



Introducción a la preservación de medios ópticos

por alex duryee
especialista en preservación digital y metadatos
avpreserve



Introducción

El papel de los medios ópticos en los archivos cambió en la década pasada desde un medio de preservación a un formato en riesgo. Cuando los modelos de longevidad estimaron una vida útil de 25 a 200 años para los medios de grabación,¹ las pruebas recientes descubrieron que ese rango era muy optimista en órdenes de magnitud. Una colección de CD-R de datos de la década de 1990 produjo una proporción de errores de 92% después de unos veinte años de almacenamiento.² Debido a que solo recientemente los archivos se acercaron a los medios ópticos como medio de preservación, existe poca literatura e investigación relacionadas con el formato. Dado que es imposible llevar la cuenta de todos los problemas de migración y preservación de la información almacenada digitalmente, existe la necesidad de un panorama general más extenso sobre los medios ópticos, el cual pretende proporcionar este documento.

Este documento cubrirá cinco de los estándares más comunes de almacenamiento de información en medios ópticos. Al hacerlo, espera informar a las instituciones que trabajan con dichos medios respecto de la naturaleza y el reto de los medios ópticos desde el punto de vista de su preservación. El autor también espera que, al ofrecer un panorama general de la preservación de medios ópticos, se originen más conversaciones e investigaciones del medio.

Obsérvese que este artículo limita deliberadamente su alcance a los formatos y usos más comunes de los medios ópticos. El total de las formas y funciones de los medios ópticos son muchos y variados, desde el video analógico hasta la información oculta a los gráficos y almacenada en códigos de control, y están fuera del alcance de este documento. Se recomienda a los archivistas que trabajan con dichos discos buscar más recursos especializados, tales como comunidades técnicas y documentación de estándares. Este documento, si bien ofrece recomendaciones, no es un recetario de flujos de trabajo ni de herramientas para trabajar con medios ópticos ya que aún se tiene que realizar la investigación necesaria para hacer tales recomendaciones.

Estructura lógica

Es crucial comprender la distribución lógica de los medios ópticos antes de intentar cualquier actividad de preservación. Dado que el primer estándar para medios ópticos fue el estándar IEC 60908 (1982) para el almacenamiento y reproducción de audio, los estándares posteriores mostraron un enfoque hacia los medios audiovisuales en la manera en que estructuran el disco. A pesar de que los estándares tales como ISO/IEC 10149 (disponibles como ECMA-130³) que refieren al almacenamiento sobre un sistema de archivos, el idioma y la estructura del Libro Rojo prevalece.⁴ Por lo tanto, aunque parece extraño discutir sobre información basada en un sistema de archivos en términos de pistas lógicas, así es precisamente como se almacena la información en un CD-ROM.

Un modelo simplificado de la estructura lógica de un disco compacto es, como tal: una serie de sesiones donde cada una contiene una serie de pistas. Una pista, como su nombre da a entender, está diseñada para que sea una pista discreta de audio. Respecto de la información que no es de audio (p. ej., CD-ROM), una sola pista puede contener cualquier cantidad de sistemas de archivos. Estas pistas están arregladas en series lineales unidas por una zona de entrada (que tiene el índice, los metadatos de ubicación y descriptivos para las pistas a continuación) y una zona de salida. Esta colección comprende una sesión. Los primeros CDs estaban diseñados como si una sola sesión estuviera en el disco; esto se expandió en 1990 para contener muchas sesiones.

¹ Range Commanders Council, Optical Systems Group. Archivo multimedia: Medios de cintas de video, discos compactos (CD), discos digitales versátiles (DVD) y Blu-Ray (BD). White Sands Missile Range, Nuevo México: Range Commanders Council (actualizado en febrero de 2010, citado el 14 de Febrero de 2014). Disponible en http://www.wsmr.army.mil/RCCsite/Documents/462-10_Multimedia%20Archiving%20-%20CD,%20DVD,%20and%20Blu-ray/462-10_Multimedia%20Archiving%20-%20CD,%20DVD,%20and%20Blu-ray.pdf

² Wilsey L, Skirvin R, Chan P, Edwards G. Capturing and Processing Born-Digital Files in the STOP AIDS Project Records. Caso de estudio. Journal of Western Archives. Citado el 14 de febrero de 2014. Disponible en <http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1026&context=westernarchives>

³ Ecma International. Intercambio de información en discos ópticos de datos (CD-ROM) de solo lectura, de 120 mm. Segunda edición. Ginebra, Suiza: Ecma International. Actualizado en junio de 1996, citado el 14 de febrero de 2014. Disponible en <https://web.archive.org/web/20130116084229/http://ecma-international.org/publications/files/ECMA-ST/Ecma-130.pdf>

