I Never Metadata I Didn't Like

Andrew Janitschek
Radio Free Asia
Washington DC

Abstract - *Data continues to grow and drive much of broadcasting. An integral part is the metadata we provide and consume. It is critical to the future growth of new technology. Understanding metadata is simple enough but advancements in machine learning, artificial intelligence, cloud storage, the Internet of Things (IoT), and more have brought an explosion of data. Using metadata, or data about data, today's media is able to be interactive and complex. Unfortunately, with many metadata standards available, it can be an overwhelming challenge to ensure your digital files will work and play well with others. This paper chronicles metadata from its beginnings, through today, and into the future. It is impossible to cover every aspect of metadata here but let this whet your interest and continue into the future. As broadcasters, we provide metadata to tell the world much about the essence of our media. New generations will thrive on that metadata and develop new consumption models, standards and management systems on their own. How we use metadata is changing and we must stay on the wave of progress or let others drive that momentum without us. Metadata is the future.*

MIND YOUR METADATA

We see it daily; use and inclusion of metadata into broadcast workflows continues to grow and drive much of our work. An integral part of all broadcasting is the metadata we provide and our audience consumes. You cannot ignore discussions of metadata without expanding it uses and ties to greater technologies of today like big data, the Internet of Things (IOT), machine learning, over the top (OTT) and more. These are topics for others and not for this paper. By reading this though, you are taking a leap to understanding the future of broadcasting. Metadata is our past, present, and our future.

Over the years metadata has evolved from just simple entries on a computer file to a plethora of entries that has grown to encompass everything from date, time, location, equipment involved, and moods to almost anything imaginable. As a sample, Figure 1 below shows the simple metadata you see when examining the properties under Windows Explorer for the song, You're the One by DJ Khaled. Many of us know and understand this use of metadata already.

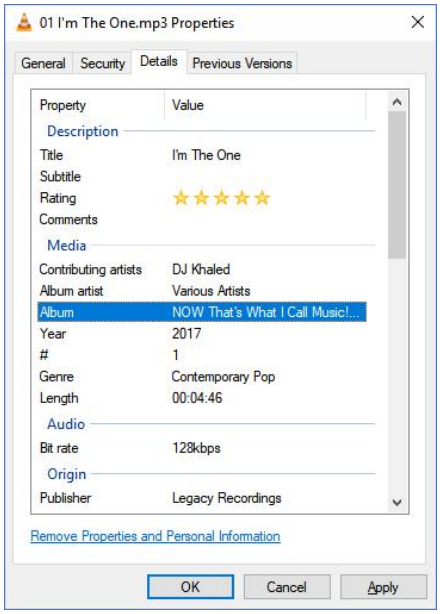

FIGURE 1: METADATA FOR 'I'M THE ONE'

At Radio Free Asia (RFA) we embrace metadata and as we dive headlong into 2019, its importance is more and more evident. As this was being written, we are working closely with all other US international broadcasters to choose a metadata standard we will all use. With many metadata standards available for most every industry imaginable, choosing the right one can be difficult. Just a few metadata standards worth evaluating are:

- IPTC's Global Standards for News Media
- European Broadcast Union's EBUCore
- Exchangeable image file format (EXIF)
- XMP

The IPTC's metadata standard for a range of media is superb for administrative, descriptive, and copyright information. At the EBU, their EBUCore is the highlight of their metadata standard; EBUCore, along with their EBU Class Conceptual Model (CCDM) provide the framework for descriptive and technical metadata. This is a strong contender for adoption. EXIF specifies formats for graphics and pictures, audio, and additional tags used by smartphones and other digital cameras, scanners and most any other equipment that records or creates images and sound. We'll look at EXIF again soon. XMP is great for creating and exchanging metadata in data sets and documents. Certainly there are pluses and minuses to all standards under consideration. From our viewpoint, we are happy with any standard with a

huge global acceptance base and works well between the agencies. The standard we chose may not be the right one for you but if our metadata can work seamlessly with yours, isn't this a great reason to consider adopting the same metadata standard? This isn't a sales pitch, but if large broadcast entities can agree on one metadata standard, that would be a positive step toward the future.

