

Formatos de preservación final

- 6.1.1.5 Debe notarse que ningún soporte digital constituye una preservación permanente, pero si se le incorpora a sistemas tecnológicos capaces de administrar el cambio inevitable, facilitarán la transferencia de los datos de audio de un sistema a otro. Sin embargo, el gran beneficio de escoger un sistema que pueda modernizarse para adaptarse al cambio puede ser también su gran desventaja. Si el soporte tecnológico no se encuentra disponible, o si los recursos para reemplazar el sistema dentro del tiempo requerido no están disponibles, entonces el almacenamiento digital puede poner las colecciones en riesgo.

6.2 Sistemas de almacenamiento masivo digital

6.2.1 Introducción

- 6.2.1.1 Un sistema de almacenamiento masivo (DMSS) puede ser un sistema comprensible, completamente automatizado y diseñado para almacenar, administrar, mantener, distribuir y preservar un complejo conjunto de objetos digitales heredados junto con los metadatos relacionados, o un sistema de respaldo y almacenamiento sencillo para un solo tipo de formato de archivo, sus copias derivadas junto con los metadatos relacionados. La decisión acerca de un sistema apropiado dependerá primeramente, entre otras cosas, del tipo y tamaño de la colección, la relación de la colección de sonido con un archivo más grande o biblioteca, la necesidad de administrar el acceso en línea y la disponibilidad de recursos económicos.

- 6.2.1.2 Hay sistemas completamente desarrollados disponibles, y algunos fabricantes prometen el suministro de soluciones completas que incluyen todo el *hardware* y el *software*. Alternativamente, archivos con soporte adecuado IT pueden ser decisivos al comprar componentes e integrarlos a un sistema que cubra necesidades específicas. La complejidad —o simplicidad— de cualquier solución particular puede variar tanto que se encuentra más allá de los alcances de este documento para definir o recomendar un enfoque particular o sistema.

- 6.2.1.1 En cambio, las siguientes secciones de estos lineamientos:

- Ofrecen una descripción somera y ligas a las especificaciones de diversos sistemas o documentos con propuestas de sistemas de biblioteca o archivo que pueden usarse como modelos.
- Incluyen lineamientos generalizados para las estrategias de protección de datos basados en los lineamientos de la UNESCO para la preservación de la herencia digital.
- Proporciona dos secciones sobre los dos componentes principales de un sistema de preservación digital, cinta de datos y discos duros respectivamente.
- Incluye una sección específica sobre sistemas de almacenamiento en pequeña escala manualmente respaldados.

6.2.2 Sistemas ejemplares

- 6.2.2.1 **La Biblioteca Nacional de Australia (NLA)** compró su sistema de almacenamiento digital de objetos (DOSS) en 2001. La colección digital de audio de la NLA incluye el archivo *Pandora*, una colección de publicaciones selectas en línea; su colección de imágenes digitales, mayormente sustitutos digitales de ítems de colección con algunos originales digitales; publicaciones electrónicas y colecciones de manuscritos electrónicos, y una colección de sonido no publicada de alrededor de 36 mil horas. El DOSS es el encargado de almacenar todos esos materiales. El sistema consta de alrededor de 10 TB (*terabytes*) de almacenamiento en y fuera de línea. El almacenamiento en línea lo suministran arreglos RAID en una red SAN usando discos Fibre Channel de 1 y 2 GB. El almacenamiento fuera de línea lo suministra una librería robotizada de 140 TB usando tecnología DLT y LTO2. Un sistema de almacenamiento jerarquizado

es usado
veces a ci
sistemas
SAN, todo

2.2.2 Al mismo
completo
bases de
ción han:

2.2.3 La informa

2.2.4 **La Biblio**
masivo di
de esta cl
en discos
creados c
digitales c
disco en e
físicament
glo virtua
sistemas c
combinad
DSM usa un
ha ocurri

2.2.5 El sistema
a perpetu
software, c
de acceso
en fuente
ligar los o

2.2.6 La red de
pios de 2
nexión 10
el control
el primer
la radio.
70 TB ser
Veritas N
almacenar

2.2.7 **La Sued**
mania, qu
integrar la
zado y en
de las col
la única s
y superar
aquella de
recursos

Formatos de preservación final

es usado para administrar el almacenamiento en y fuera de línea. Un objeto es copiado tres veces a cinta y almacenado en múltiples localidades seguras. Hay una serie de servidores para sistemas de catalogación, acceso en línea, bases de datos, administración digital y conmutación SAN, todo esto integrado en un ambiente de red de 100 MB a 1 GB.

