

## Conferencia Magistral: Criterios a considerar en la definición de un sistema de gestión y almacenamiento masivo digital

**Stefano Cavaglieri**

**Experto en tecnología de audio y digitalización, fonoteca nacional**



Esuchar conferencia:

[Original](#)

[Traducción](#)

Suiza. Quiero agradecer a la Fonoteca Nacional de México por haberme invitado a venir a este lugar tan precioso y al Seminario. Es un honor para mí estar aquí y un orgullo compartir mi experiencia con todos ustedes.

Mi presentación va a ser técnica y difícil de entender para algunos de ustedes, incluso, es difícil para mí preparar una presentación como ésta, así que por favor discúlpenme si hay algunas partes que necesiten más explicación.

Comencemos por el principio y para eso quiero hacerles una pregunta: ¿qué es un Sistema de Almacenamiento Masivo Digital? Si googlean “sistema de almacenamiento masivo digital”, lo más probable es que se desilusionen. Muchos proyectos en esta área no cuentan con información descriptiva sobre la terminología del campo, por lo que usan el acrónimo: DMSS. La situación empeora con los resultados de la búsqueda, por ejemplo: sistema de vigilancia médica, sistema de distribución maestro, software descriptivo dual, sistema de normas móviles. Después de este montón de resultados, al final sale: Sistema de Almacenamiento Masivo Digital.

Hagamos el intento por hablar acerca del contexto y definir qué es lo que necesitamos para identificar el contexto adecuado para la documentación de los elementos en forma digital. Primero se requiere de una estructura que sirva de depositario y que tenga la capacidad de resguardar ciertos objetos digitales; es decir, una secuencia lógica de ceros y unos. Además de eso, tenemos que decir que el sistema de almacenamiento típico, para ser efectivo, debe constar de diferentes elementos: controlador de flujo, sistema de administración de datos y sistema para difusión. Una vez que se tienen todos estos elementos, podemos hablar de un “sistema de archivo”, éste es el término que prefiero.

La palabra archivo se utiliza en muchos contextos pero ¿qué es un archivo? Según el diccionario Merriam Webster un archivo es “el lugar en donde se mantienen documentos o archivos históricos, o un repositorio o colección de información especialmente”.

Quiero agregar algo para tener un panorama más completo de lo que es un archivo. La Sociedad de Archivistas Estadounidenses define un archivo como “los materiales que se crean o que recibe una persona, familia u organización privada o pública al realizar sus actividades debido al valor de la información contenida en esos materiales”; por lo tanto, los materiales y su contenido son responsabilidad del creador. Aquí se aplica el principio de proveniencia, orden original y control colectivo para definirlos como registros permanentes.

También se dice que un archivo es una división dentro de una organización, dicha división debe resguardar los registros valiosos para la misma; sin embargo, pudiese ser una institución que recopila los registros de individuos como la profesión, de familias o de otros organismos, todo esto para el ordenamiento de dichas colecciones en el edificio donde se guardan. ¿Todas estas definiciones nos ayudan en nuestra investigación?

Aclaremos las dudas. ¿Cuál es la regla maestra de todos los archivos? El modelo de referencia para el Sistema Abierto de Información en Archivos (OAIS) por sus siglas en inglés, es una recomendación que se publicó para sistemas de datos guardados con diferentes normas, es un servicio de consultoría para sistemas de almacenamiento y archivos que posteriormente se convierte en lo que la ISO clasifica como código 14.721; 203.

El término archivo se utiliza para hacer referencia a diferentes sistemas y funciones de almacenamiento. Los archivos tradicionales se entienden como instituciones u organizaciones que preservan registros físicos elaborados especialmente para instituciones o corporaciones y que reúnen el material para ponerlo a disposición de los usuarios. Se hacen cargo de los registros y se aseguran de que el público los entienda. Su función es preservar el contenido y la autenticidad. Por tradición, los registros han sido libros, mapas, fotografías, películas y material que puede ser leído por el ojo humano o que pueden ser leídos por medio de algún dispositivo de ayuda para amplificar o escanear la imagen. El enfoque de la preservación de este material es procurar la estabilidad a largo plazo y controlar el acceso.

El crecimiento explosivo de los formatos digitales y de la información ha representado un reto para los archivos, no solamente tradicionales, sino también para otras organizaciones a nivel gubernamental y para organizaciones sin fines de lucro; dichas organizaciones se han dado cuenta de que está en sus manos la función de la preservación, actividad que anteriormente la hacían los archivos.

