

## **Content ID: The Elephant in the Room**

**Chris Lennon, Harris Corporation**  
**Karen Broome, Sony Pictures**  
**Harold Geller, Ad-ID**  
**Regis Flad, ISAN**

The identification of audiovisual content. While the topic is nothing new, increasingly complex user demands across global application interfaces and user languages are making new demands on content identification.

In the past, proprietary in-house numbering systems were adequate to uniquely identify audiovisual content across various local systems. But when these systems connect to vast networks of content from outside parties, the IDs generated by in-house numbering systems are useless, and it becomes easy to mistake one piece of content for another. With today's high-value content, airing the wrong spot, for instance, is too costly a mistake to make. This need, to uniquely identify audiovisual content so that multiple organizations and systems can seamlessly interchange content and metadata about that content, is the holy grail of content identification.

Without a means of uniquely identifying video content, seamless digital workflows from acquisition to archive are impossible. If content creators, distributors, and archivists can't all agree on a consistent means of identifying content, there is little hope for the goal of automated efficient workflows that serve all the participants in the audiovisual value chain. It is also important to note that these smooth workflows are more than just something imagined by a few dreamers...they are necessary in today's world of lean, often barebones staffing.

If the problem is so common, one might ask, why hasn't this need already been addressed? Those who have developed proprietary workarounds and manual processes to make up for these digital shortcomings might say that the problem has existed for so long that perhaps a global system is not needed.

However, new technologies have created new opportunities for broadcasters and studios. Content that will never find new audiences sitting in a studio archive can now be recycled and re-purposed as streaming or downloadable web content, catch-up TV, VoD (Video on Demand), and mobile broadcasts. The distribution models in the digital world can be complex -- they may span across various industry standards, languages, and locales.

Precise identification of digital content provides an anchor for associated content metadata. This helps broadcasters and studios track multiple versions of a program, distribute them securely, measure the audience over all distribution channels, and assess financial performance. An ideal content identification system would provide a way to associate diverse content versions with each other, allowing various types of hierarchies and program groupings. In addition, smaller works are sometimes used in larger works. This type of "composite" structure can be useful with content such as documentaries that often reference other licensed historical works that may have their own associated identifiers.

Interoperability between various digital media offerings requires some identifying link that can be used across diverse delivery specifications such as MPEG (Motion Picture Experts Group), SMPTE's Broadcast Exchange Format (BXF) and Material Exchange Format (MXF), CEA's (Consumer Electronics Association), OpenEPG (Open Electronic Program Guide), ATSC's A/76B Programming Metadata Communication Protocol (PMCP), Digital Cinema, archival formats, proprietary metadata specifications such as iTunes, and various fingerprinting/watermarking technologies.

## **The History of Content Identification**

In the past, content identification was easy. With each piece of content intrinsically attached to physical media, content identification was as simple as writing on a label attached to a film can or tape.

This strategy started to fail around the time of large Smart Cart Machines (LMS, ACR-225, etc). These machines allowed you to put several pieces of content on a single piece of media, complicating the relationship between content and physical media. Because the media was behind glass in a big machine and virtually inaccessible, you not only needed some form of content identification, but a database to relate the content to its media.

When video file servers came along, the time-honored crutch of physical media labeling was gone and digital asset management systems were no longer a nice-to-have, they were a necessity. Now that broadcasters have adopted file-based workflows and tapeless operations, truly unique globally managed IDs are critical to the smooth interoperability of equipment and systems. To fully realize the efficiencies promised by file-based workflows, content identification requirements cannot be ignored.

As the global need for unique content identification grows and becomes ever more complex, emerging technologies and industry efforts are working to resolve the problems of content identification in today's digital workflows.

## **Identifying Program Content**

After the success of its ISBN (International Standard Book Number) identifier, the International Organization for Standardization (ISO) launched an effort to create a similar identifier for audiovisual content. The International Standard Audiovisual Number, or ISAN, is a globally-unique, centrally managed means of issuing identifiers for program content.



Figure 1 Illustration of an ISAN in use

ISO spent over 10 years developing the ISAN content identification standard with contributions and feedback from dozens of media companies, producers, collection societies, broadcasters, and standards organizations. ISO 15706-1, published in 2002, forms the foundation of the ISAN identification and numbering system, enabling audiovisual work identification on the production level. The ISAN standard was extended in 2007 with the publication of ISO 15706-2, which is a 96-bit identifier that supports the identification of versions of a production. Today, the two ISAN standards describe a complete numbering system for the identification of productions as well as versions of those productions across languages, picture cuts, and regional broadcasters.