- 2.2.2 Al mismo tiempo, la NLA se comprometió a desarrollar el sistema de administración digital completo. La herramienta primaria de acceso a la colección es el catálogo; sin embargo, las bases de datos, ILMs, las herramientas de administración de metadatos y los sistemas de colección han sido integrados en un sistema completo.
- 2.2.3 La información detallada del sistema puede obtenerse en <http://www.nla.gov.au/dsp/doss/doss.doc>
- 2.2.4 **La Biblioteca Nacional de Noruega (NB)** comenzó desarrollando su almacenamiento masivo digital, llamado Archivo Digital de Seguridad, en 2001. Ya que la velocidad de desarrollo de esta clase de tecnología de almacenamiento es tremenda, el DSM ya ha visto varios cambios en discos, cintas y tecnología de administración. El DSM guarda todos los objetos digitales creados o coleccionados por la NB incluyendo material audiovisual, fotografías, sustitutos digitales de material impreso y sitios web cosechados. El sistema consta de dos sistemas de disco en espejo de 50 TB que emplean ambas clases de discos ATA y Fibre Channel, colocados físicamente en diferentes lugares, para el material accesible en línea y uno de 12 TB EVA (arreglo virtual empresarial), uno de 10 TB ATABeast y uno de 12 TB ATABoy2. Éstos sirven para los sistemas de administración, catalogación en línea y flujo de trabajo de producción, y todo está combinado con un sistema robótico en cinta LTO de 80 TB para la tercera copia de respaldo. El DSM usa una red SAN conmutada a 2 GB/seg. Tal como el hardware ha sufrido cambios, lo mismo ha ocurrido con el sistema de administración.
- 2.2.5 El sistema está empleando un DSM con objeto de cuidar la información en los objetos digitales a perpetuidad, suministrando funciones tales como la creación de información espejo por software, capacidades de verificación de errores, estrategias necesarias de conversión, control de acceso y administración. El DSM es creado totalmente en casa con herramientas basadas en fuente abierta (*open source*), implementando una numeración de recursos única (URN) para ligar los objetos digitales a los catálogos, corriendo además Veritas Netbackup.
- 2.2.6 La red de almacenamiento de área (SAN) fue establecida en ScreenSound, Australia, a principios de 2003, con un grupo de diez estaciones de trabajo de audio en red vía Ethernet, conexión 100 BaseT. Las estaciones de trabajo introducen el material directamente a la SAN bajo el control del sistema de administración de medios MAVIS. Imágenes fijas se siguieron durante el primer año. El sistema ahora surte señal a la televisión terrestre y vía satélite así como a la radio. Los intentos para el video de alta resolución han comenzado. La biblioteca LTO de 70 TB será modernizada a 140 TB cuando la rutina de ingestión de video comience en 2005. Veritas Netbackup se usa para el respaldo diario del trabajo en progreso, y el migrador de almacenamiento HSM se usa para archivar copias para retención a largo plazo.
- 2.2.7 **La Suedwestrundfunk (swr)**, una de las corporaciones de radiodifusión pública en Alemania, que transmite también radio y canales de TV, inició deliberaciones en 1993 sobre cómo integrar los archivos de sonido de radio en un ambiente de radiodifusión totalmente digitalizado y en red. Un proyecto piloto a dos años encontró que la automatización y digitalización de las colecciones de audio por medio de almacenamiento masivo digital de audio (AMS) era la única solución para lograr esta integración. Esta solución también hizo posible encontrar y superar exitosamente el problema inherente a la preservación que enfrenta todo archivo, aquella de la inevitable obsolescencia de los soportes y equipo de reproducción. Entre los recursos que la swr incluye están las cintas magnéticas de 1/4 de pulgada, vinilo, casetes DAT,

Formatos de preservación final

videocassetes U-matic y CD. Existe un estimado de 380 mil horas de material de audio en la colección combinada SWF. Desde 1975, bases electrónicas de datos han reemplazado los catálogos convencionales de tarjetas.

