

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

modulación, señales, pausas, qué señal está corriendo, cuándo inicia, la distorsión análoga, chasquidos, “drop-outs”, cuándo termina, pérdidas de información, la velocidad, si viene en mono o si viene en stereo, la tasa de señal a ruido, ancho de banda, niveles promedio, picos, niveles dinámicos, errores de azimut, correlación balance stereo, cliqueo, distintos tipos de sonido y huecos en el muestreo del archivo y muchos otros aspectos.

Es muy difícil controlar todo lo que está pasando en tiempo real, y sólo con ayuda del oído humano, debido a los recursos humanos y financieros; por lo tanto, estas herramientas tienen la posibilidad de ofrecer un proceso óptimo del control de calidad, para que cada una de las etapas pueda optimizarse y se pueda cubrir el mayor número de medios en el proceso de transferencia al entorno digital. La pregunta entonces es ¿cómo podemos dividir el proceso en tantos grupos sin perder la calidad? ¿Es posible hacerlo?

Con estas herramientas sí, ya que se pueden tener más estaciones de trabajo para reducir la carga, en algunos casos se necesitarán más operadores y dividir el proceso. Según sea el caso, tenemos la posibilidad de hacer grabaciones en distintas velocidades –hasta cuatro– dependiendo del material; por supuesto que no todos los materiales pueden ser sometidos a este proceso, habrá cintas que manejen velocidades menores y otras que se puedan reproducir al doble y otras requerirán de ecualización, o podemos tener un proceso de grabación multicanal con hasta ocho pistas y canales. También tenemos la posibilidad de hacer grabaciones en dos direcciones y hacer una ingesta de pistas de mono a stereo, de carrete abierto, casete, o compacto a otros formatos. Además se puede hacer una combinación de todas las dinámicas de grabación y verán cómo se ahorran mucho tiempo con la bidireccionalidad y la variación de velocidades: para un casete de 90 minutos nos tardaríamos aproximadamente 23 minutos en digitalizarlo y, por supuesto, se puede aumentar el número de módulos.

Hablemos ahora de la ingesta de medios digitales; este es un proceso de transferencia de cassetes DATs, CDs y Minidiscs a otro medio. Los mecanismos de detección de errores se aplican a través del mecanismo de reproducción, también el proceso de importación de los archivos tiene que pasar por un proceso de revisión de calidad porque las corrientes de bits se extraen y se registran en un archivo XML. Los archivos comprimidos de audio MPEG en nivel 1, nivel 2 ó nivel 3, también se pueden someter al proceso de revisión de calidad; este proceso se puede acelerar o correr de manera automatizada, al igual que la verificación de los archivos de audio con una potencia de procesamiento mucho mayor, se puede optar por utilizar múltiples mecanismos de reproducción para CDs, CDROM, utilizar altas velocidades y, además, distintas estaciones de CD o la modalidad de Jukebox para los CDs. Los medios digitales se pueden pasar por un proceso de ingesta a una velocidad mucho mayor que los medios analógicos. Los dispositivos de reproducción como el DAT a Minidisc ya son obsoletos y es imperativo que estos medios sean los primeros en transferirse por la obsolescencia tecnológica.

En términos generales, hemos hablado de cómo funciona el proceso de transferencia para medios analógicos, digitales, CDs y de la ingesta de distintos archivos; en este sistema y con la herramienta en cuestión, todo puede pasar a un archivo temporal y todos los procesos van a pasar automáticamente por control de calidad.

Ahora hablemos de la estación de control de calidad. En un proceso de digitalización altamente eficiente, no es posible tener a una persona escuchando para verificar la calidad, ni tampoco tener a un supervisor u otro operador supervisando ese trabajo; además, si ponemos las cintas o de necesitar hacer un corte o si las cintas están muy alteradas, sabemos que una persona no puede hacerlo todo, se tiene que mandar a un especialista. Dependiendo de las estaciones de trabajo, el nivel de ingesta requerido y de los medios, tendrán que disponer de una o varias estaciones de trabajo según la demanda.

Este es un ejemplo de una estación de control de calidad con cuatro funciones.

Tenemos una máquina para cada tipo de medio: primero que nada se reproduce el tipo de medio; se hace la conexión al mecanismo de reproducción; en caso de necesitar hacer reparaciones se hacen en este punto; revisamos cuáles son los archivos de audio existentes. Por ejemplo, si hay informes de error, o si vienen en formato BWF, o de haber cambios en los campos de los datos del formato BWF, se verifica la suma total de los datos que vienen en ese archivo, todo esto se puede hacer en el punto de la revisión de calidad.

Esta es la estación de restauración opcional. En algunos casos, los materiales especiales tendrán que ser reparados y restaurados; así es que el objetivo principal de la preservación es digitalizar el material en su versión original, pero, en muchos casos, tendremos que pasar primero al periodo de restauración de esos archivos. Por lo tanto, tenemos que tener especialistas en nuestro equipo para que den los pasos necesarios al proceso de la ingesta. El operador puede brincar de un paso a otro según lo decida, lo que representa una ventaja muy importante por que este sistema está integrado, nosotros le llamamos a este sistema “de un sólo clic” para pasar a la restauración y viceversa. Aquí vemos una lista de eventos y con tan sólo un clic se puede pasar a la herramienta que está en el editor y solucionar los problemas que nos presente dicho archivo. Contamos con distintas herramientas para optimizar nuestra labor y lidiar con los artefactos que tienen cierta distorsión, trabajar con ruido, interferencia y para solucionar problemas en el campo del espectro.