TO THE WABAC

I hope you caught the reference to the 2014 animated film, Mr. Peabody & Sherman. The WABAC is their fictional machine to travel back through time. Without a WABAC, we can still understand the roots of metadata and how it is maturing. And if you're interested in the movie, please check Amazon, Netflix, Hulu, or your local TV stations.

Metadata's roots date back to 280BC and the Great Library of Alexandria. Zenodotus, a Greek grammarian and the Library's first librarian, attached tags to the ends of scrolls identifying the author, title and subject of each document. This was done so each scroll could be returned to the correct location in the library without needing unroll and view the scroll. Can you imagine the toll it would take on the scrolls to have them opened and closed dozens, or hundreds, of additional times just to determine what was in the scroll, and then where it belonged in the library?

Over the centuries following, the fine tuning of categorization grew. Prior to the growth of computers, the best example of metadata we see time and time again is a library card catalog that used the Dewey Decimal System. Remember going the public library decades ago to find cabinets full of drawers containing 3x5 cards in alphanumeric order? Those cards provided more than just basic information on the books we wanted, or needed, There were two huge sets of cards; one set of cards 'by author' and the second set of cards arranged 'by title.' On each card you would find the book's alphanumeric identification number, the section and shelf where the book resided, the title, author, and the subject along with a short description. Even though the concept is simple, it was not always easy finding an author's listing or a specific book you needed. As you can imagine, there were occasional problems created by a misplaced or lost card. It was a catastrophic event when there was a fire in any library until the age of computers allowed us to make backup copies of the library catalog.

Melvil Dewey published the first version of the Dewey Decimal Classification (DDC) in 1867. The latest version was released in 2011. In time, the DDC became better known as the Dewey Decimal System. It introduced new concepts of classification and streamlined the addition of new books to any library, ensuring they were placed in the correct location based on each books subject. Since then, the DDC has been adopted for metadata standards in countless libraries in over 100 countries. It established the global standards for classifying, identifying and locating specific media. The reality of today is that metadata is not just a term for managing library assets. We now hear the term tied to all industries; broadcasting and streaming included. As the use of the cloud and virtualization grows, so does the importance of accurate metadata. Drawing from the DDC, metadata should include these minimums:

- A way to create the metadata
- Its purpose
- The author
- The date and time of creation
- Where the data is located whether on the Internet or a network

How many of you remember similar paper cards for your station's music library? I was the lone broadcaster at the American Forces Network Korea at Kunsan Air Base, Republic of Korea in 1983. Each week we would receive a new mail shipment of record album pressings from the American Forces Radio and Television Service (AFRTS) in Los Angeles. For each song, we received one 'song title' card along with a matching 'recording artist card.' Thankfully the cards were pre-printed so all we need do was to break the sheets of cards up into their individual cards, then go through the tedious process of filing each card. Anything less than filing the cards '100% accurately' was unacceptable. Unfortunately when humans were involved in the process, some cards always ended up being misfiled, or worse yet, lost. By the mid-80's, AFRTS switched to weekly CD shipments that came with electronic library cards on a disc we loaded in to a computer; ah, metadata!

So where does the term 'metadata' come from and who coined the name? Another part of metadata's history is traced back to 1967 when it was first described by David Griffel and Stuart McIntosh of MIT's Center for International Studies when they wrote: "In summary then, we have statements in an object language about subject descriptions of data and token codes for the data. We also have statements in a meta language describing the data relationships and transformations, and ought/is relations between norm and data. These meta language statements are named" [1]. A year later, Philip Bagley actually used the term, metadata, in his book, Extension of Programming Language Concepts, so let us assume Mr. Bagley is the Father of Metadata.

RFA started working with metadata two decades ago when Cart Chunk was released as the traffic data file format for the radio industry. In 2002, Cart Chunk was formally accepted by the Audio Engineering Society (AES) Standards Committee and is now better known now as AES46-2002.

FIGURE 2: CART CHUNK LOGO

For its time, the concept of Cart Chunk was groundbreaking. It let us embed common resource information we needed at our radio network, and allowed us to do that directly into audio WAVE files. Prior to CartChunk, we did this with pens and grease pencils on paper labels mounted on audio cartridges (aka carts), cassettes, and reels of tape. As a non-proprietary system, Cart Chunk established the standard.

