sup>.

**ID Architecture in Registries**

As pointed out earlier in this paper, the use of an ID administrator brings many benefits. A good administration authority will provide a publicly consultable registry of IDs as well as methods of resolving IDs and searching for IDs.

The right registry must contain a number of different record types, including Movies, TV programs, Radio programs, Shorts, Series, Seasons, Episodes, Web programs, Clips, Compilations, Composites, Supplemental programs, and Manifestations (i.e. technical variations such as resolution, encoding type etc.).

Most of these must be arranged in a hierarchical inheritance tree for each audiovisual work. At the root of the tree is the abstract Title record, which encompasses the referenced work in all its forms. Beneath that, one must include the Edit records that represent the different creative versions or cuts of the work. If one Edit is derived from another, the Edit relationship can be denoted by using child Edits (or Edits of Edits). Beneath the Edits are the Manifestation records that represent the different encodings or fixations of the Edits. Manifestations must have the flexibility of being quite detailed, so to enumerate each video or audio track with all their technical details.

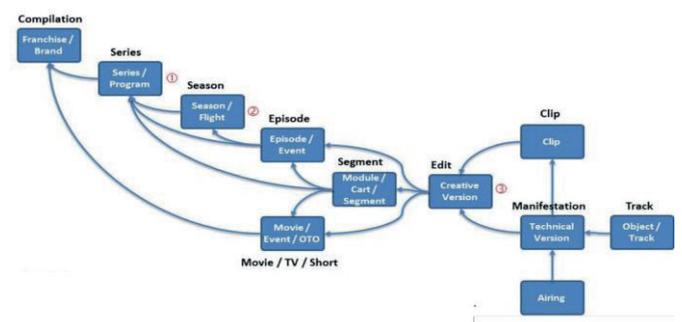


FIG 3: Broadcast Hierarchy Example for Works of Fiction

In addition to the standard hierarchical inheritance tree structure, the proper registry structure must also support a number of different non-hierarchical relationships, such as a Promotional relationship to link a trailer/teaser to the work it promotes and a Composite relationship to link a composite work (such as a “clip show”) to the source materials from which it was assembled.

**Live Programming Considerations**

News and, even more so, sports present specific challenges as depicted in the “Content Management Challenges in Broadcasting” section above. The need for clips for reciprocal exchanges with competitors down to several generation of re-editing; variable and inconsistent schedules due to weather or stoppage time, etc.; the existence of sub-events (football leagues, regional vs national cups, disciplines in track-and-field, etc.); and the atypically high number of re-runs (or Replay TV) imply the

need of additional relationship structures. Additionally, last minute program changes due to last-minute cancellations, team elimination from a knock-out round, etc. must be carried out without confusing the consumer if program guides are not properly updated.

The proper registry must include fields that provide the means to unambiguously identify works and their associated derivatives.

For instance, in the case of program information for Digital Video Recorders (DVRs), additional identifying characteristics should be included as part of a standardized registration practice, including:

- The original broadcast date (or anticipated broadcast date for a program that has not aired yet)
- Anticipated running time (useful to distinguish a full game from game highlights)
- The primary language used in the presentation (for sports, this is generally the language spoken by the commentators and used for on-screen text)
- Any alternate titles by which the program may be known (quite useful for language disambiguation and later discovery)
- Any alternate identifiers that may be available (including house production IDs and event IDs)

### The TV Network Challenge

Many Broadcast networks, Multichannel Video Programming Distributors (MVPDs) and streaming services are organized in different regions, each with their own schedules, time zones, or even languages.

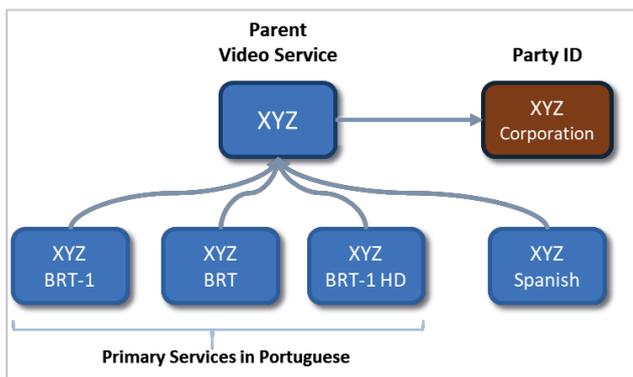


FIG 4: ID for Video Services in Networks

Complex sophisticated organization also own multiple brands or affiliates each with their own identity. The right ID and its registry will allow the inclusion of brands and station affiliations. The examples below show a structure for the Disney / ABC broadcast group.