Figure 2 ISAN Structure

The ISAN Registration Authority, ISAN-IA, is a non-profit international agency appointed by the ISO Technical Management Board to exclusively operate and promote the ISAN standard, develop and maintain the central ISAN system and database, and appoint ISAN Registration Agencies worldwide. ISAN-IA is based in Geneva, Switzerland, and was founded in early 2003 by AGICOA (Association de Gestion Internationale Collective des Oeuvres Audiovisuelle), CISAC (Confédération Internationale des Sociétés d'Auteurs et Compositeurs) and FIAPF (Federazione Internazionale delle Associazioni dei Produttori di Film).

Appointed ISAN Registration Agencies (RAs) are authorized to process audiovisual content registrations for content creators. These can be appointed to serve a region, country or a specific market. In some cases, these registration agencies provide additional metadata services on a for-profit basis; others operate as non-profits. As of today, nineteen Registration Agencies have been appointed worldwide. Twelve of these agencies are located in Europe (Austria, France, Germany, Italy, Netherlands, Poland, Portugal, Serbia, Spain, Sweden, Switzerland, and U.K.); four are in North America; and

three additional agencies serve Australia, Brazil and Iran. Other RA applicants are under consideration to serve the Asian and Latin American regions.

More than 6,000 content owners have already adopted ISAN worldwide and over 620,000 individual ISANs have been assigned. Of that total, more than 250,000 of those have been delivered to MPAA members who have registered ISANs for feature films, TV series, and Blu-ray versions.

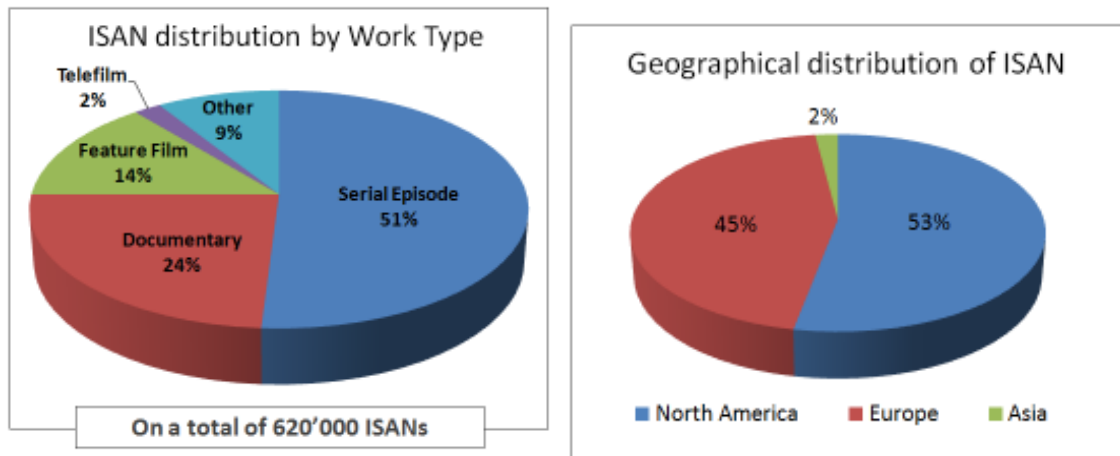


Figure 3 ISAN Distribution Charts

In comparison to other numbering systems that can contain embedded information like the broadcaster ID or broadcasting date, the ISAN is a dumb number and doesn't convey any information on its own. However, combining this globally unique number assigned by ISAN-IA with a database of descriptive metadata collected by ISAN Registration Agencies can enable a vast array of services from multilingual EPG (Electronic Program Guide) metadata services to the coordination of broadcast programs and advertising.

The ISAN central database can be securely accessed worldwide. Access models for this database range from limited free queries of the system through a web browser to paid web services and complex scheduled queries for broadcast systems and equipment that wish to communicate directly with the ISAN central database.

ISAN-IA has developed detailed specifications for over 50 descriptive metadata fields, though only a third of these fields are mandatory. Some of these fields may only be applicable for a particular type of audiovisual asset. A sporting event, for example, might have unique metadata requirements as compared to a feature film or other television program, so specific metadata attributes have been defined for each type of content to supplement the core ISAN metadata set. This specification continues to evolve with changing needs in the digital landscape, but a central core set of ISAN metadata is expected to remain stable.

