IRON MOUNTAIN ENTERTAINMENT SERVICES

WHITE PAPER

LONG-TERM MEDIA ASSET PRESERVATION

WHERE ARE YOU ON YOUR CLOUD JOURNEY?

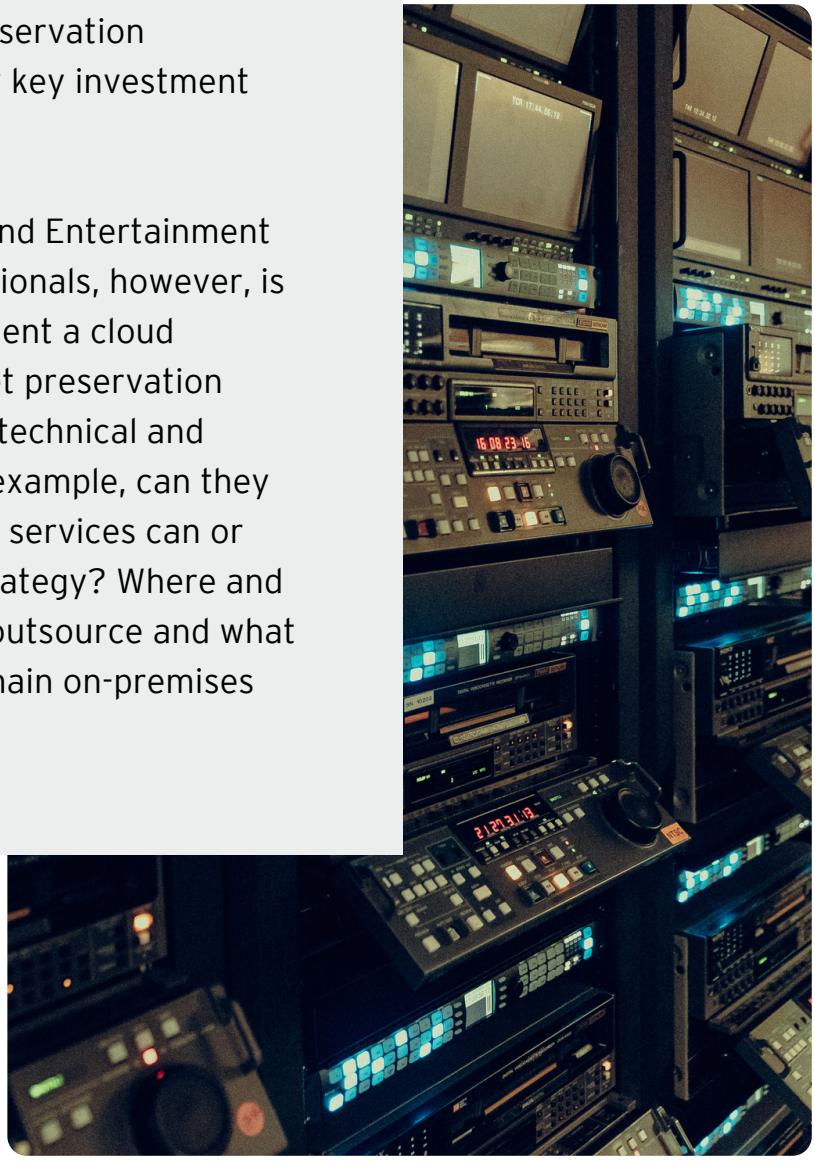
WHITE PAPER

TABLE OF CONTENTS

INTRODUCTION

The ongoing growth of cloud computing continues to challenge the role of legacy solutions and services in almost every industry, with long-term preservation use cases across the Media and Entertainment landscape among those now firmly on a pathway of digital transformation. From performance and cost to security and reliability, cloud-based services are set to compete with existing preservation technologies across every key investment criterion.

The challenge for Media and Entertainment organizations and professionals, however, is how to design and implement a cloud computing long-term asset preservation strategy that meets their technical and financial needs. How, for example, can they assess where cloud-based services can or should fit into a hybrid strategy? Where and when is it appropriate to outsource and what content assets should remain on-premises and for how long?



WHERE ARE WE TODAY?