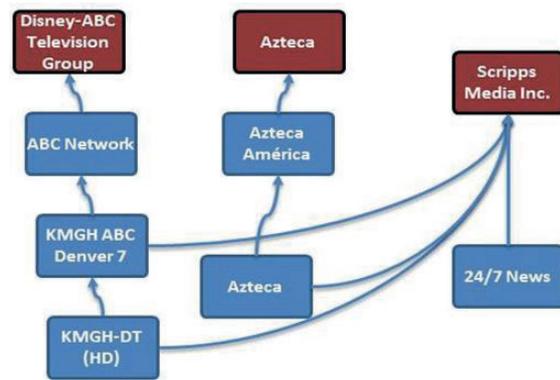
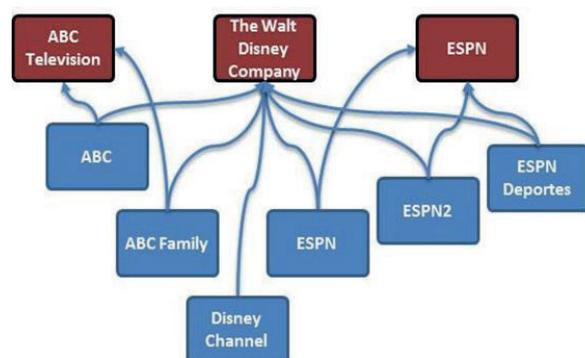


FIG 5: Example of ID for Video Services by Brand or Affiliation

All of this results in a broadcast industry need for universal video service – or channel – identification nearly as great as the need for content identification. EIDR also administers a DOI identifier for this purpose.



FIG 2b: DOI-Based Video Service ID

The EIDR Video Service ID begins with the DOI prefix 10.5239 and is followed by eight hexadecimal digits, presented as two groups of four. This has the capacity to identify 281,474,976,710,656 different content delivery services. As with EIDR Content IDs, the Video Service IDs are read-for-free including ID resolution, metadata search, and alternate ID cross-references. The individual Video Services can indicate network affiliation and content distribution relationships in addition to ownership interests tied to other Video Services or separate organizations (themselves identified by EIDR Party IDs in the form of 10.5237/XXXX-XXXX). As always, each Video Service record can contain an unlimited number of Alternate IDs (see “The Need for Nested IDs”, below), allowing the Video Service registry to act as a cross-reference service that helps to bring proprietary identification systems and supporting workflows into the global identification framework.

### The Need for Nested IDs

Broadcast organizations have often relied on legacy identifiers sometimes developed during the days of analog signals and tape recorder infrastructures.

Routinely, different stations of a broadcast network – and in some cases, different departments within an organization – have their own content identifiers. (See “Why Identify?” and “The TV Network Challenge”).

Concurrently, there is a need for application-specific identifiers used internally or with partner companies for uses such as audience measurement, human consumption (e.g. EPG, VOD thumbnails, rating etc.), and many other applications.

Therefore, a universal ID must not exist in a vacuum. The universal Registry must maintain a cross-reference service for other third-party identifiers and house identifiers from a variety of motion picture and television organizations and industry service organizations. Not only should you be able to resolve to an ID (and its associated metadata record) from one of these alternative identifiers, but you should also be given the ability to come in with one ID and go out with another.

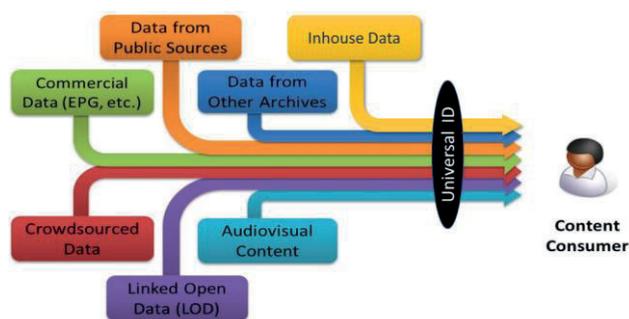


FIG 6: Universal ID as a translator to all other IDs

For example, one of EIDR’s studio members uses the EIDR Registry to translate IMDb IDs (which they already have) into Flixster IDs so they can access the Rotten Tomatoes scores for critical and consumer review ratings.

Taken as a whole, a good Registry can deliver IDs for any form of any type of audiovisual work. Once a particular object has been identified, all parties in the media ecosystem leverage that same ID for unambiguous identification and automation without further manual intervention.

By using universal IDs, a world leader in digital distribution measured a reduction of on-boarding time from 50 hours per title to 12 minutes translating into faster time to market, cost savings, and expansion of content capacity.

### Conclusion: The Positive Spiral

A few moments spent with a blank whiteboard will lead to the identification of many use cases where a universal identifier (globally unique, permanent, and universally accessible) can help reduce costs, increase accuracy, facilitate new services, and make content even more accessible and profitable. The value of universal identification grows significantly with each additional party participating in the ID ecosystem. To achieve these goals, a content identification system must have a global reach and accurately and unambiguously identify the various types of audiovisual works, from the parent level through their derived edits down to their specific representations with clips, composites, and compilations alongside. In addition, such content IDs must be accompanied by a compatible system of content delivery channel IDs to capture a complete picture of the broadcasting ecosystem. Following the rules of standardization, workflow-agnosticism, commercial neutrality, and translation to existing IDs ensures broad adoption and invaluable economies of scale.