⁴ El Libro Rojo es el nombre coloquial para el estándar IEC 60908 debido al color de su portada. Los distintos estándares que definen los medios ópticos se conocen como los Libros Arcoíris ya que cada uno tiene un color de portada diferente (p. ej., el ISO/IEC10149 es el Libro Amarillo).

| | | | | | | | | | |
|------------------|----------|----------|----------|----------------|------------------|----------|--------|-----|----------------|
| Contenido | audio | audio | audio | | Contenido | iso9660 | joliet | hfs | |
| Zona de entrada | pista 01 | pista 02 | pista 03 | Zona de salida | Zona de entrada | pista 01 | | | Zona de salida |
| Sesión 01 | | | | | Sesión 02 | | | | |

Las configuraciones de muestra a continuación son posibles gracias a la flexibilidad de la forma en que se puede arreglar la información en sesiones y pistas:

- Una sesión, muchas pistas de audio (CD de audio típico).
- Una sesión, una pista, muchos sistemas de archivos (CD de datos típico).
- Dos sesiones, una con muchas pistas de audio, la otra con una pista y muchos sistemas de archivos (CD mejorado).
- Una sesión, una pista de datos, muchas pistas de audio (videojuego).

Merece destacarse que se puede abrir o cerrar una sesión en un CD-R y en un CD-RW. Se pueden escribir pistas nuevas en una sesión mientras esté marcada como abierta; al cerrarla, evita que se agregue más información y crea el índice final. Los métodos para manejar los discos con sesiones abiertas están documentados por la comunidad forense digital debido a su importancia en la recolección de evidencia.⁵

Para los fines de este artículo, la estructura de un DVD es suficientemente parecida a la de un CD-ROM, por lo que no amerita un análisis más profundo.

ISO9660

El ISO9660 es el formato más común para transportar información en CD-ROM y aparece a menudo como el sistema de archivos básico de los discos multiplataforma. También es común que sirva como el sistema de archivos base de sistemas más sofisticados, tales como Joliet y HFS; estos, a menudo (aunque no siempre) funcionan como capa por encima de la información ISO9660 para proporcionar una funcionalidad adicional. Sus limitaciones (tales como los nombres de archivo no Unicode estilo 8.3) lo hacen compatible con casi todos los sistemas operativos comunes, ofreciendo entonces una alternativa en caso de que los sistemas de archivos más avanzados no sean compatibles. Como tal, es extremadamente común encontrar discos compactos con un sistema de archivos ISO9660 junto con una mezcla de HFS y Joliet; tal disco aporta funciones avanzadas para Windows y MacOS, pero se podría tener acceso a la información desde virtualmente cualquier sistema operativo.

Desde la perspectiva de la preservación, podemos tratar al ISO9660 de manera similar al sistema de archivos en medios magnéticos. Mientras que cada sector debe entregar información consistente en cada lectura (consideren las consecuencias de un sector sin usar cuando se ejecuta el código compilado) y cada sector debe encontrarse fácilmente en el disco; estos discos usan la estructura de sectores Modo 1. En contraste, los CD-DA usan cada byte en un sector por dato, el Modo 1 dedica 2048 bytes de datos, 16 bytes a la sincronización e identificación de sectores y 288 a la detección y corrección de errores. Estos bytes que no contienen información disminuyen la cantidad de información del usuario en el disco pero permiten, de manera crucial, que se comporte como un sistema de archivos proporcionando una búsqueda rápida y precisa de sectores y una lectura consistente de información.

⁵ Puede encontrar más información sobre esta teoría y método para recuperar información de discos con sesiones abiertas en los Análisis forenses de CD y DVD de Crowley y Kleiman.

Como tal, para fines de migración, preservación a largo plazo y acceso, un disco ISO9660 se puede tratar de manera similar a un disco magnético. Si bien no hay necesidad del hardware especializado que requieren los medios magnéticos (tales como bloqueadores de escritura), se pueden obtener imágenes de las pistas ISO9660 a nivel de byte⁶ mediante la variedad de estándares de herramientas de transferencia de datos (dd, guymager, FTK, IsoBuster, etc.)⁷. La imagen subsecuente se puede montar luego como cualquier otro sistema de archivos para fines de análisis de archivo y acceso del usuario. Previamente se exploraron modelos de acceso y flujos de trabajo post-migración, tales como el montaje remoto de imágenes y el análisis automatizado del sistema de archivos, de manera similar y paralela a la investigación en medios magnéticos.⁸

Joliet / HFS

Mientras que el estándar ISO9660 es el formato de datos más común para CD-ROM, es raro verlo por sí solo. Debido a las restricciones en las estructuras de datos, existió demanda por algún mecanismo para expandir las capacidades de los sistemas de archivos de los CD-ROM. Para atender esto, en 1995, Microsoft estableció la especificación Joliet como una extensión al ISO9660. Las principales mejoras de Joliet sobre ISO9660 son los nombres de archivos largos, los nombres de archivos Unicode y árboles de directorios con más exigencias. Ya que Joliet no maneja la información del disco, solo proporciona metadatos mejorados del sistema de archivos, típicamente está empaquetada como una capa próxima al sistema de archivos ISO9660. Una exploración de la estructura lógica del disco mostrará que un archivo ISO9660 y su contraparte Joliet apuntan al mismo sector del disco; mientras que el nombre difiere, se usa una copia de la información.