La información digital se pierde o altera con facilidad ya que está cambiando constantemente; en cuestión de un par de años muchos sistemas se vuelven obsoletos rápidamente, por lo tanto, las estructuras y formatos de almacenamiento para la preservación de la información presentan muchos retos. Ahora, la información se encuentra disponible al momento de la producción del original y las organizaciones, por consiguiente, necesitan tener una participación activa en los esfuerzos a largo plazo para la preservación de los materiales y deben utilizar los principios de OAIS para asegurar la preservación ¡A LARGO PLAZO! La participación en estos esfuerzos minimizará los costos de vida útil y aumentará la preservación efectiva de la información.

La explosión de la capacidad de procesamiento de las computadoras y los medios digitales ha cambiado en función del producto y del archivo. Ambos elementos ahora son responsabilidad de la misma entidad. Estos sistemas, conocidos como archivos activos, deben tomar en consideración los principios de preservación a largo plazo que se describen en la OAIS, ya que gran parte de las actividades de preservación pudiesen estar en conflicto con los objetivos de producción y/o difusión del material al público.

Es por eso que los diseñadores y arquitectos de los sistemas deben documentar las soluciones a las que llegan.

La función principal del modelo de referencia es facilitar el entendimiento sobre lo que se requiere para preservar y acceder la información a largo plazo; no hay que confundirnos con el almacenamiento de bits o con un almacenamiento digital con ceros y unos. Los principios definen las normas que se deben seguir para lograr la preservación a largo plazo. Un archivo OAIIS trata de preservar información para dar acceso a una comunidad designada y necesita ciertos requerimientos: la OAIIS debe negociar con los originadores cuál es la información adecuada; el nivel para asegurar la preservación a largo plazo y determinar, por sí misma o con otras organizaciones, cuáles son las comunidades designadas para entender qué información se está proporcionando; asegurar que la información que se va a preservar sea entendible de forma independiente para la comunidad específica, es decir, la comunidad debe poder entender la información sin necesidad de contar con la participación de los expertos que la produjeron; seguir las políticas y procedimientos de documentación para que la información se preserve y pueda enfrentar todas las contingencias y se pueda difundir como copias auténticas del original o que el original se pueda rastrear; la información preservada debe estar disponible para la comunidad designada; además, incluye a los archivos que deben manejar flujos y entradas periódicas de información, a los archivos con servicios de acceso sofisticados y aquellos que sólo manejan solicitudes sencillas.

La OAIIS tiene un modelo que reconoce la naturaleza altamente distribuida de los acervos digitales y la necesidad de tener instrumentaciones locales, políticas y estrategias para preservar la información; esto permite tener distintas organizaciones con responsabilidades diferentes de los archivos tradicionales para lograr la preservación.

Se espera que las organizaciones hagan el intento de preservar la información y que utilicen los conceptos de la OAIIS para lograr sus objetivos de preservación.

Para acercarnos a los criterios que debemos tomar en consideración para los Sistemas de Almacenamiento Masivo Digital, debemos desarrollar los siguientes conceptos:

las interfases entre los archivos OAIIS; las metodologías de ingreso o ingesta a los archivos; la inclusión de la información en los archivos; la entrega de fuentes digitales a los archivos; la entrega de metadatos de las fuentes digitales o físicas a los archivos; la identificación de fuentes digitales dentro de los archivos; recopilación de materiales a partir de las fuentes físicas y digitales; acceso de medios para reemplazar los sistemas de administración digital sin tener que volver a escribir en el medio, en medios físicos; migración de información entre formatos; prácticas recomendadas para el archivo; y los archivos acreditados.

Debemos evitar comparar a los archivos tradicionales con los digitales, hay diferentes razones por las cuales no se debería hacer esta comparación, algunas legítimas y otras no, pero son la base del análisis del diseño del Sistema de Almacenamiento Masivo Digital.

Entonces, la vida de todos los soportes audiovisuales depende de su estabilidad física y química, y de la disponibilidad de la tecnología de reproducción. Ésta también puede ser una

fuente de daño, por ello, la preservación audiovisual siempre requiere de la producción de copias que puedan servir como duplicados de preservación, esto en la jerga de los archivos digitales se conoce como “fuentes substitutas de preservación”.

La necesidad de migrar contenido a un nuevo sistema de almacenamiento se aplica a soportes de audio digital y audiovisuales originales que puedan estar en peligro o tengan una vida útil muy corta debido al hardware y software altamente sofisticados que se ofrecen en el mercado; después de algunos años de su introducción, estas fuentes se vuelven completamente obsoletas. Sin embargo, las mismas limitantes que se aplican al original se aplican a los formatos de preservación que requieren de reduplicación continua.