### REFERENCES & END NOTES

- [1]. Kroon, R., “Bringing Order to Digital Identifiers,” M&E Journal, Winter 2014, pp. 148-150.
- [2]. WaveSeven, “Broadcast Working Group Update – 12/15/16.”
- [3]. Kroon, Richard W. and Raymond Drewry. “EIDR for Sports Broadcasting Content Management Challenges,” The 62nd Annual NAB Broadcast Engineering Conference Proceedings. Washington, DC: National Association of Broadcasters, 2015.
- [4]. Drewry *et al.* “Content Identification for Audiovisual Archives,” International Association of Sound and Audiovisual Archives Journal, No. 45, Summer 2015.
- [5]. Library of Congress (LOC). “Library of Congress Classification.” Last modified October 1, 2014. <http://www.loc.gov/catdir/cpsol/lcc.html>.
- [6]. Online Computer Library Center (OCLC). “Dewey Decimal Classification summaries.” Accessed June 1, 2018. <https://www.oclc.org/dewey/features/summaries.en.html>.
- [7]. British Film Institute. “Advanced search.” Accessed June 1, 2018. <http://collections-search.bfi.org.uk/web/search/advanced>
- [8]. ITU-T (Telecommunication Standardization Sector of ITU). “Information technology - Procedures for the operation of object identifier registration authorities: Generation of universally unique identifiers and their use in object identifiers” Last modified October 14, 2012. <https://www.itu.int/rec/T-REC-X.667-201210-I/en>
- [9]. IEEE (Institute of Electrical and Electronics Engineers). “ST 330:2011 - Unique Material Identifier (UMID).” Last modified August 23, 2011. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7289985&contentType=Standards>.
- [10]. SMPTE (Society of Motion Picture and Television Engineers), “ST 2114:2017 - SMPTE Standard - Unique Digital Media Identifier (C4 ID).” Last modified August 16, 2017. <https://ieeexplore.ieee.org/document/8019810/>.
- [11]. Rivest, Ronald, “RFC 1321: The MD5 Message-Digest Algorithm.” Last modified April 1992. <https://tools.ietf.org/html/rfc1321>.
- [12]. Lilly, Glenn M. “US2002122554 - Device for and method of one-way cryptographic hashing.” Last modified December 7, 2004. <http://www.google.com/patents/US20020122554>.
- [13]. Statistically unique identifiers are logically unique, rather than globally unique. As such, they are not absolutely guaranteed to be unique in all cases, but it is very unlikely that the values will ever repeat.
- [14]. International DOI Foundation. “Key Facts on Digital Object Identifier System.” Last Modified August 16, 2017. <http://www.doi.org/factsheets/DOIKeyFacts.html>.
- [15]. ISO. “ISO 26324:2012 Information and documentation -- Digital object identifier system.” Last modified May 2012. [http://www.iso.org/iso/catalogue\\_detail?csnumber=43506](http://www.iso.org/iso/catalogue_detail?csnumber=43506)
- [16]. It is important to note that the Digital Object Identifier is a digital identifier of objects, not an identifier of digital objects. This is akin to the Flying Purple People Eater being an eater of purple people rather than a purple creature who eats people. (See: Song Facts. “The Purple People Eater by Sheb Wooley.” Accessed June 1, 2018. <http://www.songfacts.com/detail.php?id=25985>). The importance of this is that DOIs in general, and EIDR IDs in particular, can be used to identify media assets in the abstract or when fixed in physical or digital media.
- [17]. Entertainment Identifier Registry Association. “About EIDR.” Accessed June 1, 2018. <http://eidr.org/about-us>.

- [18]. Free public resolution services are offered at <https://dx.doi.org/> and <https://ui.eidr.org/>, respectively or via the URLs [https://ui.eidr.org/view/content?id=\[EIDR-ID\]](https://ui.eidr.org/view/content?id=[EIDR-ID]) and [https://doi.org/\[EIDR-ID\]](https://doi.org/[EIDR-ID]).
- [19]. EIDR Members span the media and entertainment ecosystem including producers, distributors, broadcasters, MVPDs (cable and satellite TV services) post-houses, archives, and data aggregators in the Americas, Europe, and Asia. (See: Entertainment Identifier Registry Association. "Members." Accessed June 1, 2018. <http://eidr.org/membership>).



**Richard W. Kroon** is the Director of Engineering for the Entertainment Identifier Registry Association (EIDR). Prior to EIDR, he held senior positions with the Motion Picture Association of America (MPAA), Technicolor and MovieLabs. Mr. Kroon is a member of the Society of Motion Picture and Television Engineers (SMPTE), International Stereoscopic Union (ISU), and the Advanced Imaging Society (AIS). He is the author of numerous technical papers and several books. Mr. Kroon holds a Bachelor of Science from the University of Southern California (USC), an MBA from Auburn University, and post-graduate certificates in Film, Television, and Multimedia from UCLA.



**François Modarresse** has over 30 years of experience in broadcast and media technologies, playing key roles in the launch of HDTV and H264. He currently runs international business development for EIDR, a consortium of the major Hollywood Studios, many of TV Networks and streaming companies. Before EIDR, François served many clients spearheading innovation in the broadcast production and metadata worlds. Prior to that, he led product marketing at Dolby, productizing major technologies such as Dolby Mobile, Atmos and Vision.