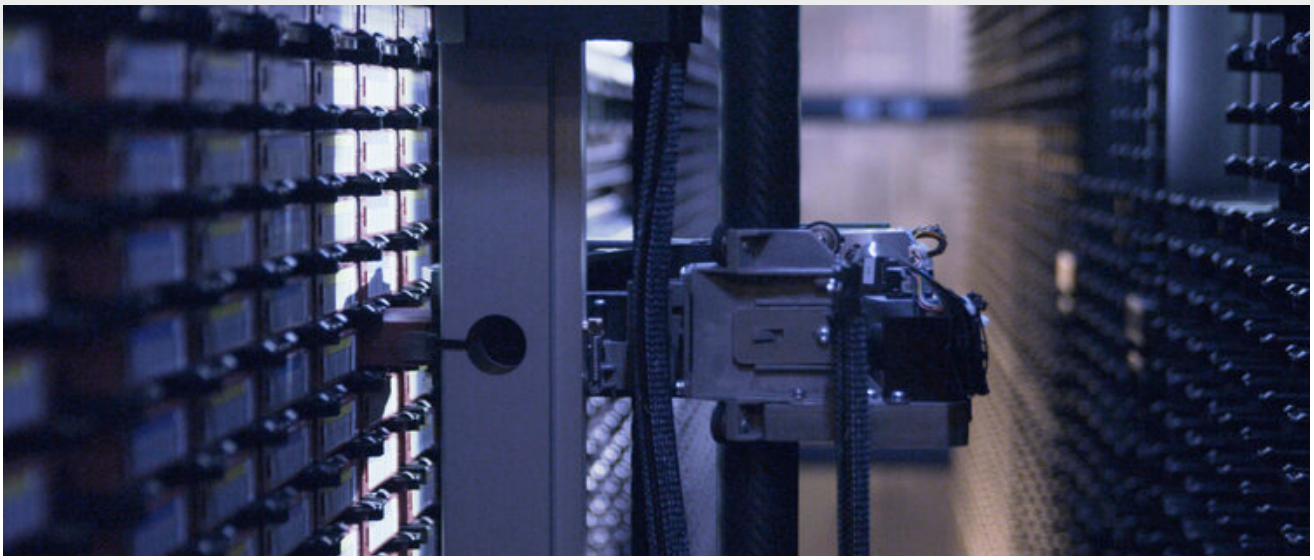
While the adoption of digital technologies has delivered widespread benefits for media organizations the world over, implementing long-term media asset preservation in the digital age is not without its significant challenges.

Indeed, the continuing reliance on well-proven legacy technologies and processes is limiting the extent to which Media and Entertainment organizations can embrace the full performance and economic benefits of key digital computing trends – particularly the adoption of cloud computing technologies and services.

But why is this important in the context of long-term media asset preservation?

Put simply, cloud computing has become one of the most important and transformational technology movements of the digital age. Organizations the world over have turned to its versatile, outsourced service provider model to deliver a more affordable, reliable and high performance technology infrastructure strategy. From the tactical adoption of services to 'all-in' fully outsourced models, shifting the burden of data and application strategy away from under-resourced in-house teams to specialist partners has seen cloud rapidly head towards becoming a trillion-dollar industry.

For Media & Entertainment organizations, however, this momentum is less well-established. In particular, the adoption of cloud computing technologies for long-term preservation use cases is not simply a matter of switching from one infrastructure strategy to another. Indeed, digital assets are notoriously difficult to preserve over the long term, not least because they were never originally designed with this objective in mind.



WHERE ARE WE HEADING?

Despite these challenges, many Media and Entertainment organizations agree that they are on a journey to the cloud but that most technology solutions need to advance further to deliver against their requirements.

In particular, any cloud solution for deep storage/long-term preservation must be economically competitive against the cost of LTO libraries (and their constant and costly migrations). More specifically, key processes such as fixity tests – confirming that a digital file has remained unchanged over time – represent the acid test of the digital preservation process, and currently must be executed once or twice a year to ensure preservation integrity.

As archivists learn more about how the cloud solutions perform on these tests and more data about durability is collected, it may become practical to adjust the frequency and associated overheads associated with carrying out these tests. In doing so, organizations will have greater confidence in the reliability of cloud services and solutions as a preferred long-term preservation strategy. To perform effectively, cloud-based fixity tests must also be integrated into the cloud storage architecture to deliver an efficient and cost-effective solution. In many current and existing cloud-based offerings, however, they remain highly resource-intensive and cost-prohibitive to run. But, the implementation of cloud-native technologies such as low-cost serverless APIs, or the potential availability of 'archive-specific cloud storage' have practical potential to resolve any storage integrity error detection into a file-based error indication.

With these challenges gradually being addressed, many Media and Entertainment organizations are likely to adopt an iterative approach to adopting cloud for long-term preservation. Indeed, the introduction of cloud services and solutions into this ecosystem need not be an all-or-nothing decision. In particular, a hybrid approach gives archivists a roadmap to adopt the cloud in a more phased manner (e.g., cloud storage as virtual tapes, loud + LTO libraries, multi-clouds, etc.).

There's no doubt that despite the general appetite to embrace the benefits of cloud computing, many industry professionals would feel more comfortable and confident with a gradual and well-thought-out transition.

The bottom line is that by addressing these challenges, organizations will be able to realize a series of crucial cloud benefits, including:

INFRASTRUCTURE

Ensuring infrastructure is maintained against obsolescence and performance issues.