La preservación se puede continuar a través de una duplicación seriada en el dominio análogo, pero esto puede tener un efecto en la señal de audio y ocasionar un impacto negativo en las siguientes generaciones. El potencial que ofrecen estas fuentes substitutas para la preservación parece proporcionar un vínculo con el acceso y la preservación; sin embargo, las decisiones que se tomen en referencia a la resolución, formatos, soportes digitales y sistemas tecnológicos, deben contemplar límites de eficacia en la preservación digital, ya que las decisiones son irreversibles.

En lo que se refiere a la calidad del audio y video que se está codificando, es indispensable la extracción óptima de señales de los soportes originales; este es el punto inicial de cualquier proceso de digitalización. La tecnología de grabación requiere de tecnología de reproducción muy específica; es por eso que se tiene que tomar en consideración el cambio de la tecnología antes de que la obsolescencia del hardware se convierta en un problema, algo que ya hemos escuchado durante el Seminario.

La habilidad para volver a copiar una copia digital sin degradación ha llevado a los documentalistas a decir que estamos hablando de una preservación eterna, pero la facilidad de producción de las copias de baja resolución limita la capacidad de los documentalistas en términos del acceso: sin dañar el elemento original, más allá de ser eternos, se necesitan prácticas de archivo efectivas para mejorar la vida e integridad del contenido audiovisual.

La conversión digital y la preservación bien manejadas pueden facilitar la realización de los beneficios que promete la tecnología digital; asimismo, un sistema mal planeado que requiera de intervención manual puede considerarse como un elemento dañino, ya que se sale de las capacidades del documentalista y puede dañar la calidad de la colección.

Un sistema bien planeado puede garantizar la automatización del proceso y lograr una preservación adecuada. Cabe mencionar que ningún sistema audiovisual puede proporcionar una solución única, todos los sistemas van a requerir de mantenimiento y migraciones futuras planeadas, todo esto sucede desde el momento en que se hace la digitalización y almacenamiento del original.

En este momento, probablemente todos estamos de acuerdo en que los archivos digitales no pueden reemplazar a los archivos físicos; pueden ser un apoyo y mejorar su capacidad de acceso. En un archivo digital fragmentamos –a esto se le llama “digitalización” – un objeto

analógico del mundo real en una serie de valores numéricos unidos oportunamente. A partir de este momento, el Sistema de Almacenamiento Masivo Digital puede utilizarse con todas sus funciones, lo que nos sirve mucho, ya que podemos hacer una réplica, cambiar el formato y transmitir el contenido a una amplia gama de usuarios.

Lo que nos sigue faltando es una serie de componentes necesarios para recrear el original ante el usuario. Por ejemplo, cuando hablamos de la grabación de sonidos provenientes de un LP (del inglés long play) ¿qué es lo que podemos digitalizar y guardar en un Sistema de Almacenamiento Masivo Digital? Imágenes escaneadas de la portada por los dos lados. El sonido en sí es un solo archivo que tal vez tenga diferentes pistas, cada una con su registro; todo esto lo podemos guardar en un solo archivo digital. Aquí decidimos cómo manejar los componentes y esto puede tener un impacto al momento de tratar de reorganizar la información y ofrecer un paquete de difusión de la información para el usuario –DIP en términos de la OAIS–. Es poco probable que podamos proporcionar exactamente lo mismo que teníamos al inicio.

El único objeto que va a permanecer sin cambios a lo largo de esta cadena es el objeto digital que se desprende del objeto físico. Esto nos lleva a otra pregunta, ¿cuántos archivos recopilan estos objetos? Aun cuando el número de este tipo de archivos se incrementa, la mayor parte de los materiales que tenemos en los archivos para la posteridad son de naturaleza analógica. Entonces, todo lo que hemos dicho es un fundamento sólido, además de esto podemos abordar aspectos prácticos para las estrategias de protección de datos.

Los siguientes criterios son elementales para el diseño de un Sistema de Almacenamiento Masivo Digital, ya que tenemos diferentes estrategias para manejar los datos en sistemas de largo plazo y partimos de la suposición en la que los datos son los que se pueden preservar, no el soporte. Para audio y video, el sistema debe transformar los datos a formato estándar o de registro reconocibles por las computadoras, como: WAF AIFF o BWI. Estos archivos, a diferencia del audio o los soportes específicos para video, definen tecnológicamente los límites de su contenido y están codificados de tal forma que reconoce la pérdida de datos a lo largo de todo el sistema.