El Sistema de archivos jerárquico (HFS) se empleó para permitir comportamientos de archivos específicos de Macintosh en los CD-ROM y para ofrecer una solución a las limitaciones de ISO9660. El HFS, específico de las máquinas Apple, es un sistema de archivos notablemente más poderoso que ISO9660, permite que los nombres de archivos más largos y los metadatos necesarios se integren de manera más natural con el sistema operativo⁹. Como tal, se usó mucho para proporcionar compatibilidad con MacOS. Por ejemplo, un documento sin los metadatos necesarios específicos del HFS carecerá de las señales de formato/software necesarias para que abra correctamente. Además, incluso en ambientes donde el sistema de archivos HFS no es más que un puntero de metadatos a la información ISO9660 subyacente, el sistema de archivos HFS es crítico para mantener la compatibilidad del disco. Típicamente, los sistemas de archivos HFS contienen información única del disco en MacOS, tal como un código de bytes compilado y documentación específica del sistema operativo, los cuales deberán migrarse como parte de cualquier flujo de trabajo de archivo.

⁶ Obsérvese que el término "a nivel de bit" carece de significado al tratarse de medios ópticos. En vez de usar ocho bits físicos por byte, todos los discos compactos usan la modulación ocho a catorce (EFM): las 256 combinaciones de catorce bits más favorables para una lectura precisa se mapean hacia los bytes de ocho bits correspondientes y en el disco se escriben las palabras de catorce bits. Como tal, una interpretación "a nivel de bit" de un disco compacto requeriría un análisis del artefacto físico. Para mayor información, consulte el enlace del estándar ECMA-130 mencionado anteriormente.

⁷ Nótese que el término coloquial "Imagen ISO" es bastante engañoso. A pesar de implicar que un archivo de imagen de disco que termina en .iso solo captura información ISO9660, en realidad se usa para describir una imagen de disco general. Como tal, a menos que el origen del archivo .iso esté documentado, no puede usarse para determinar el alcance y contenido de la migración.

⁸ Woods, Kam A. (2010). Preservando el acceso a largo plazo a los documentos del gobierno de los Estados Unidos de América en formatos digitales antiguos (disertación). Bloomington: Universidad de Indiana. Disponible en <https://web.archive.org/web/20140217161345/http://www.digpres.com/publications/Kam-Woods-Dissertation-Pre.pdf>

⁹ Por ejemplo, el código creador y el tipo de código permiten que MacOS abra un archivo automáticamente usando el software requerido.