- 6.2.2.8 La SWR posee tres localidades de producción (Stuttgart, Baden-Baden, Mainz), que están interconectadas mediante una red a 6.8 GB/seg. Desde 2003, dos de ellas están equipadas cada una con un arreglo de discos duros de 2 TB, una biblioteca en cinta robotizada de 20 TB que incluye cuatro transportes que usan cartuchos de 40 GB, y una serie de servidores (agentes) para tareas diversas. Ambos sirven como sistema de respaldo entre ellos ofreciendo así seguridad óptima de los datos. Es un sistema modular: todos sus componentes pueden ampliarse para cumplir los requerimientos de un crecimiento permanente tan rápido como sea necesario. Los sistemas locales AMS son SUN conmutados dentro de una red *fibre-channel*.
- 6.2.2.9 Los archivos de audio son creados como archivos .wav lineales (es decir, sin reducción) para la producción, postproducción y con propósitos de intercambio, así como para la escucha previa en MP3. Las tres locaciones cuentan con un número de clientes de audio en donde los usuarios pueden escuchar archivos de audio MP3 usando las bases de datos como herramientas primarias de acceso, y unas pocas estaciones de captura donde los especialistas llevan a cabo la transferencia de la colección hacia el almacenamiento masivo, audio y el material con información que lo acompaña, como son los folletos del CD.
- 6.2.2.10 La SWR también decidió digitalizar archivos de TV, iniciando en 2004 con un archivo de prueba.
- 6.2.2.11 **PARADISEC (Archivo pacífico y regional para fuentes digitales en peligro de extinción).** Es un sistema interinstitucional establecido entre las Universidades de Sydney, Melbourne y la Universidad Nacional Australiana (ANU) en 2003, al que se unió en 2004 la Universidad de Nueva Inglaterra. La arquitectura del sistema explota el amplio ancho de banda de la red de investigación y educación GRANGENet para mover datos de audio entre las instituciones participantes y los Asociados Australianos para la Computación Avanzada (APAC); Facultad Nacional localizada en el campus ANU, donde un sistema de almacenamiento masivo de 1200 TB, basado en el Sistema de cartuchos automatizado Storage Tek Powderhorn 9310, se encuentra disponible a un amplio rango de usuarios de educación superior.
- 6.2.2.12 Los archivos de audio de PARADISEC se mantienen inicialmente en un servidor de espera local en la Universidad de Sydney, donde el audio es introducido. Los datos de audio son creados en la Universidad de Sydney, normalmente como BWF a 24 bits, 96 kHz. Los contenidos del servidor se archivan periódicamente en el sistema de almacenamiento masivo en la Universidad de Sydney por medio de la impresión de una marca de agua alta-baja, la cual se crea automáticamente tan pronto como los contenidos del servidor de espera exceden 80% de su capacidad. Además, cada noche, una sinopsis automática que se encuentra en Canberra actualiza cualquier archivo nuevo o cambiado, de manera que se crea una imagen espejo de la colección en Canberra. Las auditorías periódicas de la colección incluyen metadatos, contenidos, tamaño de archivo y reportes de tiempos de reproducción; se generan automáticamente y son distribuidas vía correo electrónico al grupo de administración del proyecto para reportes y control de calidad. En el primer año de operación, PARADISEC había generado 868.3 GB de datos. La tasa de ingestión anual es estimada en 2 TB para el segundo año y 3 TB en el tercer año de operación.