Only now, as our broadcasters realize the importance of our Digital Asset Management (DAM) system are they coming forward to sip the Kool-Aid. RFA's current DAM system is ResourceSpace which is an open-source solution. It is standards-based and supports several existing metadata standards for exchanging metadata including XMP, IPTC and EXIF.

For years we saw little-to-no metadata entry by the content creators on media resources placed into the system. The primary reason given was, 'I am too busy.' Only over time, and with a hands-on push from senior management, did people stop looking at our DAM system as a burden and started seeing a resource that can continue to give time and time again. For instance, if you want to find a resources relating to the Dalai Lama, you need to put that term 'Dalai Lama' in the keyword search. If there are no resources with the proper keywords associated, our DAM system would return nothing. Without access to all available resources, writing a news story is just that much more difficult.

Today, we know better. We now have a better handle on entering our metadata. Broadcasters are forced to enter metadata fields for every asset uploaded to the DAM system. Overtime our search functions became, and are becoming stronger. Here, in 2019, we are doing a much better job of entering the metadata for each resource, and we look forward to the next 5-10 years of not only improving our metadata, but also adopting a global standard so we can share resources with other broadcasters.

HAVE YOU EVER MET A DATA?

Metadata helps you describe, use, find and manage content and data. It allows you to determine essential metadata properties needed to control and use business information. Use of standardized metadata helps support interoperability and information sharing. Getting metadata entered, automatically or manually, the right way the first time will make it easier to search for, and use, resources. It all starts with the entry of metadata. Your metadata should be descriptive, clear, relevant and easy to add. Being descriptive means to work at specifics and avoid nebulous, or vanilla, data otherwise you will have to add additional metadata afterward to ensure your resources can serve their purpose. Being descriptive goes hand-in-hand with being clear. Your metadata must be understandable in order to ensure you provide clarity and avoid redundancy. The ABCs of writing are worth mentioning here when applied to entry of metadata; accuracy, brevity, and clarity. This is a good rule to apply when writing metadata. In order to be relevant, your metadata must be germane to the resource you are labeling. Metadata must be easy to add. The easier it is to enter metadata, the lower the total cost. If we force people to spend a long time filling out line after line of metadata, they will not bother to fill it out if avoidable. It is always a tough balance to achieve the highest accuracy of your metadata and also make it simple to add. The time you take to get it correct now will pay off in multiples in the future. Frankly, fine tuning your workflow never ends.

It is generally agreed there are 3 main types of metadata. They are descriptive, structural and administrative metadata. There is also statistical and reference metadata. Each serves its own important purpose to the management of today's digital life.

Descriptive metadata is fairly easy to understand; it is information that describes media or files thereby making it easier for discovering and identifying them. It is really not too different than the library cards mentioned earlier. Descriptive metadata provides the basic elements you need for queries like title, creator, photographer, author, and keywords.

Structural metadata describes containers of data. For example, how pages of a book are ordered to form chapters. Structural metadata describes the types, versions, relationships, and other characteristics. In essence, structural metadata formatting stays the same. It applies to many files and helps you interpret data by describing what each field means, even if it is concentrating on descriptive metadata.

Administrative metadata provides information that helps manage data. These are informational tidbits like when and how the file or resource was created and who has access to it. Under administrative data, there are many subsets. Two that are often categorized as additional metadata types are right management and preservation metadata. "Rights management metadata…deals with intellectual property rights, and preservation metadata, which contains information needed to archive and preserve a resource" [2]. Statistical metadata, or process data, describes systems and procedures that assemble, process, or create statistical data [3]. Here are just a few other subsets of metadata worth knowing since you will likely work with them someday:

- Public metadata
- Private metadata
- Preservation metadata.

Public and private are self-explanatory, but preservation data is data needed to archive and preserve a resource.

STANDING UP FOR STANDARDS

In life, there are standards for everything. These are agreed to and accepted ways of doing things. There are security standards, standards of conduct, food standards and countless more. Many standards are established by organizations that support and adjust their publicized standards as needed and when needed. We rely on standards-based organizations to

create, adjust and enforce standards no matter where they are. For instance, the American National Standards Institute (ANSI) is the de facto leader of all US standards. Everything from systems, processes, services, products, and personnel in the United States, ANSI establishes and coordinates standards so American products can be sold overseas [4].