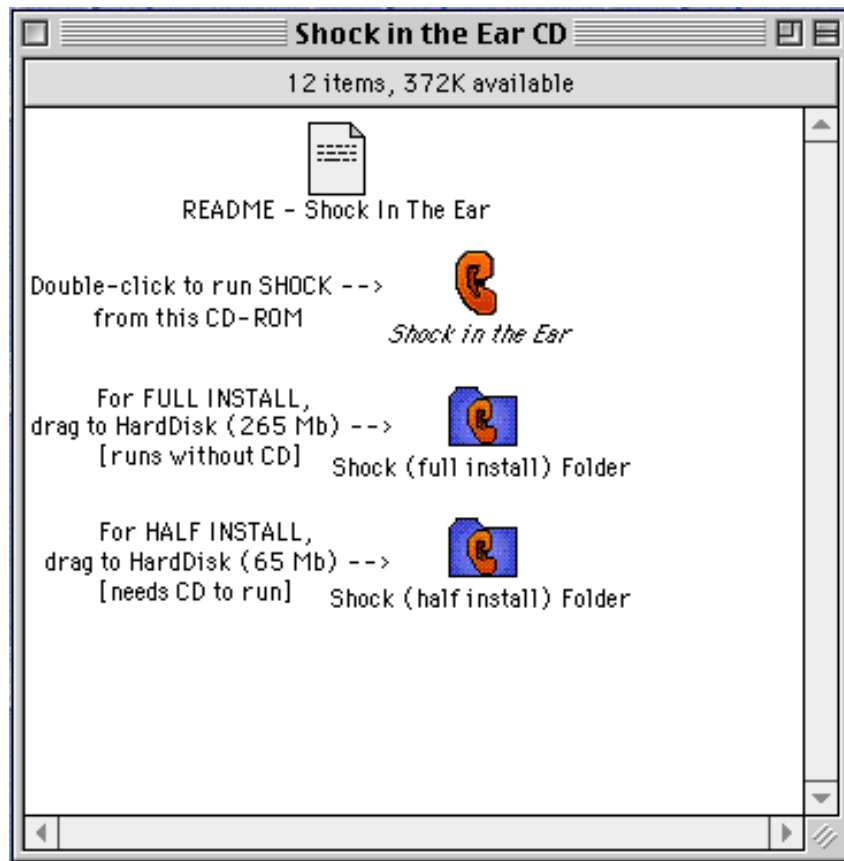


Figura 2. Un disco que contiene un sistema de archivos HFS, visto en un emulador. Observe el texto flotante en la cubierta; esto se logró creando archivos vacíos con un ícono en blanco y colocándolos dentro de la ventana del Buscador.

UDF

El UDF se diseñó como un estándar universal para el almacenamiento de datos y tiene pocas sorpresas para su preservación. Contrario al disco compacto, un DVD (salvo en circunstancias inusuales) casi siempre se comportará de manera estándar en cualquier tipo de dispositivos. A pesar de las variaciones aparentes entre los DVD de video, DVD de audio y DVD de datos, el formato de disco subyacente es idéntico. Aunque hay una variedad de peculiaridades entre las versiones e implementaciones del UDF y casos muy específicos donde sus funciones estructurales son críticas para una migración exitosa, estas están más allá del alcance de la preservación diaria.

Obsérvese que, debido a que el soporte de hardware cambia según las versiones de UDF, normalmente está asociado con un sistema de archivos ISO9660. Gracias al soporte casi universal de ISO9660, este puede usarse como "puente" para permitir que el hardware viejo lea el disco. Como tal, depende del archivo individual decidir si se debe preservar uno o ambos sistemas de archivos.

El DVD de video es una subclase importante de DVD, ya que es el formato predominante para el video de consumo masivo. Usando un enfoque de pantalla esencial, el DVD de video aparece como el equivalente de los formatos de cinta audiovisual: el usuario inserta el disco en un reproductor que muestra un menú interactivo y una colección de flujo audiovisual lineal. Sin embargo, esto enmascara la estructura interna del disco. En realidad, el DVD contiene un sistema de archivos UDF con un directorio y una estructura de archivos estandarizados (que

un reproductor de DVD reconoce y analiza en un flujo aparentemente único). En general, el sistema de archivos contiene un directorio VIDEO_TS que contiene los flujos MPEG (VOB) y los metadatos de reproducción (IFO/BUP)¹⁰. Para fines de migración, el DVD de video puede tratarse de manera similar a cualquier DVD de datos. El mayor problema que los archivos enfrentarán con los DVD de video es el Sistema de Codificación del Contenido (un esquema de protección contra copia que impide la migración del disco y, por lo tanto, su preservación) y proporcionar acceso a los archivos audiovisuales sin procesar después de la migración.

CD-DA

La estructura de un CD-DA es más similar a una cinta que a un disco de datos tradicional. En vez de dividir el área de almacenamiento en archivos discretos, la información del CD-DA se escribe como una transmisión lineal de modulación de impulsos codificados (MIC) divididos en pistas separadas. Conforme la unidad de disco lee la información (a 44,100 muestras de 16 bits por segundo, seleccionadas específicamente para estar por encima de la tasa necesaria mencionada anteriormente para su reconstrucción humanamente perfecta), el dispositivo de reproducción interpreta la transmisión MIC y genera la onda correspondiente. En este sentido, el CD-DA se acerca más a los formatos en cinta que al disco magnético tradicional, ya que no representa un sistema de archivos estructurado sino un flujo lineal de medios.



Figura 3. La modulación de impulsos codificados hacia la señal digital hacia la cadena de señal analógica

Ya que el CD-DA es, fundamentalmente, diferente de los formatos de audio, tales como ISO9660, existe una cantidad de factores a considerar. Ya que no hay “archivos” en una sesión CD-DA sino flujos de información MIC sin procesar, la lista de archivos proporcionada por el shell de un sistema operativo no refleja el contenido del disco. Windows, por ejemplo, mostrará los archivos .cda que se reproducirán mediante una aplicación para medios. Sin embargo, estos archivos no contienen información, son meramente punteros hacia los lugares dentro del CD donde empieza su pista correspondiente. La información en sí es algo que el shell no puede analizar y por ende requiere de software especial para migrar.