6.2.3 Prin

6.2.3.1 La sig
estrat
de la
incorp
digital
2003:

6.2.3.2 Aspec

6.2.3.3 Existe
almace
preserv

6.2.3.4 Asigna
ponsabi
respons
así com
colecti
respons
un encar

6.2.3.5 Infraest
y adminis
mas de a
disponibl
digital; se
determin
servicio. E
y los sele

6.2.3.6 El sistema

6.2.3.7 Suficient
construirs
datos que

6.2.3.8 Es fundam
se requier

6.2.3.9 Confiabili

6.2.3.10 La habilid
denominaci
namiento es
tecturas par
los objetos
pueden impac
diferentes a
terriblemente
los identific

Formatos de preservación final

6.2.3 Principios DMSS

- 6.2.3.1 La siguiente información está basada de manera muy cercana a los aspectos prácticos de las estrategias de protección de datos de los lineamientos de la UNESCO para la preservación de la herencia digital. Está modificada solamente para reflejar la posibilidad de sistemas que incorporen el respaldo no automatizado y las dudas en cuanto a la preservación del audio digital en un solo formato. La sección está incluida con el amable permiso del autor (Webb, 2003: 16.13).
- 6.2.3.2 **Aspectos prácticos de las estrategias de protección de datos**
- 6.2.3.3 Existe una razonable gama de estándares de estrategias empleadas para administrar datos en almacenamiento de largo plazo. La mayoría asume que el soporte en sí mismo no necesita ser preservado, solamente los datos. Lo que sigue comprende, en parte, esas estrategias.
- 6.2.3.4 **Asignación de responsabilidad.** Alguna persona concreta debe tener la absoluta responsabilidad de la administración del almacenamiento de datos de su protección. Ésta es una responsabilidad técnica que demanda un conjunto particular de habilidades y conocimientos así como de capacidades administrativas. El almacenamiento de datos y la protección de las colecciones requieren de un plan apropiado y de personal asignado específicamente para responsabilizarse de todas esas estrategias; aun la colección más pequeña debe contar con un encargado que tenga el conocimiento y la habilidad para esta tarea.
- 6.2.3.5 **Infraestructura adecuada para realizar el trabajo.** Los datos deben almacenarse y administrarse mediante los sistemas apropiados y sobre el soporte adecuado. Existen sistemas de administración de activos digitales o sistemas de almacenamiento de objetos digitales disponibles que cumplen con los requerimientos de los programas de preservación del audio digital; se discuten algunos enfoques más adelante. Una vez que los requerimientos se han determinado, deberán discutirse completamente con los suministradores potenciales del servicio. Diferentes sistemas y soportes son los más adecuados para diferentes necesidades, y los seleccionados para programas de preservación deberán cumplir con ese propósito.
- 6.2.3.6 El sistema completo debe tener capacidades adecuadas que incluyen:
- 6.2.3.7 **Suficiente capacidad de almacenamiento.** La capacidad de almacenamiento puede construirse a través del tiempo, pero el sistema debe ser capaz de administrar la cantidad de datos que se espera almacenar dentro de su ciclo de vida.
- 6.2.3.8 Es fundamental que el sistema sea capaz de **duplicar sin pérdidas los datos cuando se requiera**, y transferir datos nuevos o renovados a otros soportes sin pérdidas.
- 6.2.3.9 **Confiabilidad demostrada** y soporte técnico para afrontar los problemas rápidamente.
- 6.2.3.10 **La habilidad de mapear nombres** de archivos dentro de un esquema conveniente de denominación de archivos para su arquitectura de almacenamiento. Los sistemas de almacenamiento están basados en el nombre de los objetos. Sistemas diferentes usan diversas arquitecturas para organizar objetos. Esto quizá pueda imponer restricciones en la manera en que los objetos son nombrados dentro del almacenamiento; por ejemplo, los sistemas de discos pueden imponer un directorio con estructura jerárquica sobre los nombres de los archivos, diferentes a aquellos que podrían usarse en sistemas de cinta. El sistema debe permitir, o preferiblemente llevar a cabo, un mapeo impuesto por el sistema sobre los nombres de archivo y los identificadores existentes.