Asignación de responsabilidades: debe quedar muy claro quién será el encargado de la protección del almacenamiento de datos y para esto se requieren ciertas habilidades, conocimientos y experiencia en gestión. En todas las colecciones, la protección y almacenamiento de datos requiere de recursos exclusivos, planes y un encargado de las estrategias, incluso, las colecciones más pequeñas deben tener acceso a la experiencia, tareas específicas asignadas e infraestructura técnica adecuada para realizar las actividades; por otro lado, los datos se deben almacenar y manejar con sistemas y soportes adecuados.

Cada Sistema de Almacenamiento Masivo se basa en programas de preservación de audio, una vez que se definan los requerimientos, es necesario comunicarlos a los proveedores de diferentes soportes y sistemas para identificar cuál satisface las necesidades. En los programas de preservación debemos contar con los sistemas adecuados.

Suficiente capacidad de almacenamiento: la capacidad puede cambiar a lo largo del tiempo, pero el sistema debe poder manejar la cantidad de información que se espera almacenar en un ciclo. Una de las habilidades fundamentales es que el sistema duplique datos siempre que sea necesario y sin perder información. La transferencia a los diferentes soportes no debe presentar pérdidas, necesita gozar de confiabilidad y tener apoyo técnico para poder lidiar con los problemas de forma adecuada.

Según la arquitectura, se debe contar con la habilidad de mapear los archivos con un esquema de nombres. Los sistemas de almacenamiento se basan en nombres de objetos y esto puede ser una limitante para el manejo de los objetos almacenados; por ejemplo, los sistemas en disco que tal vez tengan una estructura jerárquica directa con diferentes nombres de archivo, son sistemas diferentes al que se utiliza en cintas, por lo que deben tener: un mapeo de los nombres de los archivos y contar con identificadores y la habilidad de manejar el almacenamiento redundante.

Los medios digitales tienen una capacidad de fallas baja y se pueden tener duplicaciones redundantes, que son necesarias en cada etapa, especialmente en las últimas etapas de almacenamiento.

La revisión de errores: se necesita tener un alto nivel de revisión de errores en los sistemas, generalmente no se utiliza la participación humana, el sistema debe identificar los errores y aplicar las acciones pertinentes. Los sistemas deben alertar a los administradores e identificar los errores con tiempo para aplicar una solución, también necesitamos el almacenamiento de metadatos y datos para el establecimiento de vínculos entre los objetos digitales. Por lo general, las operaciones mayores, necesitan sistemas grandes que se encuentren vinculados pero separados del sistema maestro de almacenamiento digital para poder manejar el número de procesos involucrados y permitir que los metadatos y las interfases se puedan cambiar, sin tener que cambiar el archivo maestro.

Sustentabilidad del sistema: toda la tecnología, ya sea de software y hardware, formatos o normas, en algún momento va a cambiar debido a las fuerzas de mercado y a las necesidades o expectativas del público. La tarea del documentalista audiovisual, a cargo del manejo del contenido digitalizado, es encontrar salidas a través de estos cambios tecnológicos, de tal forma que podamos utilizar, en el futuro, toda la información en su formato auténtico de una forma costo-efectiva y manejable.

La planeación a largo plazo para esos archivos involucra, no solamente las normas técnicas para el sistema de almacenamiento de datos, hay algunos elementos técnicos que se deben resolver con mucho cuidado. También los aspectos sociales y económicos de la administración de estos sistemas son elementos que se deben tomar en cuenta para el acceso continuo a la información.

La sustentabilidad de los datos originales; es decir, el mantenimiento de los datos analógicos en un orden lógico, los datos deben regresar al sistema sin tener ningún tipo de corrupción. Vale la pena mencionar que los sistemas informáticos pueden identificar el número, la

cantidad de mantenimiento o “refrescamiento” que se requiere y sólo el buen uso de estos sistemas informáticos puede darles buenos resultados. La cosa está de miedo, ¿no lo creen?

Formatos y reproducción: los datos solamente son útiles si se pueden presentar como audio y video en el futuro. La elección adecuada de los formatos en los archivos garantiza la reproducción de los archivos de datos. Se debe tomar en cuenta la tecnología, es decir, el no utilizar pérdidas en los algoritmos de compresión garantiza que no se altere el contenido audiovisual y sonoro.

Los metadatos, la identificación y el acceso a largo plazo: todos los archivos audiovisuales deben ser identificables y ubicables, ése es el propósito, utilizar el material y valorar el contenido.