El CD-DA también fue diseñado para maximizar el espacio perdiendo precisión; por ello, sacrifica un tercer nivel de corrección de errores en pos de más información por sector¹¹. La ventaja de este enfoque fue un marcado aumento en la capacidad y un nivel de tolerancia de errores, ya que los sectores que no se leen bien no interrumpirán una reproducción tranquila en todos los casos sino en los más atroces. Sin embargo, los errores presentados durante el proceso de lectura pueden pasar inadvertidos para la unidad. Típicamente, los errores de lectura se manifiestan como problemas de sonido tradicionales, tales como clics y pops.

En consecuencia, leer una pista de audio en una sola pasada proporcionará resultados poco confiables ya que el hardware general tendrá apenas 95% de precisión a nivel de pista.¹² Esto es inaceptable para fines de preservación, ¡imaginen las consecuencias de introducir errores en un 5 o 10% de las cintas y discos durante la transferencia! Se han desarrollado varias técnicas para

¹⁰ La especificación de DVD video también define un directorio AUDIO_TS. Esto solo se usa en el exótico formato de DVD-Audio y es más probable que no se encuentre en la mayoría de los escenarios de archivo. Para fines de DVD-video, siempre debe de estar vacío.

¹¹ Los detalles de la corrección de errores de la información en los CD-DA son lo suficientemente complejos como para exceder el alcance de este documento. Para mayor información, consulte: <https://web.archive.org/web/19970616191108/http://www.ee.washington.edu/conselec/CE/kuhn/cdmulti/95x7/iec908.htm>

¹² Lista de precisión de la unidad de CD/DVD. Actualizado el 14 de mayo de 2013. Citado el 14 de febrero de 2014. Disponible en <https://web.digpres.com/publications/Kam-Woods-Dissertation-Pre.pdf>

informar sobre las pérdidas inherentes del CD-DA. El software diseñado para una extracción confiable de CD-DA combina una variedad de métodos para asegurar que las lecturas sean consistentes en un disco dado. Por ejemplo, una herramienta de extracción podría leer cada sector del disco varias veces con la finalidad de encontrar el valor correcto o podría comparar la información contra las bases de datos en línea que contengan otros flujos de bits del mismo disco.

Si bien estos métodos son poderosos en asegurar buenas lecturas consistentes de un CD-DA, son específicos de esta estructura de datos y, por lo tanto, no se pueden portar a ISO9660 u otros formatos de CD. Esto representa un problema durante la migración de los llamados CD Mejorados (ECD, también conocidos como CD Extra) y los CD de modo mixto (frecuentes para los videojuegos), que contienen información de audio y de sistemas de archivos. Como el flujo de trabajo específico de esos discos depende de la estructura de un disco en particular, excede el alcance de este documento. Ya que es improbable que estos se encuentren fuera de las colecciones específicas en cuestión, los archivos que manejan tales medios harían bien en realizar una investigación y comunicarse con especialistas al respecto de la migración.

Aplicaciones en flujos de trabajo de preservación

El análisis más importante que un archivista puede realizar en un disco sucede antes de que este, incluso, entre a la estación de trabajo. Frecuentemente, la gente usa los discos ópticos grabables como porta archivos portátiles simples. Por ejemplo, un donador puede grabar un DVD con material para transportarlo como parte de su colección, o se puede crear un disco para compartir documentos dentro de una oficina. En estos casos, el disco sirve como un simple transporte, análogo a un sobre o paquete. En otras palabras, algo que nunca amerita preservarse. Como tal, el archivista puede decidir copiar los archivos mediante un shell (Buscador, Explorador de Windows, etc.) en vez de crear una imagen de disco ya que no es útil preservar la estructura de bajo nivel. Dependiendo de la naturaleza de los discos en una colección determinada, este podría ser el método de migración preferido.

Los aspectos físicos de los medios ópticos, si bien no son importantes para la información en el disco, merecen ser considerados al momento de realizar procesos de migración. La velocidad de lectura de un disco debe ser la más lenta posible (usando herramientas tales como la herramienta para CD-ROM SPTI o hdparm) ya que las velocidades más lentas rinden resultados más precisos. La calidad de la unidad de grabación también puede causar una gran diferencia, la información empírica ha descubierto que las unidades de grabación para consumo masivo de alta calidad son 5% más precisas a nivel de pista que las menos eficientes¹³ y que se deben tomar en cuenta al momento de tomar una decisión para elegir una estación de trabajo para el archivo. Tradicionalmente, las unidades Plextor han sido usadas por varias comunidades enfocadas hacia la preservación y, por ende, recomendadas para la migración óptica.¹⁴

Dado que un disco óptico puede tener cualquier cantidad de sistemas de archivos y estructuras de sectores, algunos de los cuales son invisibles mediante un shell del sistema operativo, es necesario usar herramientas dedicadas para el análisis. Al leer directamente la tabla de contenidos de un disco (en vez de confiar en lo que el sistema operativo reconoce), se vuelve posible ver una lista completa de pistas en un disco (y su contenido correspondiente en caso de las pistas de datos). Existe un puñado de herramientas para explorar la estructura y contenidos de un disco, tales como CD/DVD Inspector e IsoBuster (ambas son solo para Windows). Esto permitirá que un archivista comprenda de forma más amplia la estructura de un disco (sesiones/pistas, sistemas de archivos, archivos, etc.), así como encontrar cualquier disco que contenga CD-DA antes de la migración.