Formatos de preservación final

- 6.2.3.11 La habilidad de **administrar almacenamiento redundante**. Dado que el medio digital tiene una tasa de fallas pequeñas pero significativas, es necesario tener copias excedentes de los archivos en cada etapa del proceso, especialmente en la fase final del almacenamiento.
- 6.2.3.12 **Verificación de errores**. Un nivel de verificación automática de errores es normal en la mayoría de los almacenamientos en computadora. Debido a que los materiales de audio y A/V deben mantenerse por largos periodos, generalmente con muy poco uso humano, el sistema debe ser capaz de detectar cambios o pérdida de datos y ejecutar la acción apropiada. Las estrategias del lugar deben alertar a los administradores de la colección, por lo menos sobre los problemas que se puedan presentar, y que lo hagan con el tiempo suficiente para tomar la acción adecuada.
- 6.2.3.13 La infraestructura técnica debe incluir también los medios para **almacenar metadatos y ligar confiablemente los metadatos** con los objetos digitalmente almacenados. Operaciones mayores encuentran la necesidad de construir sistemas de administración de objetos que se encuentran ligados –aunque separados– con su sistema de almacenamiento masivo digital, de forma que puedan hacer frente al rango de procesos involucrados, y para permitir a los metadatos e interfases de trabajo que cambien sin tener que modificar el almacenamiento masivo.

6.3 Formatos y tipos de cinta de datos

6.3.1 Introducción

- 6.3.1.1 A continuación se describen algunos de los principales formatos en cinta de datos y sistemas de automatización que pueden emplearse para almacenar archivos A/V en forma de datos. Las cintas de datos son usadas solamente en conjunto con otros componentes de un DMSS. Es prudente iniciar esta sección de comparación de formatos de datos con la idea de que ningún soporte es permanente, y que todos, aun siendo iguales, sólo serán viables en la medida en que los sistemas de datos en los cuales se están incorporados continúen soportándoles.

6.3.2 Cinta de datos, formatos y tipos

- 6.3.2.1 La geometría del formato y las dimensiones gobiernan el funcionamiento. La velocidad de transferencia de datos –un aspecto del funcionamiento– es producto directo del número de pistas escritas y leídas simultáneamente, así como la velocidad relativa entre cabeza y cinta, densidad lineal y código de canal. Similarmente, los empaques de cinta más pequeños y ligeros hacen más rápida la robótica de la librería. La densidad de datos es producto de:

- El compromiso entre la longitud de la cinta y su espesor.
- Ancho e inclinación de la pista grabada.
- Densidad lineal de la carga de datos dentro de cada pista.

6.3.3 Recubrimientos de cinta

- 6.3.3.1 Existen dos tipos principales de recubrimientos de cinta: de partículas y evaporado. Las primeras cintas de datos usaron recubrimientos de metal óxido similares a la cinta de video, mientras que las cintas de datos más recientes usan partículas de metal (MP). Hierro puro con cerámica inerte y capas de pacificación son dispersadas sobre aglutinantes de polímero, los cuales son aplicados uniformemente a una base o sustrato PET o PEN que suministra estabilidad dimensional y fuerza bajo tensión. Algunas de las cintas de la más alta densidad de datos disponibles comúnmente en el mercado ahora usan recubrimientos delgados de aleaciones de

cobalto
rial ma
evapor
cintas
Varias
1993: 1

6.3.4 Diseñ

- 6.3.4.1 Dos est
mitir tie
mayor c

6.3.4.2 Dentro

- 3.8
- QIC
- For
- DTF.
- Stor
- El es

6.3.4.3 Los cart

- MTC
- (9, I
- DLT
- LTO
- Stor
- Son

6.3.4.4 Ningún

su durac
ejemplo,
tienen d
cinta má
la cinta
han mej
cionarse
hacia cu
rebobina
que solar
generalm
como es
especific

6.3.5 Cintas

- 6.3.5.1 Las cinta
mente c
cintas lin
provoca