Another standards organization is the International Organization of Standardization (ISO). They develop, publish and promote worldwide proprietary, industrial and commercial standards. The ISO is comprised of representatives from 163 national standards agencies, including the ANSI. You can consider the ISO as the United Nations for standards. This is the same organization that brought us the ISO 9000 series of Quality Management Principles which include customer focus, leadership, engagement of people, approaches to processes, improvement, evidence-based decision making, and relationship management. The ISO publishes, and sells, published standards that clarify requirements, specifications, guidelines or characteristics to guarantee materials, services, systems and products are suited for their intended use.

There are at least as many metadata standards as there are industries today. Chances are high you have heard of, or used, one of the major metadata standards. Here are a few in use today; I hope you recognize some of them:

- Dublin Core
- EBUCore
- PBCore
- MPEG 7 (ISO 15938-5)

The Dublin Core metadata standard has been around for over 20 years and is a small set of vocabulary terms used to describe physical and digital resources. Dublin Core focuses on network resources creating a digital version of a library card catalog system for the Internet and can be used for basic descriptions of resources to "combining metadata vocabularies of different metadata standards, to providing interoperability for metadata vocabularies in the linked data cloud and Semantic Web implementations [5]." Dublin Core was initially comprised of 15 data elements. There were two levels to start; they were simple and qualified. The original 15 data elements were a part of the Simple Dublin Core. Audience, provenance, and rights-holders comprised the elements of the Qualified Dublin Core in addition to qualifiers that helped fine tune the clarity of elements and thereby making them more valuable when discovering resources. Both the Simple and Qualified Dublin Core were combined in 2012 and renamed to the Dublin Core Metadata Initiative (DCMI) Metadata Terms and is now a single set of over 50 terms, including the original 15, and the qualifiers. Below is an example of Dublin Core metadata for a photograph. The personal information below is fictional and only meant for this paper.

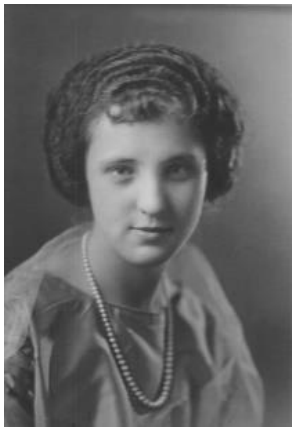

FIGURE 3: MARGARET YELINEK

Title=Margaret Yelinek

Author/Creator=Not known

Subject and Keywords=The New Jersey Archives photographs

Subject and Keywords=Mayors of Jersey City

Description=This photograph was taken in 1924 at the annual Mayors' Conference. Mrs. Yelinek was the first woman elected as the City's Mayor. She was very popular within the community and after one term, left the office. Afterward, she dedicated her life to running her family's business.

Publisher=The New Jersey Archives Digital Collections

Date=1924

Type=image

Format=image/jpg

Resource identifier=http://hdl.sample.net/1234/5678

Resource identifier=NJA123-456-7

Source=NJ Archives

Rights=This image is provided for research purposes only and must not be reproduced without the prior permission of the Archives Program, New Jersey Archives.

EBUCore, from the European Broadcasting Union, is their flagship metadata standard which is primarily for use with audiovisual metadata. Based on Dublin Core, EBUCore is another set of descriptive and technical metadata. As mentioned earlier, EBUCore is currently under evaluation by US international broadcasters. Moving from one standard to another is interesting. Some systems allow that to easily

happen. Within Radio Free Asia we too are evaluating the EBUCore standard. So what does that mean for the future of RFA's metadata? We do know that metadata is important to the future of RFA, and all of our archives. Will we have to update the metadata for older media? I am not sure, but we are making the effort today to ensure today's metadata is useful, precise, and concise for the future.

PBCore is the metadata standard established by and for US public broadcasters. PBCore defines how you organize audiovisual resources. A key advantage of PBCore is that media resources and collections can easily be shared between media systems and public broadcasters [6]. PBCore is not limited to just public broadcasters. Since its debut, many others, whom are not public broadcasters, have adopted PBCore to manage their assets. For example, AVP is a management consulting and software development firm. They have been involved with PBCore since its early development. Another commercial user of PBCore is the firm, George Blood, based on Fort Washington, PA. They specialize in digitizing archival audio, film and video.