¹³ *Ibid.*

¹⁴ Obsérvese que Plextor terminó la producción de su propio hardware a mediados de la década del 2000; cualquier unidad con el nombre Plextor creada después de ese punto fue fabricada por una compañía diferente.

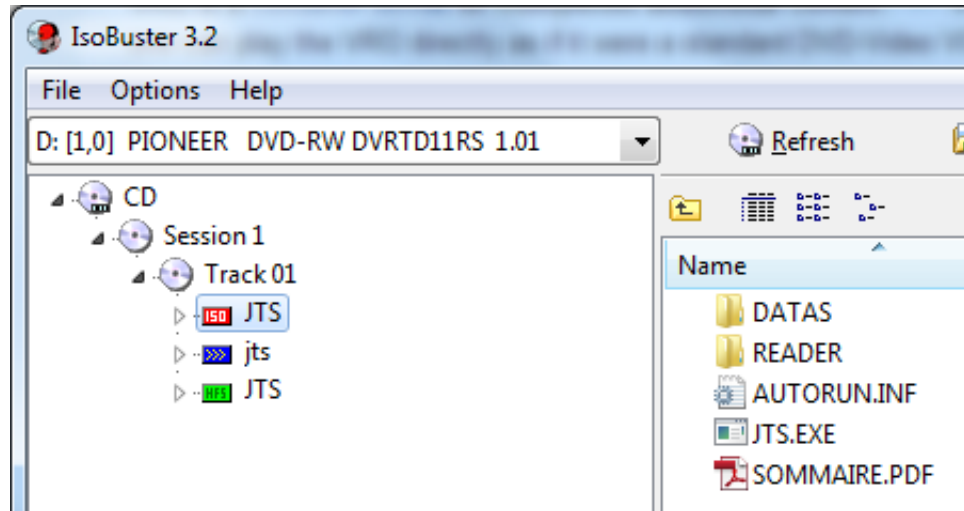


Figura 4. IsoBuster con un disco que contiene un sistema de archivos ISO9660 extendido mediante capas Joliet y HFS

Debido a la falta de corrección de errores de los CD-DA, en comparación con otros formatos de CD/DVD, requiere de un flujo de trabajo especial para migrarlo del CD adecuadamente. Una lectura lineal de un CD-DA no nos dará una transferencia de datos precisa debido a las causas mencionadas anteriormente. Además, los problemas específicos de audio (tales como colocación de silencios entre las pistas) son problemas que no se atienden en el software más general. Para solucionar estos problemas, se desarrolló software especializado que atiende los problemas que existen en la extracción de CD-DA. Las herramientas que el autor recomienda son Exact Audio Copy (Windows), dBpoweramp (Windows) and cdparanoia (Linux/OSX/Windows), que fueron diseñados específicamente para superar las barreras de la migración mediante métodos para reconocer errores de lectura de silencios, detección previa de espacios y demás. La combinación de información empírica y del análisis técnico de estas herramientas hace que sean un estándar de oro para la migración de CD.

Los formatos de destino para la migración óptica dependen de una variedad de factores. Una imagen completa que contenga cada byte en un disco no es necesariamente el formato óptimo para la preservación a largo plazo ya que va a incluir corrección de errores y datos de sincronización. Esto aumentará el tamaño de una imagen en 305 bytes por sector (96 MB en todo el disco) proporcionando una utilidad mínima y creando potenciales problemas de errores. Sin embargo, se puede replicar perfectamente un disco al usar una imagen binaria (típicamente un archivo .bin) y una hoja de entradas (archivo .cue que proporciona los metadatos necesarios para dividir el sistema binario en sesiones/pistas e interpretarlo). Esto se puede necesitar en ciertos casos para crear y preservar discos maestros, particularmente en el caso de discos complejos. Una alternativa al manejar los discos de una sola pista es una imagen de una sola pista que capturará los datos del usuario dentro de la pista y proporcionará una imagen montable y usable. En el Apéndice se encuentran los flujos de trabajo generales para la creación de discos maestros de preservación.

