

M&E JOURNAL

Media & Entertainment
Strategies. Solutions.

It's our content... and our industry's responsibility to protect it — from leaks, breaches and hacks that threaten the underpinning of our business.

WINTER 2017-18 \$25

SECURITY

ISSUE

M&E security experts offer insights on the best practices and latest tools to confront today's content protection problems.

P. 17

Smart Content

By connecting information, technology and people, content can be made "smart"... and more profitable. P. 49

Wise Workflows

From new tools to new standards, the industry is finding ways to simplify the digital supply chain. P. 78

Insider Insights

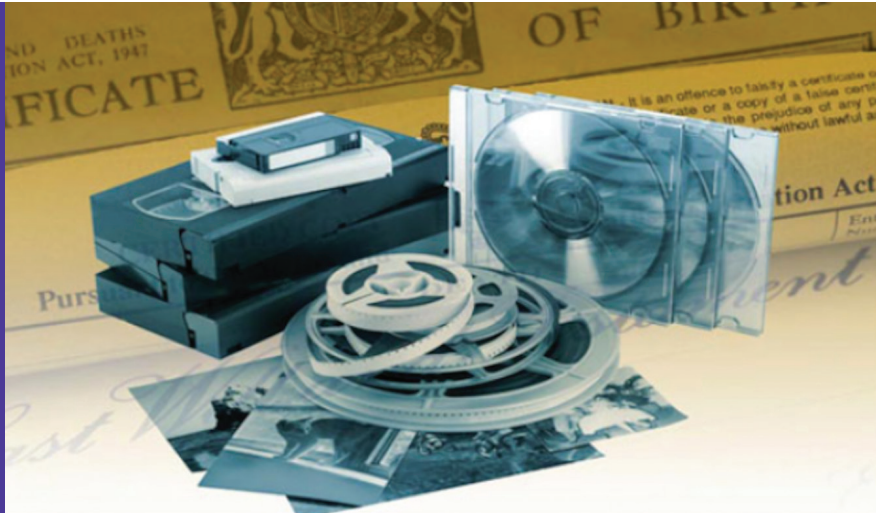
The state of IT hiring, women in technology, and the move away from leased set-tops. P. 100

MORE INSIDE:

Managing Static and Active Metadata in Support of Content Preservation

*Your investment
will only be as valid as its
metadata*

**By Linda Tadic, Founder
and CEO, Digital Bedrock**



Abstract: Digital content being preserved for future use requires careful management of the assets. Equally important to preserve are the potentially hundreds of metadata elements describing the digital object's creation, characteristics, environment, and sustainability that should also be retained along with the assets. Encapsulating critical metadata in a container along with the objects is one metadata preservation method, for example using AXF, TAR, or ZIP. Objects that are actively managed for preservation, however, will also have metadata that can change over time. This article outlines the metadata types that can be considered static or active, with recommendations on metadata management methodologies for preserving the metadata along with their related objects.

Every production requires an investment of money, time, and emotions. When the production is complete and the work released, the content created with that investment should receive as much care as the effort that went into producing it.

With so much invested, the content should be preserved, so it can be used in the future, whether for re-release, re-mastering, or restoration. If this investment is acquired by another entity, the new owner will likely require delivery of all masters and materials needed for a future release. Not everything need be preserved, but the content critical to perform any of these future actions should be carefully managed.

In today's digital reality, this means preserving not just the "essence" files or assets, but also metadata about the files and their creation, use, and preservation.

Many interpret the word "preservation" to mean backup: store files on servers, tape, or external hard drives in more than one copy, and then ignore them. Maybe the files will have the production title in the file name or directory name, and maybe there's a record about the storage media (not the files themselves) in a database or management system. But this isn't preservation. Digital objects (the combination of files and metadata that together make renderable — i.e. playable — content) must be actively managed to be properly preserved, and their bit health must be checked ("fixity checks") over time to be sure bits haven't flipped since put to storage.

They must also be migrated to new storage media every five years or so, with validation that all files are truly what they are thought to be. Obsolescence vulnerabilities must be monitored so that if a proprietary format is in danger of no longer being supported by software, it can be transcoded to a supported format. All these actions and monitoring requirements for digital preservation depend on extensive, detailed metadata.

To preserve digital content, metadata on the digital object's creation, use,

To preserve digital content, metadata on the digital object's creation, use, and preservation actions must be captured, retained, and added over time.

and preservation actions must be captured, retained, and added over time. Some of this metadata is “static,” meaning it will not change over time. Static metadata can be bundled into a container format like AXF or even TAR. In this scenario, any metadata is stored with its respective files, creating a self-contained digital object. Other metadata is “active,” and will change or be added to over time. Active metadata must be added to a database, not the container it is related to, since once the AXF or container file is created with its own checksum, that digital object can't be changed.

Static metadata

Static metadata is information that won't change, covering:

Creation history: The object's creation history is used to verify its authenticity. This will include descriptive metadata that won't change over time (i.e., information on the content, such as title, director, producer, cast, etc.), as well as information on the digital object itself.

- **WHO** created it (Who authorized creating the digital content? Who created it – vendor, artist, in-house?)
- **WHAT** created it (System, software, hardware, camera. Some of this metadata can be pulled from the file)
- **WHEN** (Date created and dates modified)
- **WHERE** (Geographic location. This can be GPS data from a camera.)
- **WHY** (What's the purpose or function of the object? Is it camera RAW? DI? Master?)