MPEG-7 is a multimedia content standard that uses XML, or Extensible Markup Language, to accumulate metadata. The metadata can reference timecode to tag particular events, or find a song by searching for a few keywords. Also known as the Multimedia Content Description Interface, MPEG-7 does not address encoding of media resources like MPEG-1, -2 and -3, but "it provides metadata for and describes low-level features of multimedia components and resources, such as images, audio and video contents, and 3D scenes, and their spatial, temporal, and spatiotemporal segments, and supports timecode-based annotation [7]."

Now let's take a quick look at a two media related metadata schemas that show relationships between metadata elements:

- IPTC Photo Metadata
- EXIF

The International Press Telecommunications Council's (IPTC's) Photo Metadata schema is used extensively "because of its universal acceptance among photographers, distributors, news organizations, archivists, and developers [8]." The schema defines structure of the metadata, the properties, and fields so images are described to the highest levels humanly possible and then easily retrieved. There is also the IPTC's Video Metadata (VMD) Hub which is very useful when using metadata across existing standards. The VMD Hub performs processes in a well-defined, reliable and efficient way. It can describe the audio and visual content of a video and when determining data for copyrights, VMD Hub provides the video's administrative details and technical characteristics. All the VMD Hub properties can be expressed through the use of the EBUCore, XMP and JSON technical standards.

The Exchangeable Image File Format, or EXIF, stores metadata in graphics and image files like GIF, JPEG, PNG, and TIFF files. EXIF lets you store a range of valuable information like camera make and model, data and time of the photo, shutter speed, white balance and more. I'm happy to say RFA's DAM system, ResourceSpace, is EXIF compatible. I find one of the most key bits of information it provides is geolocation metadata. This is information stored in the header of the image indicating where the photo was taken. When viewing an image in ResourceSpace, it displays a Google map, with a location pointer, where the camera and photographer were when the photo was taken. An excellent associated application is the ExifTool by Phil Harvey which is freely available online and lets any user read and write a file's metadata. This is great tool for adding, editing, or deleting metadata as needed. The ExifTool supports many different metadata formats and the metadata that is automatically written to picture files by most major camera manufacturers. This is an excellent option when you need to update metadata in an older resource.

Resource Description Framework (RDF) is a standard model for data interchange used to communicate facts about data resources on the World Wide Web. Its main focus is to show metadata about data resources and their relationships with other assets. RDF is one of the building block standards of the Semantic web and can be expressed in various formats such as; XML and JSON. Not only does RDF allow connections between data resources, but also encourages the creation relationships between them [9].

Enter the Recording Information Notification (RIN) standard. RIN will let gear and digital audio workstation manufacturers enable users to capture and store metadata, making it all-the-more important for long term use. This will also help drive next-generation broadcast systems letting media consumers personalize their experiences, no matter where or when they use the media. RIN is actually a metadata file format that is designed for machine-to-machine communication. Interpretation of RIN XML files is accomplished by digital audio workstations and metadata collection applications [10]. Soundways RIN-M software, now called Sound Credit, is a way to mine for technical information, album art, liner notes and production credits from any audio session. The software is a free plug-in that works on all DAW systems, but also works as a standalone program [11].

THAT'S BIG

The amount of global data, in its entirety today, grows by magnitudes. According to Forbes magazine, there were 2.5 quintillion bytes of data created daily in mid-2018 and the rate is increasing [12].

The creation of metadata continues to grow at phenomenal rates too. As long as storage is available, systems can process metadata and we can continue to have access to it. We are not necessarily experts in the use of metadata at RFA, but we feel comfortable with its use and we are taking

steps to ensure it's available for future generations. As proof of the importance we place on metadata, it is supported and promoted from the highest levels in our company. The use of our DAM system, ResourceSpace, has been supported, is supported, and will continue to be supported.

As you know, metadata is information used to describe the digital resource we are accessing; metadata IS data. Describing the content and context of any computer file makes it more valuable and beneficial. The main reason we want metadata is to help us, as providers and end users, find information and help us discover resources we did not know existed. Metadata also helps us organize our digital resources, delivers identification information, and it supports long term archiving and future use of the digital resources.