ALIGNMENT

Cloud ensures long-term media preservation is better aligned with production as it also moves to the cloud.

SCALE

A core benefit of cloud computing in general, it is inherently scalable, ensuring resources can be quickly allocated to better meet demand.

STORAGE

Cloud solutions offer storage options that are difficult to implement on-premises, such as redundancy and geographic separation.

ACCESSIBILITY

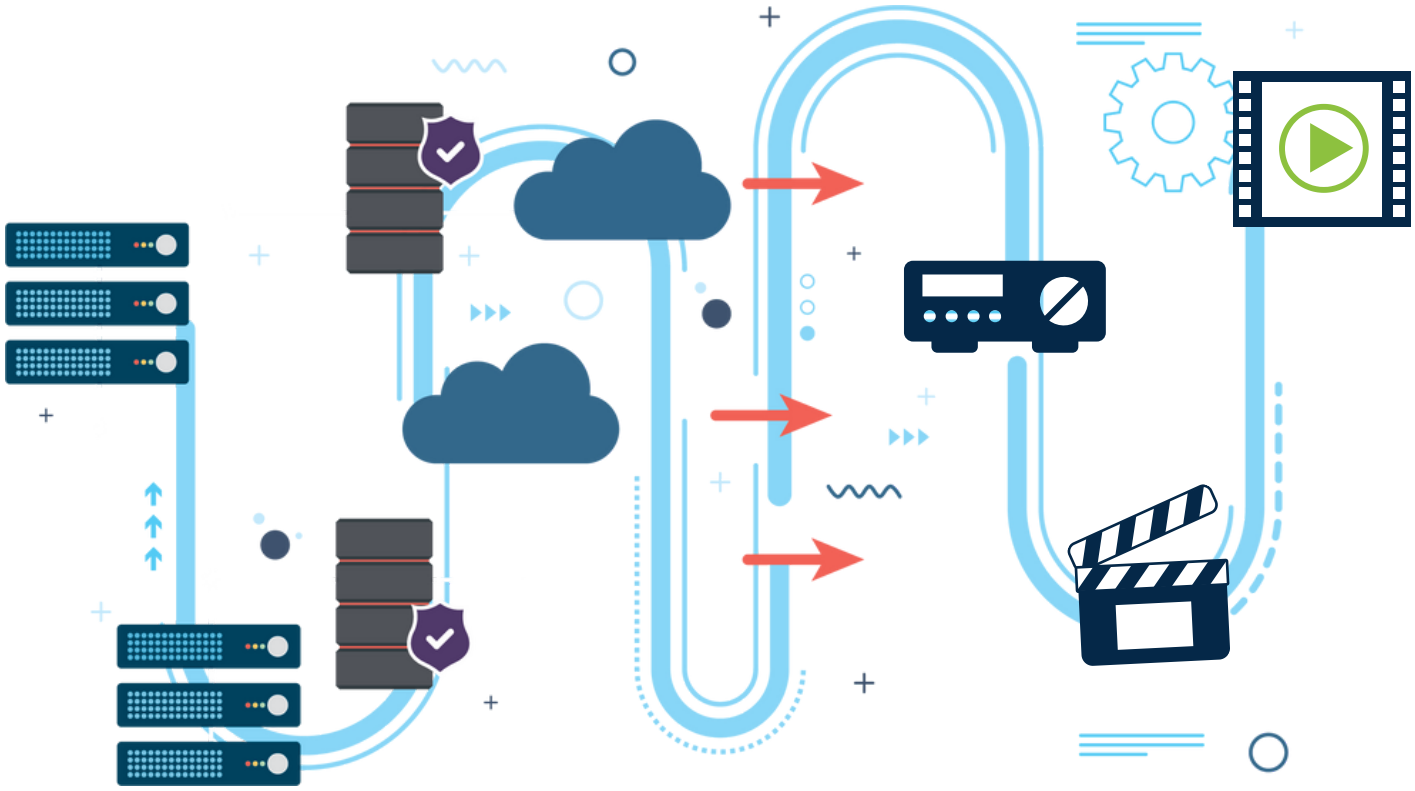
Cloud offers significant potential for improved/faster accessibility via storage structuring, together with access to computing resources for constant metadata enrichment.

HOW TO ASSESS THE ROLE OF THE CLOUD IN A DATA PRESERVATION TECHNOLOGY STACK

Since LTO tape and cloud storage have different implementation characteristics, it makes sense that their durability characteristics would also be very different. As a result, there is increasing justification for reconsidering the precise modalities of fixity checks as they relate to cloud storage (such as frequency and general methodology).

Unfortunately, there is a dearth of experimental data on how long cloud storage will last. The CSP provides "numbers of 9s" as targets for each of its storage classes, but archivists really need experimental proof, including details on the methods used to calculate the numbers and whether they hold up in real-world applications.