Information associated with an ISAN may include program or film titles, the release year, the work's duration, its participants and cast, the type of work, the first broadcasting date,

rating codes, and related identifiers such as Ad-IDs. ISAN-IA accepts multilingual metadata and provides search functions across various user languages.

Although ISAN is a voluntary numbering system, some government agencies require it for film release. For example, Screen Australia, the Australian Government's principal agency for funding the production of film and television in Australia, now requires the ISAN for all its production funding agreements. Some industry anti-piracy initiatives also rely on ISAN. For example, the AACCS (Advanced Access Controlled System) consortium requires ISAN as the content identifier for Blu-ray discs, enabling managed copy and user interactivity through online connections.

ISAN remains a young standard. Nevertheless, the strong support provided by producers' associations, studios, broadcaster associations, and industry standards indicate a bright future for this emerging standard.

### **Advertising Content**

As digital markets and channels for audiovisual works increase, so does the volume of advertisements. A typical one-hour program may contain dozens of advertisements, which have their own needs for identification.

Historically, American advertisers used ISCI (Industry Standard Commercial Identifier) codes. ISCI was a manual advertising asset coding system developed by the American Association of Advertising Agencies (4 As) and the Association of National Agencies (ANA). This system has been in use since 1969. While some say ISCI was better than nothing, and did offer a certain level of identification, they came with a lot of baggage.

Because there was no requirement that these IDs be globally unique, anyone could tag a commercial with an ISCI of their own invention. This could result in collisions if two parties decided to use the same code. The industry began generating house numbers to compensate for ISCI's shortcomings, but this strategy was bound to fail in the long term. With advertising assets shared across broadcast properties, identification schemes without any guarantee of uniqueness outside of a particular facility pose significant operational risks.

Here are just a few examples of errors that have resulted from code collisions:

- A global pharmaceutical corporation had duplication in codes when its two agencies were creating codes for the same prefix, but the prefix was being used for two different brands.
- A multi-million dollar spirits corporation stopped using their 26 ISCI prefixes because the coding history could not be tracked and they were afraid of code duplication

- A major national advertiser had a 5-year-old commercial run in a major market in primetime because the agency recycled an ISCI code.

The cost of fixing these errors is great, ranging from \$50 to re-slate an ad to hundreds of thousands of dollars if the wrong ad runs in a major market in primetime.

ISCI was formally withdrawn from the marketplace in October 2007, and has been replaced with Ad-ID, which is a joint venture of the 4 AS and ANA.

Where previously an advertiser and/or their agencies used spreadsheets, proprietary databases, or even loose-leaf binders to keep track of their codes, Ad-ID provides a code guaranteed globally unique and generated from a secure central source. Ad-ID also serves as an advertisement metadata repository, including information about the advertiser, the product, and other descriptive and operational fields. Ad-ID provides web services built on an advertiser opt-in model that is accessible to anyone who needs access to information.

There are currently over 700 advertisers signed up for Ad-ID, including over 74 of the top 100 advertisers in the US, with new users coming onboard regularly. Ad-ID is an advertiser pays model and there is no cost for the use of its web services.



Figure 4 Ad-ID Structure

There are currently three user roles within Ad-ID: Creator, Injector, and Reader.