Para los CD-DA, el formato de migración de destino es, típicamente, el formato WAVE de 16-bit/44,1 kHz. Dado que la información MIC sin procesar es el equivalente al archivo WAVE sin encabezados, el encabezado se puede agregar sin cambios a los datos de audio. Luego este archivo puede manipularse con las especificaciones de un archivo para transcodificar, transmitir metadatos WAVE, etc. Si la estructura del disco es importante, el formato bin/cue es una alternativa a los archivos WAVE discretos que preservan el tiempo y la disposición del disc.¹⁵ En el apéndice se sugiere un flujo de trabajo que se usa para una copia de audio exacta.

¹⁵ El ejemplo canónico de bin/cue que es necesario para el CD-DA es la música diseñada para bailar, que normalmente no usa espacios entre pistas. Dado que el formato bin/cue permite la manipulación directa de espacios, esto habilita que la sesión se reproduzca como se pretende.

Siendo que un DVD de video está estructurado de manera diferente a otros DVD de datos, la pura imagen binaria del disco es suficiente para su preservación. Dado que no hay necesidad real de convertir flujos MPEG dentro de los archivos VOB para su preservación, la imagen puede almacenarse tal cual y montarse cuando se le solicite acceso. Sin embargo, debido a la variedad de necesidades y métodos para preservar, describir y acceder a los videos digitales, es imposible sugerir un solo formato de destino; por ejemplo, un sistema de administración de activos puede requerir un códec específico para su reproducción. Mientras que los archivos VOB son estables y generalmente reproducibles, puede existir la necesidad de transformar los archivos para que se pueda tener un acceso más amplio/fácil (p. ej., concatenando los archivos divididos en un solo objeto). El archivo deberá establecer políticas para manejar la protección contra copia, transcodificación, almacenamiento y para proporcionar acceso a los medios.

Investigación posterior

El estado de preservación actual con respecto a los medios ópticos es escueto, ya que limita su alcance al artefacto físico. Al ser relevante en el corto plazo, la inevitabilidad del deterioro del disco, junto con la disminución de la calidad de las unidades de grabación por parte de los fabricantes, hacen que sea urgente e imperativa la investigación y práctica de migraciones. Siendo que se ha realizado poca investigación en los campos aliados tales como la aplicación de la ley y las ciencias computacionales, existe una mayor necesidad de que los archivistas establezcan prácticas y conocimientos relacionados con los medios ópticos.

Las áreas para una futura investigación pueden incluir:

- La documentación de las mejores prácticas relacionadas con proveedores ópticos conocidos y desconocidos.
- Establecer las técnicas y flujos de trabajo para medios dañados.
- Mejorar la migración de medios ópticos y equilibrar las mejores prácticas con eficiencia.
- Describir las propiedades significativas de las clases de medios ópticos.
- Estándares de metadatos para describir los medios ópticos y la información almacenada como tal.
- Exploración de los métodos para preservar los formatos ópticos más exóticos, tales como los formatos basados sobre CD+G y LaserDisc.

El autor anticipa que, conforme el alcance del acervo empieza a abarcar la era del almacenamiento óptico, se incrementará la necesidad de flujos de trabajo establecidos y de investigación avanzada. El uso de medios ópticos como objetos para la preservación de archivos está limitado a proyectos específicos (documentos gubernamentales y arte digital) y a comunidades externas (particularmente los coleccionistas de música y videojuegos). Los archivistas pueden prepararse para la siguiente generación de retos de la conservación digital aplicando y adaptando el conocimiento existente para generalizar la conservación óptica.

Apéndice

Flujo de trabajo sugerido para CD-ROM

- Analizar el disco con ISO Buster y determinar el flujo de trabajo (con un enfoque particular en CD-DA, número de pistas/sesiones y errores iniciales).
- Si el disco solo es ROM, analizar su estructura (ISO, HFS, híbrida, etc.) y describirla de acuerdo con el estándar de metadatos.
- Extraer la imagen del CD -> Información del usuario, o si son necesarios los bytes que no son de datos (para alineación a nivel byte), RAW (sin procesar).

Esto capturará todas las sesiones/pistas del CD-ROM en un disco en una imagen binaria. Si se seleccionó la información del usuario, esta se puede montar como un disco dentro del sistema operativo. Si se seleccionó el formato RAW, es posible que se necesite extraer la información del usuario antes de montar la imagen. Las imágenes RAW se pueden traducir en información del usuario para montarla usando la herramienta bchunk en Linux/OSX.