In order to properly determine how we may achieve fixity in the context of cloud storage, particularly in terms of frequency, a complete analysis of the data durability of cloud storage using fixity methodology as a measurement instrument is required.



HOW TO BUILD A CLOUD-CENTRIC LONG-TERM PRESERVATION STRATEGY

To ensure Media & Entertainment organizations and asset preservation objectives can fully benefit from the capabilities of cloud, adopting a gradual and smooth transition process is key.

In particular, any cloud-based archival implementation will need to effectively and reliably implement fixity checks. Without these, organizations simply won't be able to build the levels of confidence and trust that are fundamental to investing in long-term preservation infrastructure.

Moreover, first-generation cloud-based archival implementations should allow for rigorous comparison of fixity results between different storage classes. In doing so, fixity reports can be used over time to determine the level of durability of various cloud storage classes.

In addition, cloud data integrity checking is done at a lower abstraction level, and the data integrity algorithm has no visibility into the filesystem structure, whereas fixity checks are performed at the filesystem level in the storage stack (whether it be for a single file or group of files). There isn't a compute-free way to get from cloud data integrity verification data to the kinds of file-based fixity reports that are currently accepted in the business. Although both strategies aim to ensure data persistence, they work on different storage concepts, making it hard to switch between them without any disruption.

This implies, in particular, that more conventional techniques modified for the cloud must be used to establish file-based fixity.

For example: - Using cloud computing resources to build a method to access the archive files and carry out the hash computation and comparison (which will be visible as such to a cloud computing resource as this is the layer at which they work).

Another option is to utilise the specialized services offered by cloud vendors that carry out the same function (after validation). Alternatively, the idea of serverless hashing, which is a minimal-footprint implementation of a process, is of relevance. Although there may be some execution time and process footprint restrictions with serverless computing, these are more than sufficient for tasks like computing a file digest as part of a fixity procedure.

A bolder call to action would be to construct "archive-specific cloud storage," which would enable the resolution of any storage integrity fault detection into a file-based error indication similar to a fixity fail.

It is important to remember that existing fixity verification techniques are built on current knowledge and experience with the longevity of the underlying storage media, such as LTO tape. However, it is unrealistic to expect cloud storage to share the same characteristics, and each cloud storage type (online, nearline, deep, multi-region, etc.) is actually likely to have unique durability properties. Therefore, to ascertain the precise methodology and frequency needed, a thorough examination of the unique durability characteristics of each environment will be necessary.

Introducing cloud into asset preservation offers industry professionals several options, including:

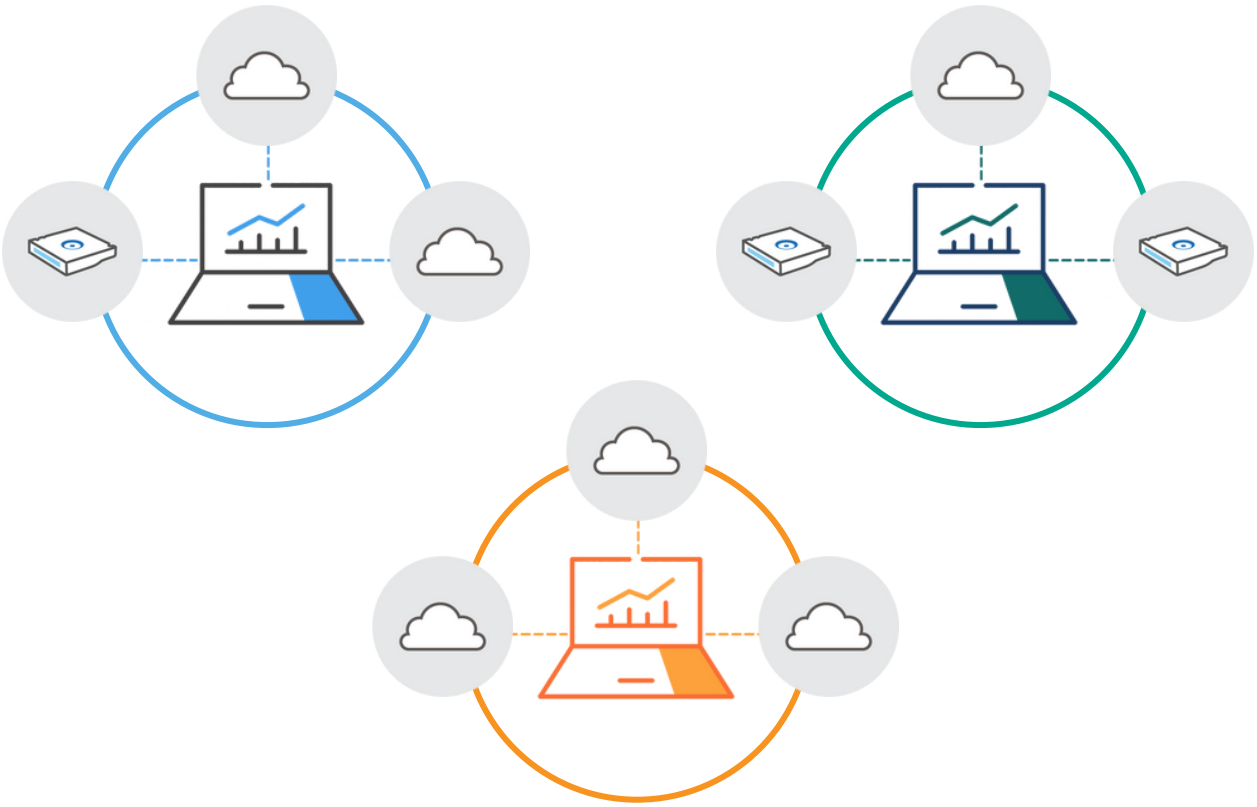
A HYBRID APPROACH

in this situation, one tape copy is replaced with cloud storage (in a 3 copy scenario). The benefits of this architecture is that it retains local archives, while also allowing for experimentation.

Integrating cloud storage into an archive solution does not have to be all or nothing. In reality, using the cloud as one of the archive pieces in a multi-element approach could be a great strategy, especially in the early stages. For example, make two tape copies and one cloud copy rather than three tape copies.

This can be accomplished by hiding the cloud copy behind a virtual tape library, but there may also be the option of shifting the archive's cloud section to cloud object storage and working on resolving the particular issues that arise in that situation (namely, establishing a clear and stable connection between archive elements and Cloud objects).

In general, a hybrid classic+cloud archive solution is a secure way to integrate the cloud into the archive, allowing for appropriate testing of the technology without jeopardising the entire archival strategy.



A CLOUD-BASED VIRTUAL TAPE LIBRARY

This enables the use of an existing archival application, providing a platform for precise comparison between cloud and physical tape. The challenge here, however, is that this approach can lead to architectural complexity and makes the use of cloud features more difficult.

Virtual tape library (VTL) technology can be a great starting point for gradually integrating a cloud storage layer into an existing preservation solution based on physical tape libraries.

There are a number of solutions on the market right now that make it possible to replace actual tapes with cloud-based (or other storage class-based) 'virtual tapes', and they can be used to maintain all of the front-end tape workflows and apps used by contemporary organisations. The same procedures are followed throughout, and the tape abstraction layer can deal with various formats in a transparent manner. Therefore, with just minor changes from an original, "all physical tape" deployment, it is possible to create an archive deployment that uses both Practical Cloud Archive physical and virtual tapes. The virtual tape copy is really an additional asset copy, with concepts like asset repository, multiple copy management, and integrity checking continuing to apply in the same way.

Today, a lot of CSPs offer services that make changing the endpoint of an archive workflow rather simple. Additionally, a wide range of managed services have evolved that can help with the transition of current archives to cloud-based alternatives.

The fact that a virtual tape library is an emulation layer by definition, adding complexity to the overall design, is one potential drawback of utilising one. A VTL deployment also has some limitations. For instance, a VTL may make it more difficult to use checksum services offered by CSP that are designed for object storage. Therefore, in order to do a checksum validation in the cloud, the fixity processes themselves must either be virtualized or containerized, or the virtual tape's data needs to be retrieved to where the fixity validation process resides.

The ability to continue using the fixity tests and fixity reports now used to evaluate tape-based archives on both sides of the archive is a particularly intriguing aspect of a hybrid, physical+virtual tape library solution (cloud and physical tape).

It is, therefore, possible to create an exact apples-to-apples comparison testbed for a comparison study of fixity in the cloud vs fixity on an LTO-based archive, as stated above, by selecting a tape work unit on the cloud side that corresponds to what is utilised on the physical side. The provision of reliable experimental data on cloud storage, appropriately related to the same data entities currently utilised in fixity verifications, would meet a fundamental need in terms of verifying cloud storage as an archive.

This method's disadvantage is that it's likely to be a lengthy procedure. Both LTOs and cloud storage have strong durability properties, and it is anticipated that data failures will be rare and appear slowly. In light of the fact that this strategy would be an application of the longitudinal study previously stated, it would be a multiyear procedure.

CLOUD-NATIVE STORAGE

This approach requires adapting the application stack and places a special focus on packaging, asset repository (different asset instances/packages). It also opens up the option of using serverless computing – a cloud computing execution where the cloud provider allocates machine resources on demand – to deliver effective fixity capabilities.

Performing fixity checks on assets stored in the cloud has clear operational advantages. For instance, cloud storage is often available alongside cloud computing resources, some of which also provide optimised processes to do checksum computations – a crucial step in fixity tests. Therefore, once adequate process validation has been accomplished, it makes sense to move fixity checks to the cloud.