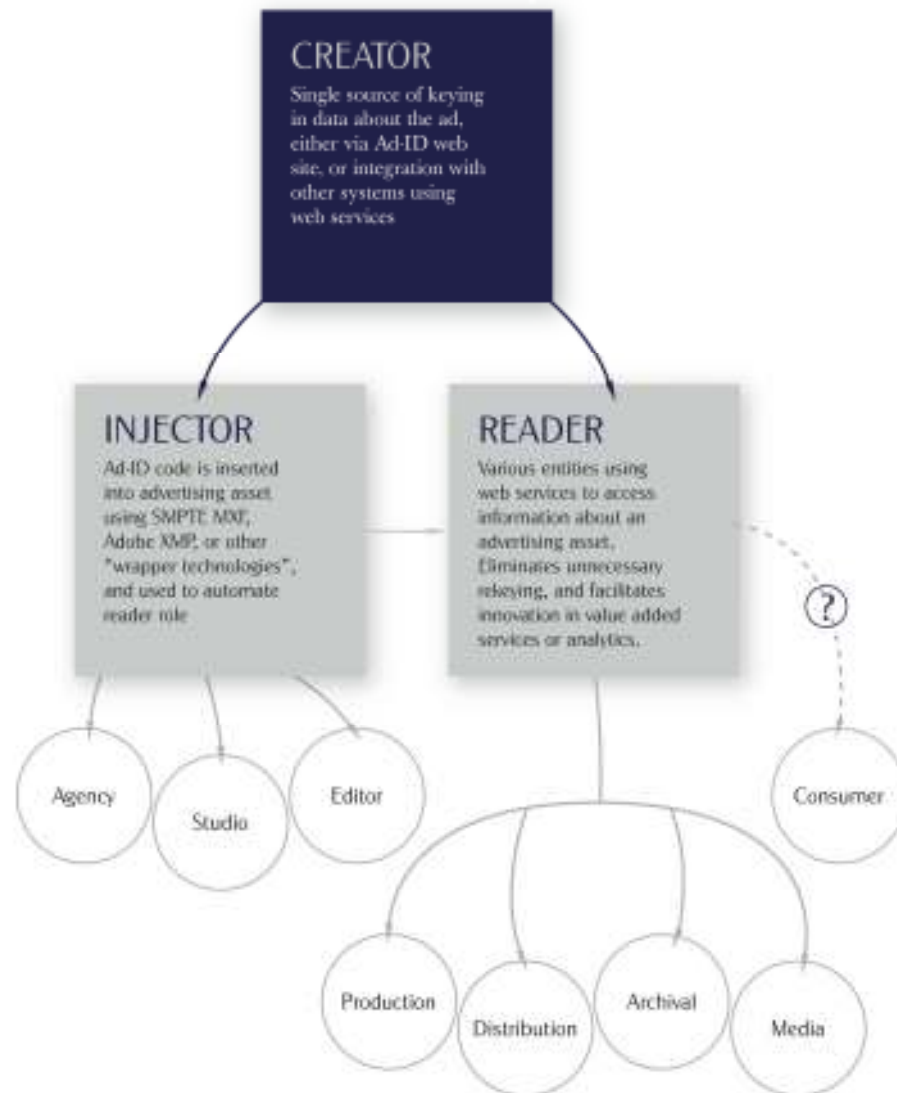


Figure 5 Ad-ID Roles

The Creator inputs the metadata into the system through the Ad-ID website or via web services. Much of the slate metadata (Advertiser Name, Brand Name, and Product Name) can be standardized. The company "Procter & Gamble," for example, can be represented in a variety of ways -- "P&G," "P & G," "P and G," "Procter and Gamble," etc. This role is generally filled by the Advertising agency or the Advertiser.

The Injector role allows authorized third-party entities to inject necessary Ad-ID metadata into formats like MXF and BXF through web services, eliminating the need for manual re-entry and its resulting inefficiencies.

The Reader role is a user who needs the descriptive or slate metadata that would normally have been manually input from source documents, e-mail, faxes, and handwritten notes.

Examples of candidate Ad-ID “readers” could range from a production house creating a slate to a special effects house or post-production facility. Advertisers can also use Ad-IDs and associated metadata for billing and tracking purposes, and they can be associated with the actual essence files contained in digital asset management systems. Ad-ID is made up of a digital slate and eleven categories of metadata which include over twenty data fields.

From the time that an advertiser gives approval to create an ad, to the time that the ad is aired and is invoiced to the agency, the slate information and related metadata can be rekeyed up to 20 times in today’s world.

There is a fourth role envisioned for Ad-ID -- this is of the role of the Contributor. This could be a talent payment company or advertising clearance body that needs to submit selected data to the system, such as the clearance status of an ad, detailed information about the performers in a commercial, or more detailed information on talent cycles.

### **Standards and Content Identification**

#### Universal Labeling

In SMPTE’s Data Element Dictionary, RP210, ISAN has been assigned a Universal Label value of

06.0E.2B.34.01.01.01.01.01.01.11.01.00.00.00.00.00

This metadata identifier can also be used with ancillary data space specifications such as RP214.

In RP210 version 12, Ad-ID will be assigned a Universal Label value of

06.0E.2B.34.01.01.01.0C.01.01.11.0B.00.00.00.00

### **BXF**

The Broadcast eXchange Format (SMPTE-2021) tackled the issue of content identification when it was created. After much debate, the decision reached by the committee that created BXF was to adopt a structure allowing for an infinite number and variety of Content Identifiers on a single piece of content.