Ahora hablemos del procesamiento de lotes. En un proceso de ingesta masiva, mandaremos un gran número de archivos de audio al Sistema de Almacenamiento Masivo: primero se tienen que reparar, verificar y hacer todos los procesos necesarios de reparación “trim and fade” y, en algunos casos, hacer la codificación en MP3. Antes de llegar a la división automática de archivos necesitamos un proceso automatizado para manejar los metadatos. Muchas veces se utilizan algunos mecanismos de compresión que hacen que se pierda el menor número de información que se pueda, como: Flag y MLP. Los métodos para hacer uso del procesamiento de lotes y ahorrar espacio son muy diversos: por ejemplo, el análisis del contenido, el procesamiento, la conversión de los archivos, la seguridad y la codificación del archivo sonoro son herramientas que nos ayudarán a tomar las mejores decisiones con respecto a los metadatos.

Además, en nuestro sistema se puede crear un diagrama de flujo con tan sólo seleccionar y arrastrar las herramientas; se pueden utilizar procesos conectados con un audio virtual y con la transferencia de metadatos a través del sistema. Los metadatos técnicos nos ayudan a mejorar el proceso de ingesta para decidir cuál es el proceso que se adaptará mejor a las necesidades

de dicho archivo, por lo tanto, tenemos un editor para definir las reglas y qué documento es. De tal manera que se le dará prioridad a la transferencia de los metadatos encontrados en otro punto e incorporará dichos elementos en nuestro proceso.

De aquí pasamos a la estación de catalogación. En un proceso de digitalización masiva podemos hacer que nuestro mantenimiento y el proceso vayan de la mano, o por su cuenta, del proceso de ingesta de metadatos. El mantenimiento hace uso de las herramientas y no necesariamente tiene que ir sincronizado con el proceso de metadatos. En nuestro proceso se pueden manejar las dos cosas separadas e incluirlas en el momento que se indique; aquí lo que hacemos es reproducir los archivos sonoros en ambas estaciones de trabajo. La estación de catalogación exige que el personal tenga conocimiento del contenido del audio, es muy importante que estas personas se manejen de esta forma.

Pasemos al almacenamiento temporal. Los sistemas de almacenamiento masivo no son parte de las soluciones del grupo de trabajo. El objetivo de un grupo es proporcionar paquetes de datos para exportarlos automáticamente a una base de datos existente, un sistema de radio, por ejemplo. Ya que éste es un proceso totalmente separado del sistema de almacenamiento masivo digital –lo que vimos en la presentación anterior–, estos archivos no se van directo al sistema de almacenamiento masivo digital y hay distintas razones por las que un documento sonoro se tenga que modificar en algún momento.

Primero que nada se hacen verificaciones aleatorias de calidad desde la estación de control debido a las advertencias que nos salgan en la lista de errores y habrá veces que se determine grabar una cinta una segunda vez, y por lo tanto, se decida eliminar el documento sonoro y también generar copias en una resolución menor como un MP3.

Todos los archivos se guardan en un sistema de almacenamiento temporal para poder conservarlos por lo menos dos semanas en este punto, antes de pasar a la siguiente etapa.

Aquí el sistema nos da la oportunidad de tener un disco duro con todas las capacidades RAID con cinco a 10 terabytes para evitar que los datos se pierdan. Lo que resulta una estrategia muy buena de respaldo temporal en cintas. Dicho sistema de respaldo sirve para respaldar los archivos a la larga y al momento de decidir qué hacer con el documento se puede hacer uso de estas herramientas para que no tengan que instalar una entrada al sistema masivo digital. Con esta herramienta tenemos la posibilidad de esperar a pasar a una etapa posterior para transferir al sistema de almacenamiento masivo digital todos los archivos que ya se hayan trabajado.

En cuanto a la interfase, no todos los metadatos tienen que exportarse a los campos de datos del MAM, muchos de los datos técnicos no le sirven a los investigadores y ni a las fonotecas; por lo tanto, se tiene que decidir qué campos del sistema MAM se van a utilizar, ya que se puede mandar como archivo XLM y así no se pierde información y se puede recuperar según se vaya solicitando en el futuro.

Ahora hablemos de la solución del flujo de trabajo; ésta tiene dos objetivos: primero que nada, servir como mecanismo de gestión en todas las labores de digitalización. El orden de las

necesidades de digitalización se va a decidir de acuerdo a la antigüedad, al número del documento, en qué anaquel se encuentra y al grado del daño también.

Entonces se tiene que dar prioridad a aquellos documentos que están a punto de desaparecer y es muy importante tener un registro de todo lo que se va a digitalizar. La herramienta nos da la oportunidad de tener un registro, además de identificar aquellos medios que pudiesen representar un problema en el proceso. En segundo lugar, tenemos la gestión de todos los metadatos recolectados y que deben quedar registrados, para eso necesitamos identificadores únicos en el sistema de flujo de trabajo. Este sistema también puede manejar archivos adjuntos y habrá que seleccionar y preparar los metadatos para exportarlos al MAM de la institución o al sistema de radio automatizado.

Para concluir, por más de diez años Cube-Tec, continuamente ha desarrollado productos para satisfacer las necesidades de los sistemas de ingesta masiva de la actualidad. Hemos desarrollado Audio Cube, para restauración y Cube Word Flow, que canaliza y da apoyo a todas las actividades que se realizan en un grupo de trabajo; también tenemos a Cuadriga, con más de 10 años y se utiliza en todo el mundo, además de ser la norma para los procesos de digitalización en masa y Doving, motor de procesamiento en lotes, para el procesamiento masivo de documentos.