Flujo de trabajo sugerido para CD-DA

Debido a la complejidad (y necesidad) de configurar adecuadamente una copia de audio exacta (EAC), esto excede el alcance de un documento general. En Hydrogen Audio Knowledge Base, se puede encontrar una guía sobre la variedad de software y opciones de unidades de grabación.¹⁶ Obsérvese que, mientras que las opciones de software para EAC pueden generalizarse para su preservación, cada unidad de grabación es única y, por ende, requiere de una configuración individual. También es prudente alinear la unidad de grabación con la base de datos AccurateRip para comparar la precisión de cada unidad de grabación contra otras del mismo modelo.

- Analizar el disco con ISO Buster y determinar el flujo de trabajo.
- Si todo el disco contiene pistas de CD-DA, abrir el EAC y leer el disco desde ahí.
- Detectar espacios previos y buscar desviaciones inusuales (típicamente los espacios son de 2 segundos cada uno).
- Probar y copiar las pistas seleccionadas -> Sin comprimir.

Esto generará archivos WAVE que cumplen con las especificaciones de las preferencias de la EAC. Debido al énfasis de la EAC en la lectura confiable de CD-DA (usando métodos tales como comparar varias lecturas entre ellas) la migración de la información del disco será muy precisa. Si hay necesidad de mantener la estructura en el disco mediante una hoja de entradas, esta se puede crear dentro de la EAC usando las configuraciones de espacio actuales.

Glosario

CD: Disco compacto. Formato diseñado para contener audio y datos (y más tarde expandido a datos generales) en un disco de plástico de 12 cm, usando un láser para leer una serie de pozos y llanos como datos binarios.

CD-DA: Disco compacto. Audio digital estándar usado para definir las estructuras lógicas, físicas y de datos de los discos de audio.

CD-R: Disco compacto grabable. Formato WORM (una escritura, varias lecturas) que permite que los discos compactos se creen usando hardware para consumo masivo.

CD-ROM: Disco compacto, memoria de solo lectura. Estándar usado para definir las estructuras lógicas, físicas y de datos de los discos de datos.

CD-RW: Disco compacto regrabable. Un formato similar al CD-R que permite volver a escribir información en el disco.

DVD: Disco digital versátil. Formato diseñado para almacenar películas con duración cinematográfica en un disco de 12 cm. Esto llevó a la creación de un disco que ofrece 6 a 10 veces la capacidad de un CD-ROM con el mismo factor de forma.

DVD-Video: Sistema de archivos y formatos de archivo estándares para distribuir películas en DVD.

Enhanced CD: Formato para almacenar información de CD-DA y CD-ROM en un solo disco. Originalmente esto era imposible, pero se logró en 1995 con el estándar del Libro Azul.

HFS: Sistema de Archivos Jerárquico. Sistema de archivos estándar que se usa en MacOS y que se usó en un CD-ROM para proporcionar compatibilidad y funciones mejoradas.

¹⁶ Opciones EAC. Actualizado el 3 de agosto de 2013. Citado el 14 de febrero de 2014. Disponible en http://wiki.hydrogenaudio.org/index.php?title=EAC_Options

Preservación de medios ópticos

ISO9660 (“ISO”): Estándar usado para definir la ubicación del sistema de archivos y la estructura de archivos en discos compactos. A pesar de que es muy limitado, es casi ubicuo de los CD-ROM.

Joliet: Una extensión de ISO9660 diseñada por Microsoft, que proporciona una capa de metadatos para superar las limitaciones de ISO9660.

Modo 1: Un subestándar de la estructura de datos de CD-ROM. El Modo 1 define 2048 bytes de información del usuario por sector, usando el resto del sector para la corrección de errores y bytes de sincronización. El Modo 1 se usa mucho más que el Modo 2, que fue diseñado para usarse en formatos más exóticos tales como CD-i y Video-CD.

MIC: Modulación por impulsos codificados. Método para almacenar información de audio analógica en una serie de valores binarios.

UDF: Formato de disco universal. Un estándar diseñado para sustituir al ISO9660 quitando algunas de sus restricciones. El UDF se creó con la finalidad de reemplazar los miles de formatos de CD de datos con uno solo.

AVPreserve es una empresa de consultoría con servicio completo de archivos de medios y administración de datos. Nos asociamos con archivos, museos, organismos gubernamentales, corporaciones, medios y entretenimientos, y otras organizaciones que crean o recolectan medios para ayudarles a administrar, acceder y conservar sus valiosos activos y datos. Nuestros servicios abordan el ciclo de vida completo de las colecciones, desde la evaluación y planificación de preservación para materiales analógicos, incluyendo la administración de proyectos de actividades de digitalización, hasta los distintos aspectos de la preservación digital y administración de archivos, incluyendo la selección del monitoreo de actividad en bases de datos (DAM), el desarrollo de la taxonomía, las políticas y los flujos de trabajo, y el desarrollo de soluciones de software que apoyan la preservación y el acceso.