In this way, BXF does not take a stance as to which Content Identifier is the correct one to use on a particular piece of content. Instead, it allows the applications exchanging messages about content to associate any number of identifiers with that content, which facilitates the mapping of content identifiers across applications. ATSC (Advanced Television Systems Committee) metadata standards take a similar approach and highlighted the use of ISAN in their A/76B Standard (PMCP).

### **MXF and Digital Cinema**

ISAN identification is also permitted in KLV structures, such as MXF (Material eXchange Format). MXF is a SMPTE file exchange format used in production, post-production, and archival applications. In addition, it is the underlying standard used by Digital Cinema implementations, which also use ISAN for identification of content in DCPs (Digital Cinema Packages).

### **UMID**

SMPTE's Unique Material Identifier (UMID) standard is used by technical operations staff in production and post-production work to identify specific instances of media files. UMIDs can be used to identify either audio or video bitstreams.

Unlike ISAN identifiers, UMIDs do not require a central registration authority so they can be generated on-the-fly. While a completed work would typically have a UMID assigned by the final post-production workstation, this identifier is not likely to be used outside of the organization that generated it. A published work may well have both an ISAN and a UMID, but the UMID would not likely be widely distributed or publicized.

### **Other Digital Media Identifiers**

Sometimes digital media packages include a number of digital assets, including subtitles, audio tracks, sound recordings, and other ancillary files. There are other industry standard identifiers that can be used with these assets.

### **ISRC**

The International Standard Recording Code (ISRC) is used primarily to identify sound recordings, rather than the works that appear in those recordings. However, these codes can also be assigned to audiovisual works such as music videos, which are also eligible for an ISAN. Because of this, a music video may have both an ISAN and an ISRC.

### **ISWC**

The International Standard Musical Work Code (ISWC) is used primarily to identify musical works, rather than specific recordings of a musical work that would be identified with ISRCs.

### **ISTC**

The International Standard Text Code is designed to identify written works, such as subtitle files, scripts, and other written materials. This standard is still in development by ISO's TC 46/SC 9. These codes could also be associated with metadata about the written work.

### **UPC**

The Universal Product Code (UPC) is used to identify retail products. Its numbering and barcode standard were initially used in the grocery industry, but have since expanded to include the entire commercial product space. DVDs, CDs, and Blu-ray discs may have UPC codes in addition to ISANs.

## **EAN/JAN**

The EAN (European Article Numbering) and JAN (Japanese Article Numbering) standards are the equivalent of the UPC in Europe and Japan. The main difference between the two standards is in the number of digits each employs.

## **CRIDs and Other URNs**

CRIDs (Content Reference Identifiers) are used with the TV-Anytime standard and can also be used to represent product hierarchies and groups. CRIDs can be generated without the help of a central registrar and resemble a URL. This makes it easy to use CRIDs with any device that connects to an IP-based network.

An example of a CRID is:

`crid://spe.sony.com/casinoroyale/english`

There are other proprietary schemes that use URNs to ensure that identifiers assigned are globally unique and can be accessed over by web services. In some cases, these identification schemes may actually combine several types of identifiers with a common prefix to group them together within a system environment.

## **Conclusion**

Content identification solutions have come a long way from the days of scribbling on film can or tape labels. While the dream of a single content identification scheme for all audiovisual materials across all media has not yet been realized, new content identification standards are now addressing the needs of networked systems that span a variety of languages and locales. It is clear that the demands of today's file-based workflows both within and between organizations require new content identification approaches. Some of these identification systems support the complicated product hierarchies found in the motion picture and television industries. The number of content creators relying on internal identification schemes alone are rapidly decreasing as more global solutions, such as ISAN and Ad-ID, emerge.

## **Bibliography and References**

- ISO 15706:2002 Information and documentation -- International Standard Audiovisual Number (ISAN)
- ISO 20924-1:200x, Information and documentation – Version identifier for audiovisual works (V-ISAN) – Part 1: Format and use
- A/76B:2008, “Programming Metadata Communication Protocol, Revision B”, Advanced Television Systems Committee, Washington, DC, 14 January 2008
- SMPTE-2021, “Broadcast eXchange Format (B XF)”, Society of Motion Picture and Television Engineers, 1 April 2008

- RP210, “Metadata Dictionary Registry of Metadata Element Descriptions”, Society of Motion Picture and Television Engineers, 7 Novemeber 2002