Ahora hablemos acerca de la economía de los archivos audiovisuales, esto incluye: a la institución que soporta el almacenamiento audiovisual y el repositorio; así como aquellas organizaciones que reciben un valor de la información almacenada. El costo de mantenimiento de una colección o archivo es continuo y debe haber un plan y presupuesto realista, que tome en cuenta los aumentos, para la preservación a largo plazo de las colecciones, la curaduría y la administración de la colección audiovisual. La preservación digital es un tema económico y tecnológico, se trata de encontrar fuentes confiables para garantizar la sustentabilidad del contenido digital, las tecnologías y sistemas de repositorios. Se debe dar mantenimiento a estos sistemas mientras sea requerido, en otras palabras, se les debe tener para siempre.

Almacenamiento, manejo y preservación alternativos: debido a que el ambiente tecnológico es volátil, se deben establecer acuerdos entre archivos e instituciones en lo que se refiere al almacenamiento como archivos de última instancia; esto requiere de acuerdos sobre el formato, la organización de los datos, elementos sociales y económicos para el manejo del contenido. Como pueden ver, poco a poco voy llegando al punto central de mi presentación, éste es mi último punto.

Herramientas, software y planeación a largo plazo: los sistemas, el hardware y el software necesitan de mantenimiento externo, ya que son solamente herramientas para preservar el contenido. El software del repositorio T-Space, por ejemplo, no se describe como una solución de preservación, es útil para permitir a las instituciones mantener información y activos valiosos para ellos. El software del repositorio es una herramienta, al igual que los distintos componentes, y se diseña para ayudar en la operación: son procesos para manejar y utilizar los metadatos. La planeación a largo plazo debe permitir que se actualicen los sistemas sin dañar el contenido y esto es lo más difícil.

En conclusión, cuando se define un sistema de almacenamiento masivo digital, no solamente tenemos que contactar al proveedor y decirle qué hardware o qué software queremos, porque no vamos a encontrar una solución “de llave en mano” en el mercado.

Necesitamos comenzar con un proyecto, con fecha límite, asignar los recursos, involucrar a todo el personal interno y buscar los flujos de trabajo y especificaciones para entrar a la etapa de planeación, crear alianzas en la medida que se pueda, entender quién tiene la información

para desarrollar el sistema, hacer que el sistema sea a prueba de fallas, redactarlo en papel, probarlo en un ambiente de prueba y después pasarlo al ambiente de producción; es decir, el ambiente del archivo.

Entonces, hay varios libros que ahondan un poco más en el tema de esta presentación, especialmente el modelo de referencia de la OAIS y el documento TC04 de IASA.

## Paradigmas de los sistemas de gestión y almacenamiento masivo digital

### Tom Lorenz

#### Socio directivo de cube-tec international GmbH

Alemania. Primeramente quisiera darles las gracias por la oportunidad de venir a dar esta presentación ante ustedes, agradecer a los mexicanos por su hospitalidad y a los organizadores quisiera felicitarlos por la magnífica organización de este Seminario.

Voy a hablar de los elementos de una solución para archivos que cuenten con actividades de trabajo en grupo; con esto me refiero a personas que trabajan en un archivo y todas sus actividades se orientan a digitalizar el material. Tratar de diseñar el flujo de trabajo es una tarea bastante difícil y en esta presentación hablaré de las generalidades, del proceso de digitalización masiva de archivos de audio y de cómo se administran las labores en grupo.

Hablemos de cuáles son los elementos útiles para optimizar el proceso. El resultado del proceso será tener archivos sonoros en WAV o BWF, archivos en XML con metadatos descriptivos, técnicos y también metadatos en distintos campos de la base de datos. En términos generales estos son los elementos y explicaré cada uno.

Empecemos con la ingesta de datos analógicos: necesitamos que la transferencia sea de óptima calidad, ya que pasa del dominio analógico al digital. En la imagen vemos lo que se necesita para poder salvaguardar los registros sonoros históricos. La estación de trabajo se encarga de monitorear el sonido con el oído humano y el operador debe mantener los estándares de calidad para poder hacer la transferencia del dominio analógico al digital; para cumplir con los parámetros de calidad se requieren protocolos y para ello tenemos un sistema, desarrollado por nosotros mismos, que verifica la calidad analógica y digital y tenemos un inspector de errores.

Estas son herramientas que funcionan en tiempo real para detectar en la corriente de audio qué es lo que está sucediendo. En cuanto al tiempo, tenemos la posibilidad de identificar en qué momento se están dando las cosas y plasmarlas en una lista que se llama "Informe" (report list) y hacer un análisis estadístico del promedio de todo lo que va corriendo en el proceso de transferencia; además, tenemos la posibilidad de identificar eventos automáticamente. Esta es la ventana en que saldría todo lo que se va registrando y va apareciendo en esta lista. Así tenemos la posibilidad de identificar: el inicio y el final de la