Metadata is the key to finding resources through the use of relevant benchmarks. The benchmarks are items like keywords, date and time of creation, the identity and location of resources, geolocation data and more. As news outlets have reported, the metadata from all our communications, include Internet traffic, has been and continues to be widely collected by various governments, organizations and commercial businesses like Facebook. This data is used for traffic analysis, marketing, and in some instances, mass surveillance too [13]. We now realize there can be dangers involved in the use of metadata.

The creation and use of metadata is a constant balancing act between quality levels of data versus the costs of making and storing the data; the greater the granularity, or detail provided in metadata, the greater the cost to generate it. When metadata is created using low-cost or free systems, it is more likely the metadata will reflect a lower level of precision.

METADATA BASICS

To understand metadata better, one needs to understand some of what it takes to use it. While not all encompassing, below are parts of the metadata puzzle you need to know.

A **metadata standard** institutes a common understanding of the meaning of the data, in order to guarantee the accurate use of the data and that it is understood by all that use it.

A **metadata schema** shows the logical relationships between metadata. It is also called an element set. The schema creates guidelines for metadata use and determines how it is managed. It also defines semantics, syntax, and available options. "If your data matches a developed schema, the use of that schema will result in the best metadata for your data [14]."

A **data dictionary** this tells us the context of a collection of data. This is where you document what the columns of a data table mean. First you need to determine the names of the columns of your table so you can search. Then you need the definition, so you know what each column means. You might also want to record the data type (text, numeric, etc.) so it can be interpreted correctly, discover the amount of space it takes up in the database, and keep a list of possible values. The dictionary explains codes used to represent data. One example is ANSI/NISO Z38.87-2006 Data Dictionary – Technical Metadata for Digital Still Images. This defines a set of metadata elements for raster digital images.

Use of **comma separated values** (CSV) is a way of storing data. You use commas to separate values while you can also use quotation marks to identify one specific value. For example, 1,5 would normally represent a value of 1 in one column and a value of 5 in the next column; "1,5" could be used to represent a single value of 1 ½. You can also use tabs generated by your keyboard so the data can be interpreted correctly. The tab mark indicates spaces between different values; these are better known as tab separated values (TSV). Another option is to use fixed width files. In this case all columns are a certain width.

The **semantic web** is an additional part of the World Wide Web Consortium's (W3C) web standards which endorse common data formats and exchange protocols. The Resource Description Framework (RFD) is the foundation of their endorsement.

Ontologies are representations for organizing information and concepts into a structured system. In the end, ontologies may or may not classify data. In other words, an ontology is a model of knowledge [15].

METADATA AUTOMATION

Let's face it; the best way to create and gather metadata is when it is done automatically and without human interaction. When the creation of metadata is error free, there is no limit to the amount of metadata we can gather. As mentioned earlier, there were over 2.5 quintillion bytes (10^{18}) created daily in mid-2018; that's 2,500,000,000,000,000,000 bytes and the numbers are climbing. "By 2020, it's estimated that for every person on earth, 1.7 MB of data will be created every second [16]."

Many of the major software companies and cloud service providers already have tools that will convert audio-to-text and text-to-audio. They also provide tools that look deeply into your digital images and create metadata to accurately describe the contents. If you thought artificial intelligence (AI) was something only for the future, you are wrong; AI is here. Using technology like neural networks, the tools make a 'best guess' when evaluating your media files. As time goes on, their accuracy increases with the greater use and as users and developers help these tools learn and increase their precision [17].

At CES 2019, metadata did not garnish huge headlines but it nonetheless was on the minds of many. According to ThinkAnalytics CEO Gabriel Berger, "[their] solutions understand metadata and viewer behavior [in order] to personalize the viewing experience and boost subscriber engagement…Data mining led to machine learning and that has made way now to artificial intelligence (AI) [18]."

Meanwhile, TiVo Corporation announced that it signed a multi-year intellectual proper license renewal with Minerva Networks which provides service management infrastructure for the delivery of pay television services. This is part of a

new global deal to use TiVo as the primary metadata provider for Minerva's *Your TV Now* which is a cloud video platform for pay TV operators [19]. Lastly Gracenote's VP of Product Management, Greg Gentschev, said during CES 2019 they are "already getting a lot of interest from customers" about their next-generation descriptive metadata solution, called Video Descriptors that uses machine learning. According to Greg, this "kind of confirms our idea that richer metadata is one of the things that's been holding search and discovery back [20]." He made these comments to the Media & Entertainment Services Alliance (MESA) at CES 2019. Gracenote, a business unit of the Nielsen Company, is an entertainment data company specializing in music, video and sports metadata. Gracenote also provides automatic content recognition technologies to companies worldwide.