Technical characteristics and embedded

creation metadata: This metadata is gleaned from within the file. Technical characteristics include metadata created by the creating hardware or software. Think of a camera's EXIF data. Embedded metadata is added to a file by

a person, using software on the set or in the editing room. Technical and embedded metadata is unstructured metadata. In other words, since it's hidden in a file, it can't be searched in a structured database unless first extracted and indexed.

Technical metadata provides information on how the file was created (camera, software, lens, etc.) and its characteristics (size, bit depth, codec, format, etc.). These characteristics, with much of the validation information coming from the file header, are used to empirically validate that a file is what it is meant to be: vitally important information to know to transcode the format in the future, as well as to be able to monitor for obsolescence vulnerabilities.

Basic technical metadata commonly extracted from digital asset management and other management systems includes approximately a dozen metadata elements, including file size, format, codec, compression, bit rate, bit depth, color, frame height and width.

Extended technical metadata can run into the hundreds of data elements, depending on the nature of the file: camera model, UMID, lens, focal length, etc. Since most systems don't automatically extract this metadata, a system must be configured so the metadata becomes indexed and searchable.

Embedded metadata added by a human usually consists of content identification: title, director, camera operator, scene, take, etc. Editing systems can extract this metadata, but it is not usually pulled into digital asset management systems (DAMs) for automatic retrieval.

Original preservation information: To preserve this content in the future, information that provides the context for the file and its digital “fingerprint” must be captured.

The original directory structure for where the digital object was stored before being archived provides clues to the object's context, purpose, and relationship to other objects.

A hash (checksum) should be created immediately after the file is created, or if not, as soon as possible. The initial checksum is validated over time, and every time the file or object is disseminated, to verify its bit integrity. Hash information to be retained should be the algorithm used (MD5, SHA-256, SHA-512) and the value.

A digital file must be validated to ensure it conforms to a specified format. Technical metadata captured (see previous section) helps to validate a format.

Active metadata

Active metadata includes information that can change over time, if an active preservation policy is implemented. Since it is active, it lives in a database outside the digital object. The active metadata can reference the entire digital object, or individual streams or files within the object.

Managing this metadata outside the file requires persistent linking between the metadata record and the files. Care must be taken when migrating metadata between systems that this link isn't broken.

Critical active metadata includes:

- Obsolescence vulnerabilities (monitoring format deprecation, software/OS support, etc.)
- Preservation events (fixity check, transcode to a preservation format, media migrations)
- Access events (delivery, transcode for fulfillment purposes)
- Storage locations (which exact storage media (LTO tape, server, HDD), and where it is physically located (cloud, physical storage))
- Descriptive and rights metadata (descriptive or “content” metadata can be added over time; rights metadata can change)

Bundling digital objects and active preservation

Archiving digital objects in container for-



Linda Tadic has 30 years' experience leading preservation, metadata, and digital production operations at organizations such as ARTstor, HBO, and the Media Archives and Peabody Awards Collection at the University of Georgia. Currently an adjunct professor in UCLA's Media Archival Studies, she consults and lectures on digital asset management, audiovisual and digital preservation, copyright, and metadata. She is co-author of the book "Descriptive Metadata for Television: An End-to-End Introduction." ltadic@digitalbedrock.com

mats such as AXF and TAR is a static approach with advantages and disadvantages. Many digital storage system users may not be aware that their system is creating AXF containers, so if you don't know, you should find out.

The advantage: Useful for long-term preservation of static objects, since the object and metadata about the object are retained in one container. This is valuable for transmitting full packages of data.

The disadvantage: Ongoing "active" preservation work is made more complex. A system must parse out the individual streams to monitor fixity and obsolescence. With a bundled container like AXF, it isn't possible to add metadata to the container file once the container is

built. The active metadata can be added to a database or system, but that negates some of the value of the container file. They can still be used in tandem, but the user must be aware that should the AXF be transmitted to another user, the active metadata must be pulled from the system.

Another approach is to keep all metadata active in a system. That way, active preservation management can be more easily performed. This can be accomplished so long as care is taken that all files and metadata related to the full digital object are properly linked together

Whichever approach is selected, it's critical that extensive metadata — both static and active — be captured and preserved along with the digital content.

In short, your investment will only be as valid as its metadata. ■



DIGITAL BEDROCK

AFFORDABLE, SECURE, MANAGED DIGITAL PRESERVATION SERVICES

Developed from experience and familiarity with the needs of media entertainment professionals to meet your digital preservation requirements.

WE ARE YOUR STAFF AND INFRASTRUCTURE:

- NO software licenses
- NO annual subscription fees
- NO staff training
- NO large-scale hardware or infrastructure investments

DIGITAL BEDROCK PRESERVES CONTENT THROUGH:

- Detailed file analysis
- Extensive metadata extraction
- Active obsolescence monitoring
- Scheduled SHA-512 fixity checks
- Offline storage with built-in future migration
- Three copies geographically dispersed

Supporting complex and proprietary formats in an open architecture environment.

INFO@DIGITALBEDROCK.COM | 888.938.7386 | WWW.DIGITALBEDROCK.COM