Another intriguing feature of a cloud archive is that it might occasionally be easier to access than a tape archive. This is particularly true for older tape formats where access to drives can be a problem, but it can also just be a matter of time to access being faster with cloud storage. However, that is not always the case, as reading a tape from a local library can be faster than cloud retrieval from deep storage, although the jury is still out on which method offers the fastest access.

In any case, it is entirely conceivable to imagine using the cloud, rather than an existing tape set, as the source for archive migration to a new tape set in scenarios where a tape copy is necessary. This holds true provided that a strong level of trust in the integrity of cloud storage for archive purposes has been achieved in advance.

Likewise, an organisation can use cloud archival elements as a source to execute fixity checks before migrating to new tape media, as opposed to retrieving physical tape archive elements.

SEARCHABILITY CHALLENGES

Searchability is the single most common problem associated with archives and long-term preservation. Understanding the specifics and importance of a given asset at the time of development is challenging enough, and it should not be surprising, therefore, that after any significant passage of time, these challenges tend to grow.

Generally speaking, having a naming convention for files is no longer considered "good enough", because if you don't know which version of an asset was approved, for instance, knowing that you have 12 versions of it is useless. This means there are some key questions to consider: Are you certain, for example, that the version used for the final output was the most recent version in the collection? Which behind-the-scenes videos were approved for usage by the talent? Who is the asset's owner?

In addition, details relating to decisions made during the lifecycle of a piece of content may be required when returning to a preservation asset repository, depending on how a preservation store is set up. It is crucial, therefore, to consider the information you will need to supply depending on your production, its associated strategy, and its preservation aims.

In the end, in order for an archive to be complete, you must either save assets whose names contain all the information a user might possibly need to use them or store the information required to understand the data it contains.

MULTI-CLOUD ARCHIVE

The advantage of this approach is that it avoids the issue of single cloud lock-in, where users are unable to access the choice of provider and solution because it is technically difficult or economically prohibitive to change. Additionally, this strategy offers the possibility of 'on-premises cloud storage', where hardware that is related to cloud services or activities is nonetheless located on-site at the client's physical business location.

This strategy forgoes conventional archive methodologies and depends solely on the data protocols employed by the cloud vendors. It also assumes that the likelihood of losing the integrity of the same bits from the same files is extremely remote across clouds.

One intriguing challenge presented by all the multi-entity storage solutions, such as LTO + cloud and LTO + Polycloud, is the requirement to maintain the various entities properly in balance with one another, not only in relation to data sync but also to ensure proper location equivalences between the various instances of a given asset. For instance, the path of a given file may not necessarily be the same across multiple cloud vendors, which then poses an interesting challenge as it will require a tape index in addition to a path, at the very least.

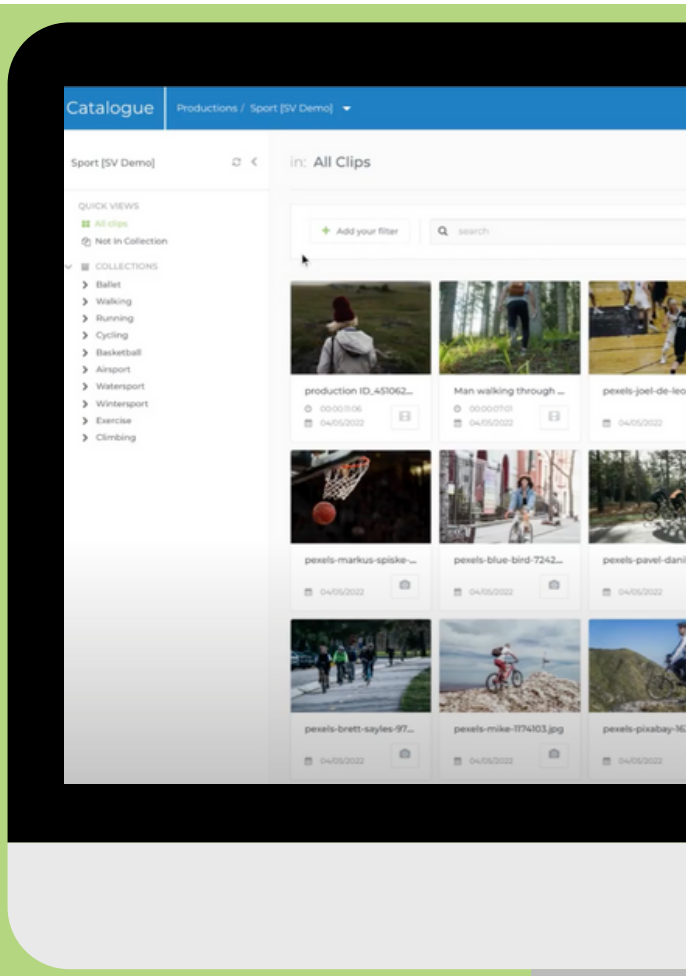
This raises the idea of a central archive asset metadata repository that acts as a type of multi-architecture location clearinghouse that stores fixity data in addition to asset metadata.