As we see, automated creation and capture of metadata is here. Who are some of the leaders and where are they today? For brevity, we will only look at the top three; Amazon Web Services, Microsoft, and Google.

Amazon Web Services (AWS) is the global leader in providing cloud as a service. Metadata is easily updated to a Simple Storage Service (S3) object through the standard AWS console. But the real power of AWS comes from services like AWS Rekognition. Rekognition is a deep learning-based visual analysis service that helps you search, verify, and organize millions of images and videos. It can automatically label objects, concepts and scenes in your images, and when done, it presents you with a confidence score. If you have an Amazon account, visit their AWS site to log in; new users qualify for a year of free-tier storage which provides up to 5 GB of their S3 standard storage. Without making any commitment to AWS, you can view the Rekognition demos and upload your own JPEG or PNG that is less than 5 MB, to see the results. Rekognition Video is easy to use through the AWS API and offers real-time analysis of streaming video and facial analysis. This lets developers easily add visual analysis to existing applications and automatically add/retrieve metadata where none existed before. When using Rekognition Video, it automatically extracts metadata, which can then be used to create a searchable library and personalize your metadata. According to Amazon, you can also start creating useful metadata from media resources automatically by using their Media Analysis Solution. This provides AWS CloudFormation templates you can use within a few short minutes to begin pulling metadata. Processing files through AWS is a simple as uploading your files through the appropriate AWS web page. The Media Analysis Solution uses Rekognition for facial recognition, Amazon Transcribe to create a transcript, and Amazon Comprehend to analyze the transcript. It is also powerful enough that you can train Rekognition to recognize specific people. Some users of AWS Rekognition include CSPAN, the Scripps Networks Interactive, Motorola Solutions, and PopSugar. It seems like the future of automating metadata creation has already arrived with AWS.

The InfoLibrarian API with Microsoft Azure provides a metadata management server that lets you automate metadata cataloging, collaboration with others, and access to your information at any time. It automates the collection of metadata from a multitude of sources. You use the InfoLibrarian web portal which has all the documentation needed for you to do your work. Some of the power in InfoLibrarian is its ability to provide business intelligence on a self-serve basis. This reduces the demand on your internal and external IT resources, and gives you the confidence that you have all the correct data. InfoLibrarian catalogs and manages your metadata to ensure you receive precise search results and accurate data for analysis. Some users are Fuji Film, the United States Postal Service, Unilever, UPS and the University of Arizona. Additionally, Power BI from Microsoft, scheduled for launch in April of this year, is Microsoft's business intelligence and analytics tool that scans hundreds of sources for metadata and transforms it in to dashboards and reports that will constantly update the data from your resources. Power BI solves many business challenges letting you share your data and post the business intelligence to your website or app.

As part of Google, the Cloud Video Intelligence API makes videos searchable, and discoverable by extracting metadata. You can search an entire video file in your collection. Using Google's library of 20,000 labels and your files in Google Cloud Storage, it marks your videos and helps identify key entities and when they occur. A most impressive aspect is that Cloud Video Intelligence can tell the difference between noise and audio signals. To do this, it examines all relevant information from the entire file on a 'per frame' basis. The power lies in that it lets you retrieve insights from your video files without the need of using any machine learning [21]. Another addition to Google's metadata toolbox is the Cloud Vision API. Since its 2016 launch it has helped Google's customers and clients extract metadata from over 1 billion images. Cloud Vision recognizes millions of entities and offers enhanced optical character recognition (OCR) capable of mining text from documents like books, letters, and contracts [22]. One successful user of Cloud Vision is Realtor.com. They use it to pull metadata from any image of For Sale signs taken on a mobile app so the Realtor website can provide more details on the property [23]. Some other users of Google's Cloud API are CBS Interactive and Disney.

WHERE TO FROM HERE?

There is no stopping the wave of metadata. What we must do is ensure we manage it smartly. With much advancement in technology like big data, machine learning, and artificial intelligence that rely on metadata, it is easy to feel overwhelmed. The current offerings from the big 3 mentioned above, are only the tip of the iceberg. Many companies are selling services to help us control the flood of metadata, and use it effectively. The future holds greater speeds, larger and cheaper storage, and wider searches. It also foretells of simplification in coding, processing and analyzing data. It is today's metadata that will speak to future generations; we are giving them the keys to life as we know

it and we are making the future better because of our work today. Where will metadata be in 5 years? How about 10 years? It will be growing, intelligent, multilingual, and everywhere.

SUMMARY

The volume of data worldwide continues to grow and drive broadcasting. An integral part of the entirety of global data is metadata. It is of ultimate importance to the future growth of new technology. Understanding metadata was easier until newer technologies brought about an explosion of data. Today's media can be interactive and complex, but it still needs metadata. We have chronicled metadata from its beginnings, through today, and into the future. How we use metadata is changing and we must stay on the wave of progress or let others drive that momentum without us. Metadata is an important to key to the future of media, technology, and humanity.

REFERENCES

[1] Griffel, D, M, and McIntosh, S, D. "ADMINS – A Progress Report", January 1967, p. 2.A.

[2] Zeng, M, "Metadata Basics: Metadata Types and Functions", http://marciazeng.slis.kent.edu.

[3] Dippo, C, S, and Sundgren, B, "The Role of Metadata in Statistics", U.S. Bureau of Labor Statistics, Jul 19, 2008, p. 4.

[4] "Introduction to ANSI" and "About ANSI", American National Standards Institute, http://www.ansi.org.

[5] "Dublin Core", http://en.wikipedia.org/wiki/Dublin_ Core.

[6] "What is PBCore?", http://pbcore.org/what-is-pbcore .html.

[7] "The MPEG-7 Standard", http://mpeg7.org/mpeg-7-standard.

[8] "Photo Metadata", https://iptc.org/standards/photo-metadata.

[9] Baden, D, M, and Reck, R, "Using the Resource Description Framework (RDF) to Simplify Content Management" April 2009, p. 1.

[10] DDEX Secretariat, "Recording Information Notification (RIN, Version 1.1): Introduction", Nov 30, 2018, https://kb.ddex.net/pages/viewpage.action?pageId=9504457.

[11] "Give Credit Where Credit is Due: About", http://soundcredit.com/about.

[12] Marr, B, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", Forbes, May 21, 2018.

[13] Ma, A, "China is Building a Vast Civilian Surveillance Network - Here are 10 Ways it Could be Feeding its Creepy Social Credit System", Business Insider, Apr 29, 2018

[14] University Librarian, "Research Data Management: Metadata Schemas and Examples", Australian National University, http://libguides.anu.edu.au/c.php?g=465056 &p=3180683.

[15] Drucker, J, "Intro to Digital Humanities – DH 101", UCLA Center for Digital Humanities, http://dh101. humanities.ucla.edu, 2013.

[16] "Data Never Sleeps 6.0", Domo, http://www.domo.com/ learn/data-never-sleeps-6.

[17] Coughlin, T, "Metadata is a Communication Channel to the Future", Forbes, Aug 3, 2017.

[18] Berman, J, "CES 2019: Think Analytics Talks Metadata, Viewer Behavior Solutions for the Viewing Experience," MESA Smart Content News, Jan 10, 2019.

[19] O'Neil, L, "TiVo Renews Multi-year Patent License and Signs Global Metadata Deal with Minerva Networks", Business Wire, Jan 9, 2019.

[20] Berman, J, "CES 2019: Gracenote Reports Strong Early Interest in Next-Gen Descriptive Metadata Solution," MESA Smart Content News, Jan 11, 2019.

[21] "Cloud Video Intelligence – Precise Video Analysis", http://cloud.google.com/video-intelligence/.

[22] Li, F, "Announcing Google Cloud Video Intelligence API, and more Cloud Machine Learning Updates", Google Cloud, Mar 8, 2017.

[23] Biskupski, B, "Powerful Image Analysis with Google Cloud Vision and Python", https://www.smashing magazine.com/2019/01/powerful-image-analysis-google -cloud-vision-python/, Jan 9, 19.