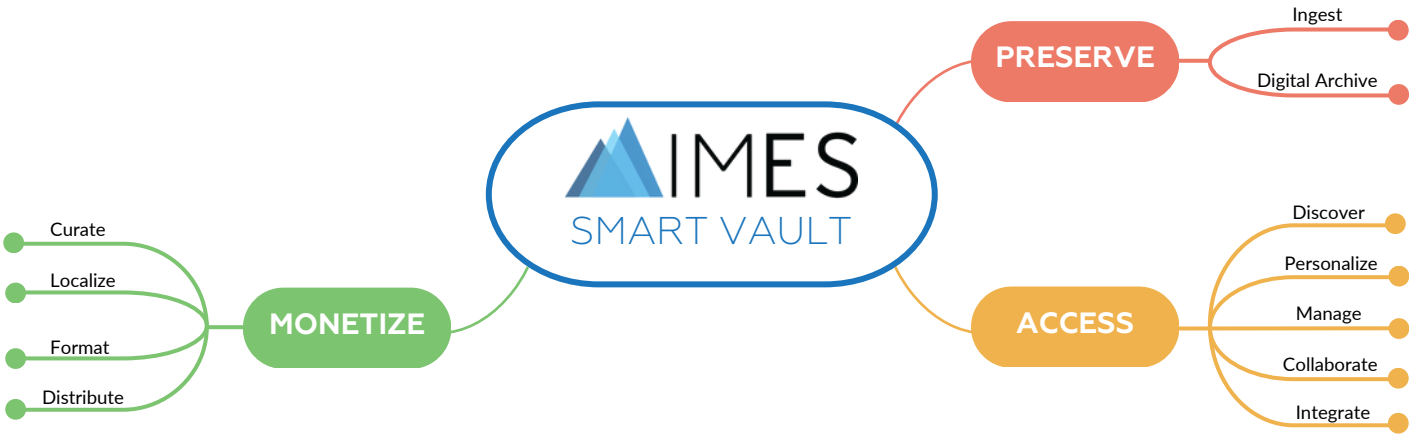
THE IMES APPROACH

Smart Vault from Iron Mountain Entertainment Services (IMES) Media Asset Management solution provides broadcasters, production companies, brands, sports rights holders, and any media organisations involved in the production of content with a powerful solution for content management, storage, and archive.

It enables content owners to use the latest cloud-native technology to enable ease of access to their content – whenever and wherever they want it, providing swift visibility and easy access to large volumes of media with an inbuilt video player.

As well as streamlining and bringing multiple efficiencies to the elongated processes used in managing media assets, powerful AI/ML capabilities enable users to enrich the media stored in the platform, providing monetization opportunities for existing and future archive content. Additional functionality allows users to share media, contribute to the platform, and create custom workflows for media production and distribution, increasing agility through the chain.





CASE STUDY | THE DIGITAL TRANSFORMATION OF CANAL+'S HISTORIC TAPE ARCHIVE

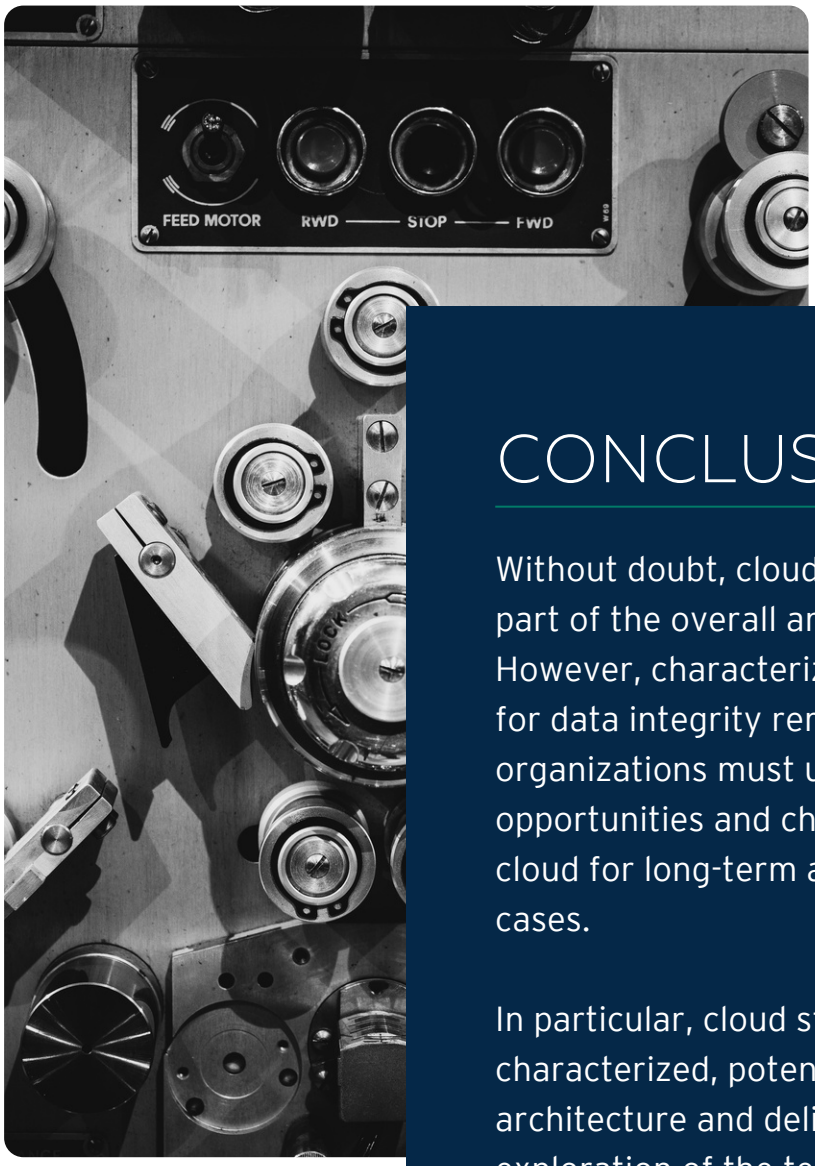
CANAL+ is a French premium television channel that boasts more than 21 million subscribers. The channel was the first to exclude paid advertising in France and today offers several kinds of programming. Iron Mountain Entertainment Services (IMES) has stored CANAL+'s half million archived tapes for the past 20 years and was recently awarded the work to digitise this historic collection to support the expansion of their streaming services.

Specifically, the challenge was to digitise, index and migrate a large portion of CANAL+'s massive collection to make it easily accessible and searchable as well as provide urgent digitisation on an as-needed basis.

To meet these needs, IMES digitised 100,000 tapes over three years and ingested content and metadata to IMES' Smart Vault for secure and easy access and distribution, and provided emergency digitisation services for immediate use.

As a result, CANAL+ is now able to repurpose valuable content for its 21 million viewers due to IMES' partnership as a physical and digital content repository that ensures CANAL+'s physical content is safe and secure, while its digital copies are easily accessible and distributed for future monetization.

To read the full IMES/CANAL+ case study, click here.



CONCLUSION

Without doubt, cloud has huge potential as part of the overall archival landscape. However, characterization of cloud storage for data integrity remains a concern and organizations must understand the opportunities and challenges of implementing cloud for long-term asset preservation use cases.

In particular, cloud storage will need to be characterized, potentially via hybrid architecture and deliberate, gradual exploration of the technology. In doing so, organizations can focus on delivering a win-win, whereby they can have all the performance, versatility and cost benefits enjoyed by cloud-centric businesses around the world, but geared towards their unique needs.

IMES

IRON MOUNTAIN ENTERTAINMENT SERVICES

WRITTEN BY:

DENIS LECONTE

VICE PRESIDENT OF TECHNOLOGY,
IRON MOUNTAIN ENTERTAINMENT SERVICES

Denis Leconte has over 25 years' experience working for Iron Mountain, NASA, Disney, Sony, and several post production facilities, designing and deploying database and media handling systems and large scale, high-performance computing and storage architectures. He has over 15 years experience in post production and digital media and data lifecycle management. At Iron Mountain Entertainment Services, he oversees all technology issues over several facilities that cover North America and Europe, as well as many special projects dealing with media and data preservation and restoration. Leconte has won an Academy Award for Scientific and Engineering achievement in the field of film preservation, and more recently has served as a Surrogate with the Academy's Science and Technology Council. He also co-chaired the 2021-2022 session of the ETC' (Entertainment and Technology Center) Archive Working Group. You can contact Denis directly at denis.leconte@ironmountain.com

ABOUT IRON MOUNTAIN ENTERTAINMENT SERVICES

Iron Mountain Entertainment Services (IMES), the media and entertainment division of Iron Mountain Incorporated®, is the go-to physical and digital media archiving service for the media and entertainment industries. IMES partners with clients ranging from the broadcast, film, music and sports sectors to brand archives and individual artists' collections and estates, to protect and activate media archives to preserve our collective cultural heritage. A global leader in media preservation, restoration and archive management, IMES offers industry-leading archive, storage, chain-of-custody processes and physical and digital content preservation solutions to ensure the complete protection of some of the world's most treasured and iconic assets.

US-WP-011122A